Texas Law Review

March 2014 Vol. 92 No. 4 Pages 773 to 1052

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

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

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

Single issues in the current volume may be purchased from the *Texas Law Review* Publications Office for $10.00 per copy plus shipping. Texas residents, please add applicable sales tax.

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

Editorial Offices: *Texas Law Review*
727 East Dean Keeton Street, Austin, Texas 78705
(512) 232-1280 Fax (512) 471-3282
tlr@law.utexas.edu
http://www.texaslawreview.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.*
MARIA M. ARRELLAGA, B.S.; *Assistant Dean for Communications.*
ELIZABETH T. BANGS, A.B., J.D.; *Assistant Dean for Student Affairs.*
KIMBERLY L. BIAR, B.B.A.; *Assistant Dean for Financial Affairs, Certified Public Accountant.*
MICHAEL G. HARVEY, B.A., B.S.; *Assistant Dean for Technology.*
MONICA K. INGRAM, B.A., J.D.; *Assistant Dean for Admissions and Financial Aid.*
TIMOTHY A. KUBATZKY, B.A.; *Executive Director of 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

HANS W. BAADE, A.B., J.D., LL.B., LL.M.; *Hugh Lamar Stone Chair Emeritus in Civil Law.*
RICHARD V. BARNDT, B.S.L., LL.B.; *Professor Emeritus.*
WILLIAM W. GIBSON, JR., B.A., LL.B.; *Sylvan Lang Professor Emeritus in Law of Trusts.*
ROBERT W. HAMILTON, A.B., J.D.; *Minerva House Drysdale Regents Chair Emeritus.*
DOUGLAS LAYCOCK, B.A., J.D.; *Alice McKean Young Regents Chair Emeritus.*
J.L. LEBOWITZ, A.B., J.D., LL.M.; *Joseph C. Hutcheson Professor Emeritus.*
JOHN T. RATLIFF, JR., B.A., LL.B.; *Ben Gardner Sewell Professor Emeritus in Civil Trial Advocacy.*
JAMES M. TREECE, B.A., J.D., M.A.; *Charles I. Francis Professor Emeritus in Law.*

PROFESSORS

JEFFREY B. ABRAMSON, B.A., J.D., Ph.D.; *Professor of Government and Law.*
DAVID E. ADELMAN, B.A., Ph.D., J.D.; *Harry Reasoner Regents Chair in Law.*
DAVID A. ANDERSON, A.B., J.D.; *Fred and Emily Marshall Wulff Centennial Chair in Law.*
MARILYN ARMOUR, B.A., M.S.W., Ph.D.; *Associate Professor.*
MARK L. ASCHER, B.A., M.A., J.D., LL.M.; *Joseph D. Jamail Centennial Chair in Law.*
RONEN AVRAHAM, M.B.A., LL.B., LL.M., S.J.D.; *Thomas Shelton Maxey Professor in Law.*
LYNN A. BAKER, B.A., B.A., J.D.; *Frederick M. Baron Chair in Law, Co-Director of Center on Lawyers, Civil Justice, and the Media.*
MITCHELL N. BERMAN, A.B., M.A., J.D.; *Richard Dale Endowed Chair in Law, Professor of Philosophy, Co-Director of Law and Philosophy Program.*
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.*
NORMA V. CANTU, B.A., J.D.; *Professor of Education and Law .*
LOFTUS C. CARSON, II, B.S., M. Pub. Affrs., M.B.A., J.D.; *Ronald D. Krist Professor.*
MICHAEL J. CHURGIN, A.B., J.D.; *Raybourne Thompson Centennial Professor.*
JANE M. COHEN, B.A., J.D.; *Edward Clark Centennial Professor.*
FRANK B. CROSS, B.A., J.D.; *Herbert D. Kelleher Centennial Professor of Business Law, Professor of Law.*
WILLIAM H. CUNNINGHAM, B.A., M.B.A., Ph.D.; *Professor.*
JENS C. DAMMANN, J.D., LL.M., Dr. Jur., J.S.D.; *William Stamps Farish Professor in Law.*
JOHN DEIGH, B.A., M.A., Ph.D.; *Professor of Philosophy and Law.*
MECHELE DICKERSON, B.A., J.D.; *Arthur L. Moller Chair in Bankruptcy Law and Practice.*
GEORGE E. DIX, B.A., J.D.; *George R. Killam, Jr. Chair of Criminal Law.*
JUSTIN DRIVER, B.A., M.A., M.A., J.D.; *Professor.*
JOHN S. DZIENKOWSKI, B.B.A., J.D.; *Dean John F. Sutton, Jr. Chair in Lawyering and the Legal Process.*
ZACHARY S. ELKINS, B.A., M.A., Ph.D.; *Associate Professor.*
KAREN L. ENGLE, B.A., J.D.; *Minerva House Drysdale Regents Chair in Law, Co-Director of Bernard and Audre Rapoport Center for Human Rights and Justice.*
KENNETH FLAMM, A.B., Ph.D.; *Professor.*
JULIUS G. GETMAN, B.A., LL.B., LL.M.; *Earl E. Sheffield Regents Chair.*
CHARLES E. GHOLZ, B.S., B.S., Ph.D.; *Associate Professor.*
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., Ph.D., M.A., Ph.D.; *Associate Professor.*
CHARLES G. GROAT, B.A., M.S., Ph.D.; *Professor.*
PATRICIA I. HANSEN, A.B., M.P.A., J.D.; *J. Waddy Bullion Professor.*
HENRY T. C. HU, B.S., M.A., J.D.; *Allan Shivers Chair in the Law of Banking and Finance.*

BOBBY R. INMAN, B.A.; *Professor.*
DEREK P. JINKS, B.A., M.A., J.D.; *The Marrs Mcl.ean Professor in Law.*
STANLEY M. JOHANSON, B.S., LL.B., LL.M.; *James A. Elkins Centennial Chair in Law, University Distinguished Teaching Professor.*
CALVIN H. JOHNSON, B.A., J.D.; *Andrews & Kurth Centennial Professor.*
SUSAN R. KLEIN, B.A., J.D.; *Alice McKean Young Regents Chair in Law.*
JENNIFER E. LAURIN, B.A., J.D.; Professor.
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.*
VIJAY MAHAJAN, M.S.Ch.E., Ph.D.; *Professor.*
BASIL S. MARKESINIS, LL.B., Ph.D., D.C.L., LL.D.; *Jamail Regents Chair in Law.*
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.*
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.*
ROBERT J. PERONI, B.S.C., J.D., LL.M.; *The Fondren Foundation Centennial Chair for Faculty Excellence.*
H. W. PERRY, JR., B.A., M.A., Ph.D.; *Associate Professor of Government and Law.*
LUCAS A. POWE, JR., B.A., J.D.; *Anne Green Regents Chair in Law, Professor of Government.*
WILLIAM C. POWERS, JR., B.A., J.D.; *President of The University of Texas at Austin, Hines H. Baker and Thelma Kelley Baker Chair in Law, University Distinguished Teaching Professor.*
DAVID M. RABBAN, B.A., J.D.; *Dahr Jamail, Randall Hage Jamail and Robert Lee Jamail Regents Chair, University Distinguished Teaching Professor.*
ALAN S. RAU, B.A., LL.B.; *Mark G. and Judy G. Yudof Chair in Law.*
DAVID W. ROBERTSON, B.A., LL.B., LL.M., J.S.D.; *W. Page Keeton Chair in Tort Law, University Distinguished Teaching Professor.*
JOHN A. ROBERTSON, A.B., J.D.; *Vinson & Elkins Chair.*
MARY ROSE, A.B., M.A., Ph.D.; *Associate Professor.*
WILLIAM M. SAGE, A.B., M.D., J.D.; *Vice Provost for Health Affairs of The University of Texas at Austin, James R. Dougherty Chair for Faculty Excellence.*
LAWRENCE G. SAGER, B.A., LL.B.; *Alice Jane Drysdale Sheffield Regents Chair.*
JOHN J. SAMPSON, B.B.A., LL.B.; *William Benjamin Wynne Professor.*
CHARLES M. SILVER, B.A., M.A., J.D.; *Roy W. and Eugenia C. McDonald Endowed Chair in Civil Procedure, Professor of Government, Co-Director of Center on Lawyers, Civil Justice and the Media.*
ERNEST E. SMITH, B.A., LL.B.; *Rex G. Baker Centennial Chair in Natural Resources Law.*
DAVID B. SPENCE, B.A., J.D., M.A., Ph.D.; *Professor of Business, Government and Society, and Law.*
JAMES C. SPINDLER, B.A., J.D., M.A., Ph.D.; *Sylvan Lang Professor of Law.*
JANE STAPLETON, B.S., Ph.D., LL.B., D.C.L., D. Phil.; *Ernest E. Smith Professor.*
JORDAN M. STEIKER, B.A., J.D.; *Judge Robert M. Parker Endowed Chair in Law, Co-Director of Capital Punishment Center.*
MICHAEL F. STURLEY, B.A., J.D.; *Fannie Coplin Regents Chair.*
GERALD TORRES, A.B., J.D., LL.M.; *Bryant Smith Chair in Law.*
JEFFREY K. TULIS, B.A., M.A., Ph.D.; *Associate Professor.*
GREGORY J. VINCENT, B.A., J.D., Ed.D.; *Professor, Vice President for Diversity and Community Engagement of The University of Texas at Austin.*
SRIRAM VISHWANATH, B.S., M.S., Ph.D.; *Associate Professor.*
WENDY E. WAGNER, B.A., M.E.S., J.D.; *Joe A. Worsham Centennial Professor.*
LOUISE WEINBERG, B.A., 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 of Law.*
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.*
SEAN H. WILLIAMS, B.A., J.D.; *Professor.*
ZIPPORAH B. WISEMAN, B.A., M.A., LL.B.; *Thos. H. Law Centennial Professor.*
PATRICK WOOLLEY, A.B., J.D.; *Beck, Redden & Secrest Professor in Law.*

ASSISTANT PROFESSORS

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

SENIOR LECTURERS, WRITING LECTURERS, AND CLINICAL PROFESSORS

ALEXANDRA W. ALBRIGHT, B.A., J.D.; *Senior Lecturer.*
WILLIAM P. ALLISON, B.A., J.D.; *Clinical Professor, Director of Criminal Defense Clinic.*
WILLIAM H. BEARDALL, JR., B.A., J.D.; *Adjunct 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.; *Adjunct Professor, Director of Legislative Lawyering Clinic.*
KAMELA S. BRIDGES, B.A., B.J., J.D.; *Lecturer.*
CYNTHIA L. BRYANT, B.A., J.D.; *Clinical Professor, Director of Mediation Clinic.*
JOHN C. BUTLER, B.B.A., Ph.D.; *Clinical Associate Professor.*
MARY R. CROUTER, A.B., J.D.; *Lecturer, Assistant Director of William Wayne Justice Center for Public Interest Law.*
MICHELE Y. DEITCH, B.A., M.S., J.D.; *Senior Lecturer.*
TIFFANY J. DOWLING, B.A., J.D.; *Clinical Instructor, Director of Actual Innocence Clinic.*

LORI K. DUKE, B.A., J.D.; *Clinical Professor.*
ARIEL E. DULITZKY, J.D., LL.M.; *Clinical Professor, Director of Human Rights Clinic.*
ELANA S. EINHORN, B.A., J.D.; *Lecturer.*
TINA V. FERNANDEZ, A.B., J.D.; *Lecturer, Director of Pro Bono Program.*
LYNDA E. FROST, B.A., M.Ed., J.D., Ph.D.; *Clinical Associate Professor.*

AHMED GHAPPOUR, B.S., J.D.; *Clinical Instructor.*
DENISE L. GILMAN, B.A., J.D.; *Clinical Professor, Co-Director of Immigration Clinic.*
KELLY L. HARAGAN, B.A., J.D.; *Lecturer, Director of Environmental Law Clinic.*
BARBARA HINES, B.A., J.D.; *Clinical Professor, Co-Director of Immigration Clinic.*
HARRISON KELLER, B.A., M.A., Ph.D.; *Senior Lecturer, Vice Provost for Higher Education Policy at the University of Texas at Austin.*
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.*

TRACY W. MCCORMACK, B.A., J.D.; *Senior Lecturer, Director of Advocacy Programs.*
F. SCOTT MCCOWN, B.S., J.D.; *Clinical Professor, Director of Children's Rights Clinic.*
ROBIN B. MEYER, B.A., M.A., J.D.; *Lecturer.*
RANJANA NATARAJAN, B.A., J.D.; *Clinical Professor, Director of Civil Rights Clinic.*
SEAN J. PETRIE, B.A., J.D.; *Lecturer.*
RACHAEL RAWLINS, B.A., M.R.P., J.D.; *Senior Lecturer.*

WAYNE SCHIESS, B.A., J.D.; *Senior Lecturer, Director of David J. Beck Center for Legal Research, Writing and Appellate Advocacy.*
RAOUL D. SCHONEMANN, B.A., LL.M., J.D.; *Clinical Professor.*
STACY ROGERS SHARP, B.S., J.D.; *Lecturer.*

PAMELA J. SIGMAN, B.A., J.D.; *Adjunct Professor, Director of Juvenile Justice Clinic.*
DAVID S. SOKOLOW, B.A., M.A., J.D., M.B.A.; *Distinguished Senior Lecturer, Director of Student Life.*
LESLIE L. STRAUCH, B.A., J.D.; *Clinical Professor.*

MELINDA E. TAYLOR, B.A., J.D.; *Senior Lecturer, Executive Director of Center for Global Energy, International Arbitration and Environmental Law.*
HEATHER K. WAY, B.A., B.J., J.D.; *Lecturer, Director of Entrepreneurship and Community Development Clinic.*
ELIZABETH M. YOUNGDALE, B.A., M.L.I.S., J.D.; *Lecturer.*

ADJUNCT PROFESSORS AND OTHER LECTURERS

ELIZABETH AEBERSOLD, B.A., M.S.
WILLIAM R. ALLENSWORTH, B.A., J.D.
ANDREW W. AUSTIN, B.A., M.Phil., J.D.
MARJORIE I. BACHMAN, B.S., J.D.
CRAIG D. BALL, B.A., J.D.
SHARON C. BAXTER, B.S., J.D.
KARL O. BAYER, B.A., M.S., J.D.
JERRY A. BELL, B.A., J.D.
ALLISON H. BENESCH, B.A., M.S.W., J.D.
CRAIG R. BENNETT, B.S., J.D.
JAMES B. BENNETT, B.B.A., J.D.
MURFF F. BLEDSOE, B.A., J.D.
WILLIAM P. BOWERS, B.B.A., J.D., LL.M.
STACY L. BRAININ, B.A., J.D.
ANTHONY W. BROWN, B.A., J.D.
JAMES E. BROWN, B.S., LL.B., J.D.
JEREMY M. BROWN, B.A., J.D.
TOMMY L. BROYLES, B.A., J.D.
PAUL J. BURKA, B.A., LL.B.
W. AMON BURTON, JR., B.A., M.A., LL.B.
ERIN G. BUSBY, B.A., J.D.
AGNES E. CASAS, B.A., J.D.
RUBEN V. CASTANEDA, B.A., J.D.
EDWARD A. CAVAZOS, B.A., J.D.
LINDA BRAY CHANOW, B.A., J.D.
JEFF CIVINS, A.B., M.S., J.D.
ELIZABETH COHEN, B.A., M.S.W., J.D.
JAMES W. COLLINS, B.S., J.D.
KEVIN D. COLLINS, B.A., J.D.
DANLEY K. CORNYN, B.A., J.D.
KASIA SOLON CRISTOBAL, B.A., M.S., J.D.
PATRICIA J. CUMMINGS, B.A., J.D.
KEITH B. DAVIS, B.S., J.D.
SCOTT D. DEATHERAGE, B.A., J.D.
DICK DEGUERIN, B.A., LL.B.
ADAM R. DELL, B.A., J.D.
RICHARD D. DEUTSCH, B.A., B.A., J.D.
STEVEN K. DEWOLF, B.A., J.D, LL.M.
REBECCA H. DIFFEN, B.A., J.D.
DENNIS B. DRAPKIN, A.B., B.E., LL.M., J.D.
CASEY D. DUNCAN, B.A., M.L.I.S., J.D.
PHILIP DURST, B.A., M.A., J.D.
JAY D. ELLWANGER, B.A., J.D.
LISA R. ESKOW, A.B., J.D.
EDWARD Z. FAIR, B.A., M.S.W., J.D.
SEAN P. FLAMMER, B.A., J.D.
JOHN C. FLEMING, B.A., J.D.
KYLE K. FOX, B.A., J.D.
DAVID C. FREDERICK, B.A., Ph.D., J.D.
GREGORY D. FREED, B.A., J.D.
FRED J. FUCHS, B.A., J.D.
RYAN M. GARCIA, B.G.S., J.D.
GRETTA G. GARDNER, B.A., J.D.
MICHAEL S. GOLDBERG, B.A., J.D.
MICHAEL J. GOLDEN, A.B., J.D.
DAVID M. GONZALES, B.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.
CHRISTOPHER S. HARRISON, Ph.D., J.D.
WILLIAM M. HART, B.A., J.D.
JOHN R. HAYS, JR., B.A., J.D.
SUSAN J. HIGHTOWER, B.A., M.A., J.D.
KENNETH E. HOUP, JR., B.A., J.D.
RANDY R. HOWRY, B.J., J.D.
MONTY G. HUMBLE, B.A., J.D.
DIRK M. JORDAN, B.A., J.D.
JEFFREY R. JURY, B.A., J.D.
PATRICK O. KEEL, B.A., J.D.
DOUGLAS L. KEENE, B.A., M.Ed., Ph.D.
CHARI L. KELLY, B.A., J.D.
ROBERT N. KEPPLE, B.A., J.D.
MARK L. KINCAID, B.B.A., J.D.
ALICE L. KING, B.A., J.D.
MARGARET A. KIRKENDALL, B.A., J.D.
MICHAEL R. KRAWZSENEK, B.S., J.D.
AMI L. LARSON, B.A., J.D.
LARRY LAUDAN, B.A., M.A., Ph.D.
JODI R. LAZAR, B.A., J.D.
KEVIN L. LEAHY, B.A., J.D.
DAVID P. LEIN, B.A., M.P.A., J.D.
ANDRES J. LINETZKY, J.D., LL.M.

JAMES-LLOYD LOFTIS, B.B.A., J.D.
JIM MARCUS, B.A., J.D.
HARRY S. MARTIN, A.B., M.L.S., J.D.
FRANCES L. MARTINEZ, B.A., J.D.
LAURA A. MARTINEZ, B.A., J.D.
PHILIP K. MAXWELL, B.A., J.D.
PETER C. MCCABE, B.A., J.D.
ANN M. MCGEEHAN, B.A., J.D.
BARRY F. MCNEIL, B.A., J.D.
MARGARET M. MENICUCCI, B.A., J.D.
JO ANN MERICA, B.A., J.D.
RANELLE M. MERONEY, B.A., J.D.
ELIZABETH N. MILLER, B.A., J.D.
JONATHAN F. MITCHELL, B.A., J.D.
DARYL L. MOORE, B.A., M.L.A., J.D.
EDWIN G. MORRIS, B.S., J.D.
SARAH J. MUNSON, B.A., J.D.
MANUEL H. NEWBURGER, B.A., J.D.
DAVID G. NIX, B.S.E., LL.M., J.D.
JOSEPH W. NOEL, B.S.E., J.D., M.S.L.S.
JANE A. O'CONNELL, B.A., M.S., J.D.
PATRICK L. O'DANIEL, B.B.A., J.D.
M. ARIEL PAYAN, B.A., J.D.
MARK L. PERLMUTTER, B.S., J.D.
ELIZA T. PLATTS-MILLS, B.A., J.D.
WILLIAM C. POLLARD, B.A., J.D.
JONATHAN PRATTER, B.A., M.S.L.I.S., J.D.
VELVA L. PRICE, B.A., J.D.
BRIAN C. RIDER, B.A., J.D.
ROBERT M. ROACH, JR., B.A., J.D.
BRIAN J. ROARK, B.A., J.D.
BETTY E. RODRIGUEZ, B.S.W., J.D.
JAMES D. ROWE, B.A., J.D.
MATTHEW C. RYAN, B.A., J.D.
KAREN R. SAGE, B.A., J.D.
MARK A. SANTOS, B.A., J.D.
JAMES J. SCHESKE, B.A., J.D.
MICHAEL J. SCHLESS, B.A., J.D.
SUSAN SCHULTZ, B.S., J.D.
AMY J. SCHUMACHER, B.A., J.D.
SUZANNE SCHWARTZ, B.J., J.D.
RICHARD J. SEGURA, JR., B.A., J.D.
DAVID A. SHEPPARD, B.A., J.D.
HON. ERIC M. SHEPPERD, B.A., J.D.
A. HAAG SHERMAN, B.B.A., J.D., C.P.A.
RONALD J. SIEVERT, B.A., J.D.
STUART R. SINGER, A.B., J.D.
HON. BEA A. SMITH, B.A., M.A., J.D.
LYDIA N. SOLIZ, B.B.A., J.D.
JAMES. M. SPELLINGS, JR., B.S., J.D.
KACIE L. STARR, B.A., J.D.
MATTHEW R. STEINKE, B.A., M.L.I.S., J.D.
WILLIAM F. STUTTS, B.A., J.D.
MATTHEW J. SULLIVAN, B.S., J.D.
GRETCHEN S. SWEEN, B.A., M.A., Ph.D., J.D.
JEREMY S. SYLESTINE, B.A., J.D.
BRADLEY P. TEMPLE, B.A., J.D.
SHERINE E. THOMAS, B.A., J.D.
MICHAEL J. TOMSU, B.A., M.B.A., J.D.
TERRY O. TOTTENHAM, B.S., LL.M., J.D.
MICHAEL S. TRUESDALE, B.A., M.A., J.D.
TIMOTHY J. TYLER, B.A., J.D.
SUSAN S. VANCE, B.B.A., J.D.
LANA K. VARNEY, B.A., J.D.
DEBORAH M. WAGNER, B.A., M.Arch., J.D.
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.
MARK B. WILSON, B.A., M.A., J.D.
CHRISTINA T. WISDOM, B.A., J.D.
STEPHEN M. WOLFSON, B.A., M.S., J.D.
HON. PAUL L. WOMACK, B.S., J.D.
LUCILLE D. WOOD, B.A., J.D.
DENNEY L. WRIGHT, B.B.A., J.D., LL.M.
DANIEL J. YOUNG, B.A., J.D.

VISITING PROFESSORS

OWEN L. ANDERSON, B.A., J.D.
ANTONIO H. BENJAMIN, LL.B., LL.M.
PA DE LARRANAGA MONJARAZ, LL.B., LL.M., Ph.D.
VICTOR FERRERES, J.D., LL.M., J.S.D.
EMILY C. HAMMOND, B.S., J.D.
ANDREW KULL, B.A., B.A., M.A., J.D.
GRAHAM B. STRONG, B.A., J.D., LL.M.

* * *

Texas Law Review

Volume 92 Number 4 March 2014

* * *

Texas Law Review

Volume 92, Number 4, March 2014

ARTICLES

BOOK REVIEWS

NOTES

Articles

Treaty Termination and Historical Gloss

Curtis A. Bradley*

The termination of U.S. treaties provides an especially rich example of how governmental practices can provide a "gloss" on the Constitution's separation of powers. The authority to terminate treaties is not addressed specifically in the constitutional text and instead has been worked out over time through political-branch practice. This practice, moreover, has developed largely without judicial review. Despite these features, Congress and the President—and the lawyers who advise them—have generally treated this issue as a matter of constitutional law rather than merely political happenstance. Importantly, the example of treaty termination illustrates not only how historical practice can inform constitutional understandings but also how these understandings can change. Whereas it was generally understood throughout the nineteenth century that the termination of treaties required congressional involvement, the consensus on this issue disappeared in the early parts of the twentieth century, and today it is widely (although not uniformly) accepted that presidents have a unilateral power of treaty termination. This shift in constitutional understandings did not occur overnight or in response to one particular episode but rather was the product of a long accretion of Executive Branch claims and practice in the face of congressional inaction. An examination of the way in which historical practice has shaped the constitutional debates and understandings concerning this issue can help shed light on some of the interpretive and normative challenges associated with a practice-based approach to the separation of powers.

---

* William Van Alstyne Professor, Duke Law School. For helpful comments and suggestions, I thank Joseph Blocher, Jamie Boyle, Kathy Bradley, John Coyle, Bill Dodge, Martin Flaherty, Jean Galbraith, Jack Goldsmith, Oona Hathaway, Larry Helfer, Duncan Hollis, Suzanne Katzenstein, David Moore, Eric Posner, Jeff Powell, Barak Richman, Stephen Sachs, Chris Schroeder, Neil Siegel, Peter Spiro, Ingrid Wuerth, Ernie Young, and participants in a faculty workshop at Duke Law School and in the American Society of International Law's annual International Law in Domestic Courts interest group held at Yale Law School. For excellent research assistance, I thank Kristina Alayan, Amy Feagles, and Chad Turner.

Introduction

Historical practice is frequently invoked in debates and decisions concerning the Constitution's distribution of authority between Congress and the President. On issues ranging from the President's authority to make recess appointments, to the role of Congress in authorizing military operations, to the validity of "executive agreements" with foreign nations, the way in which the government has operated over time is invoked as evidence of constitutional meaning.[1] Such governmental practice is sometimes referred to as "historical gloss," after Justice Frankfurter's contention in the *Youngstown*[2] steel-seizure case that "a systematic, unbroken, executive practice, long pursued to the knowledge of the Congress and never before questioned . . . may be treated as a gloss on 'executive Power' vested in the President by § 1 of Art. II."[3]

This Article presents a detailed case study of historical gloss, focused on presidential authority to terminate treaties. Treaty termination is an especially rich example of how governmental practices can inform and even define the Constitution's separation of powers. The authority to terminate treaties is not addressed specifically in the constitutional text and instead has been worked out over time through political-branch practice. This practice, moreover, has developed largely without judicial review. Despite these features, Congress and the President—and the lawyers who advise them—have generally treated this issue as a matter of constitutional law,

1. *See generally* Curtis A. Bradley & Trevor W. Morrison, *Historical Gloss and the Separation of Powers*, 126 HARV. L. REV. 411 (2012) (considering the role of historical practice in debates and decisions relating to the separation of powers); Michael J. Glennon, *The Use of Custom in Resolving Separation of Powers Disputes*, 64 B.U. L. REV. 109 (1984) (same).

2. Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579 (1952).

3. *Id.* at 610–11 (Frankfurter, J., concurring).

not merely political happenstance. Legal scholars, too, have long discussed and debated the issue in legal terms. At the same time, there has been a recognition that the constitutional law in this area is not entirely distinct from politics, and that it both is informed by and shapes political contestation.

The example of treaty termination illustrates not only how a constitutional gloss on governmental authority can develop but also how it can change. As will be seen, the center of gravity of the debate over treaty termination has shifted substantially over time, from whether the full Congress or merely the Senate needs to *approve* a termination to whether Congress or the Senate can even *limit* the President's unilateral authority to terminate. One can identify a pattern of change, the contours of which may apply to other issues of constitutional law relating to presidential authority: First there is a consensus, both among the governmental actors and in the scholarly community. Then deviations take place with a potentially limited scope. The Executive Branch proceeds to articulate broader theories of the deviations. Congress's resistance is intermittent, depending on whether it objects to the deviations on policy grounds. Practice then builds up around low-stakes examples. Eventually a more controversial example arises and the President pushes forward successfully, thereby consolidating the changed understanding.

In developing the case study, this Article makes three contributions. First, it presents the most complete and accurate account to date of the historical practice of U.S. treaty terminations. In addition to reviewing various publicly available materials, such as congressional hearings and presidential proclamations, this Article considers a number of internal legal memoranda obtained from the State Department archives. Second, this Article recovers a nineteenth-century understanding of treaty-termination authority that has largely been lost from modern considerations of the issue, pursuant to which the termination of treaties, like the making of treaties, was generally understood by both Congress and the President as a shared power. Most modern accounts acknowledge vaguely that treaty terminations have been accomplished in a variety of ways throughout U.S. history but fail to appreciate the sharp contrast between the modern presidential unilateralism and the nineteenth-century practices and understandings. In endorsing a unilateral presidential power to terminate treaties, for example, the American Law Institute's *Restatement (Third) of the Foreign Relations Law* notes in passing that "[p]ractice has varied" without acknowledging that presidential unilateralism is almost entirely a twentieth-century development.[4] Third, this Article uses this historical

4. *See* RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 339 reporters' note 2 (1987) ("Practice has varied, the President sometimes terminating an

record as a window into the nature of a practice-based approach to constitutional interpretation and some of its limitations and challenges.

Part I of this Article provides the legal and theoretical background needed to understand and assess the historical practice of U.S. treaty terminations. It describes both the allowable grounds under international law for terminating a treaty, as well as the textual and structural arguments relating to the Constitution's assignment of treaty termination authority. It also considers some of the reasons why historical practice has played a significant role in constitutional debates surrounding this issue. Parts II and III review the practice of treaty termination throughout U.S. history. Part II shows that, at least until the late nineteenth century, it was generally understood that presidents needed the agreement of Congress or the Senate in order to terminate a treaty. Part III recounts how this understanding changed in the twentieth century, a process that occurred over the course of decades as a result of repeated claims and actions by the Executive Branch in the face of congressional inaction. Part IV assesses the implications of the case study, both with respect to the specific question of treaty-termination authority as well as the more general issue of the proper role of historical practice in the separation of powers area. It concludes by reflecting on the relationship between law and politics for practice-based norms of institutional authority.

I. Legal and Theoretical Background

This Part provides the legal and theoretical background needed to understand and assess the historical practice of U.S. treaty terminations. It begins by explaining the circumstances under which international law allows a nation to terminate a treaty. It then considers the textual and structural considerations that are relevant to determining which actors in the United States have the constitutional authority to terminate treaties. Finally, it describes why historical practice plays an especially important role in constitutional debates concerning this issue.

A. International-Law Standards

Treaties are binding on nations as a matter of international law. Ultimately, therefore, whether a nation's treaty commitments are terminated is determined by international law, not U.S. law.[5] As a result, before considering the U.S. constitutional issues, it is important to understand first what international law provides about treaty termination. The modern rules

agreement on his own authority, sometimes doing so when requested by Congress or by the Senate alone.").

5. *See generally* Laurence R. Helfer, *Terminating Treaties*, *in* THE OXFORD GUIDE TO TREATIES 634 (Duncan B. Hollis ed., 2012) (describing the international-law standards governing treaty termination).

on this subject are set forth in the Vienna Convention on the Law of Treaties,[6] which took effect in 1980 and has now been ratified by over 110 nations.[7] Although the United States is not a party to the Convention, Executive Branch officials have stated at various times that they regard the Convention as largely reflective of binding rules of international custom,[8] and U.S. courts also regularly refer to the Convention.[9] In addition, the International Court of Justice has specifically observed that "in many respects" the Vienna Convention's provisions on the suspension or termination of treaty provisions reflect binding custom.[10]

Under the Convention, there are a variety of circumstances that can render a party's consent to a treaty invalid. Some of these circumstances merely make the treaty voidable at the party's discretion. For example, "[i]f a State has been induced to conclude a treaty by the fraudulent conduct of another negotiating State, the State may invoke the fraud as invalidating its consent to be bound by the treaty."[11] Other circumstances automatically void the treaty. For example, "[a] treaty is void if its conclusion has been procured by the threat or use of force in violation of the principles of international law embodied in the Charter of the United Nations."[12] When a treaty is deemed void, it will be considered never to have created obligations.[13]

Whereas some circumstances will allow a party to void even its past treaty obligations, other circumstances will allow it to terminate or suspend its treaty obligations going forward. For example, a party may suspend or terminate its obligations under a bilateral treaty if the other treaty party has materially breached the treaty.[14] In addition, "[a] party may invoke the impossibility of performing a treaty as a ground for terminating or

---

6. Vienna Convention on the Law of Treaties, May 23, 1969, 1155 U.N.T.S. 331 [hereinafter VCLT].

7. Chapter XXIII of Multilateral Treaties Deposited with the Secretary-General, UNITED NATIONS TREATY COLLECTION (last updated Feb. 3, 2014), https://treaties.un.org/Pages/ViewDetailsIII.aspx?&src=TREATY&mtdsg_no=XXIII~1&chapter=23&Temp=mtdsg3&lang=en.

8. *See* RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES pt. 3, intro. note (1987) (documenting Executive Branch statements); *see also* Maria Frankowska, *The Vienna Convention on the Law of Treaties Before United States Courts*, 28 VA. J. INT'L L. 281, 286 (1988) ("[A]ccording to a widespread *opinio juris*, legal conviction of the international community, the Vienna Convention represents a treaty which to a large degree is a restatement of customary rules . . . .").

9. *See, e.g.*, Mora v. New York, 524 F.3d 183, 196 n.19 (2d Cir. 2008) ("Although the United States has not ratified the Vienna Convention on the Law of Treaties, our Court relies on it 'as an authoritative guide to the customary international law of treaties,' insofar as it reflects actual state practices.").

10. Gabčíkovo–Nagymaros Project (Hung. v. Slovk.), 1997 I.C.J. 7, para. 46 (Sept. 25).

11. VCLT, *supra* note 6, art. 49.

12. *Id.* art. 52.

13. *See id.* art. 69, para. 1 ("The provisions of a void treaty have no legal force.").

14. *Id.* art. 60, para. 1.

withdrawing from it if the impossibility results from the permanent disappearance or destruction of an object indispensable for the execution of the treaty."[15] Furthermore, under narrow circumstances, a party may invoke a fundamental change of circumstances as a basis for suspending or terminating a treaty.[16] Treaty obligations can also be suspended or terminated if the parties expressly agree to such suspension or termination or act to conclude a new superseding treaty, or if the treaty expressly provides for suspension or termination after a certain period of time or in response to certain events.[17]

Finally, nations may also withdraw from (or "denounce") a treaty that expressly provides for a right of withdrawal.[18] Such withdrawal clauses are common in modern treaties and often include a required notice period before the termination will take effect.[19] In some instances, a right of withdrawal will be implied. The Vienna Convention states that:

> A treaty which contains no provision regarding its termination and which does not provide for denunciation or withdrawal is not subject to denunciation or withdrawal unless: (*a*) it is established that the parties intended to admit the possibility of denunciation or withdrawal; or (*b*) a right of denunciation or withdrawal may be implied by the nature of the treaty.[20]

When there is an implied right of withdrawal, the Vienna Convention states that the party seeking to withdraw from the treaty shall give at least twelve months' notice.[21] A nation that withdraws from a treaty is bound by any obligations that arose before the effective date of the withdrawal.[22]

The rules of treaty termination that existed at the time of the constitutional founding were less developed and incorporated distinctions that are no longer relevant, such as a distinction between treaties that obligated only the particular monarchs making them and treaties that obligated their nations in perpetuity.[23] Nevertheless, these rules encompassed certain grounds for terminating a treaty that we would recognize today, such as a material breach by the other party.[24] It is worth

---

15. *Id.* art. 61, para. 1.

16. *Id.* art. 62, para. 1.

17. *Id.* arts. 54, 57, 59.

18. The terms "denunciation" and "withdrawal" are often used interchangeably to refer to a voluntary act of treaty termination. Helfer, *supra* note 5, at 635.

19. *See* Laurence R. Helfer, *Exiting Treaties*, 91 VA. L. REV. 1579, 1581–82, 1597–98 (2005).

20. VCLT, *supra* note 6, art. 56, para. 1.

21. *Id.* art. 56, para. 2.

22. *Id.* art. 70, para. 1(b).

23. *See, e.g.*, 1 E. DE VATTEL, THE LAW OF NATIONS OR THE PRINCIPLES OF NATURAL LAW bk. 2, §§ 187–197 (Charles G. Fenwick trans., 1916) (1758).

24. *See, e.g.*, JEAN-JACQUES BURLAMAQUI, THE PRINCIPLES OF NATURAL AND POLITIC LAW 524 (Petter Korkman ed., Thomas Nugent trans., Liberty Fund, Inc. 2006) (1763); 2 HUGO GROTIUS, DE JURE BELLI AC PACIS LIBRI TRES 405 (Francis W. Kelsey trans., Clarendon Press

noting, however, that although clauses in treaties allowing for unilateral withdrawal are now common, they were not common at the time of the founding. Indeed, it appears that the United States did not become a party to a treaty containing a unilateral withdrawal clause until 1822.[25]

What international law did not address then, and still does not address, is how treaty termination decisions are to be made internally by each nation. For the United States, that internal issue is a matter of U.S. constitutional law.

B. Textual and Structural Considerations

Article II of the Constitution sets forth the process by which the United States is to conclude treaties. It provides that the President "shall have Power, by and with the Advice and Consent of the Senate, to make Treaties, provided two thirds of the Senators present concur."[26] The vast majority of international agreements concluded by the United States in the modern era do not go through this process and are instead concluded as "congressional–executive agreements" (approved before or after the fact by a majority of Congress) or "sole executive agreements" (approved solely by the President).[27] Nevertheless, some of the United States' most significant agreements are still concluded as Article II treaties. To take just a few examples, the United Nations Charter, the Geneva Conventions, and the International Covenant on Civil and Political Rights were all concluded

---

1925) (1646); 2 SAMUEL PUFENDORF, DE JURE NATURAE ET GENTIUM LIBRI OCTO 1339–40 (C.H. Oldfather & W.A. Oldfather trans., Clarendon Press 1934) (1688); VATTEL, *supra* note 23, § 202; *see also* Ware v. Hylton, 3 U.S. (3 Dall.) 199, 261 (1796) (Iredell, J.) ("It is a part of the law of nations, that if a treaty be violated by one party, it is at the option of the other party, if innocent, to declare, in consequence of the breach, that the treaty is void.").

25. The treaty concerned the imposition of customs duties, and it provided that it was to remain in force for two years "and even after the expiration of that term, until the conclusion of a definitive treaty, or until one of the parties shall have declared its intention to renounce it; which declaration shall be made at least six months before hand." Convention of Navigation and Commerce, U.S.–Fr., art. 7, June 24, 1822, 8 Stat. 278; *see also* Memorandum from William Whittington, Termination of Treaties: International Rules and Internal United States Procedure 3 (Feb. 10, 1958) [hereinafter Whittington Memorandum] (on file with author) (noting that the 1822 treaty was the first treaty concluded by the United States containing a unilateral withdrawal clause). The withdrawal clause was included at the request of France, which explained that, "As their object is to make an experiment, it should be so established as not to press too heavily upon whichever of the two parties may, on experience, be found to have erred in the calculation." Letter from G. Hyde de Neuville, Envoy Extraordinary, to John Quincy Adams, Sec'y of State (May 15, 1822), *in* 5 AMERICAN STATE PAPERS: DOCUMENTS, LEGISLATIVE AND EXECUTIVE, OF THE CONGRESS OF THE UNITED STATES 210, 211 (Asbury Dickins & James C. Allen eds., Wash., Gales & Seaton 1858).

26. U.S. CONST. art. II, § 2.

27. *See* CURTIS A. BRADLEY, INTERNATIONAL LAW IN THE U.S. LEGAL SYSTEM 74–75 (2013); Oona A. Hathaway, *Treaties' End: The Past, Present, and Future of International Lawmaking in the United States*, 117 YALE L.J. 1236, 1287–88 (2008). The constitutional issues implicated by the termination of executive agreements are potentially distinct from those implicated by the termination of Article II treaties, and they are not considered here.

through the senatorial advice and consent process. To the extent that the United States ever becomes a party to treaties such as the Law of the Sea Convention, the Rome Statute of the International Criminal Court, or the Convention on the Elimination of All Forms of Discrimination Against Women, it is expected that it will do so pursuant to the Article II process.[28] The Constitution does not specifically address, however, the way in which the United States is to go about terminating treaty commitments.

Some proponents of unilateral presidential authority to terminate treaties rely on what has been referred to as the "Vesting Clause Thesis." According to this thesis, the first sentence of Article II of the Constitution, which provides that "[t]he executive Power shall be vested in a President of the United States of America,"[29] conveys to the President all authority that is "executive" in nature, regardless of whether that authority is specifically mentioned in the remainder of Article II, unless the Constitution specifically conveys that authority to another institutional actor.[30] This thesis supports a unilateral presidential authority to terminate treaties, it is argued, because the termination of treaties is executive in nature and is not specifically assigned to an actor other than the President.[31] The Vesting Clause Thesis, however, is highly controversial.[32] Moreover, supporters of this thesis vary in what authority they contend is conveyed by the clause.[33]

A variety of structural considerations are also potentially relevant to determining who has the treaty termination power in the United States, but these considerations do not point in a single direction. On the one hand, the

---

28. To the extent that presidents have proposed moving ahead with ratification of these treaties, they have always suggested that the process would be the one set forth in Article II. For the treaty establishing the International Criminal Court, Congress has specifically mandated that it can be ratified by the United States only through the Article II process. 22 U.S.C. § 7401(a) (2012).

29. U.S. CONST. art. II, § 1.

30. For arguments in support of the thesis, see, for example, Saikrishna B. Prakash & Michael D. Ramsey, *The Executive Power over Foreign Affairs*, 111 YALE L.J. 231 (2001), and John C. Yoo, *War and the Constitutional Text*, 69 U. CHI. L. REV. 1639, 1677–78 (2002). *See also* MICHAEL D. RAMSEY, THE CONSTITUTION'S TEXT IN FOREIGN AFFAIRS 158 (2007) (applying this thesis to the issue of treaty termination).

31. *See* RAMSEY, *supra* note 30.

32. For criticism of the thesis, see, for example, Curtis A. Bradley & Martin S. Flaherty, *Executive Power Essentialism and Foreign Affairs*, 102 MICH. L. REV. 545, 551 (2004), and Robert J. Reinstein, *The Limits of Executive Power*, 59 AM. U. L. REV. 259, 263–64 (2009). *See also* Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 640–41 (1952) (Jackson, J., concurring) ("If [the Vesting Clause Thesis] be true, it is difficult to see why the forefathers bothered to add several specific items [in Article II], including some trifling ones.").

33. For example, unlike Professors Prakash, Ramsey, and Yoo, Steven Calabresi and Kevin Rhodes contend simply that "the Clause grants the President the power to supervise and control all subordinate executive officials exercising executive power conferred explicitly by either the Constitution or a valid statute," and they do not make "the more ambitious (and far more doubtful) claim" that it conveys substantive authority. *See* Steven G. Calabresi & Kevin H. Rhodes, *The Structural Constitution: Unitary Executive, Plural Judiciary*, 105 HARV. L. REV. 1153, 1177 n.119 (1992).

making of treaties might be viewed as analogous to the appointment of Executive Branch officers. Even though such appointment requires, as for treaties, senatorial advice and consent (albeit a majority of the Senate rather than two-thirds), it is well settled that presidents have some unilateral authority to remove executive officers.[34] On the other hand, the making of treaties could be viewed as analogous to the making of federal statutes, since both are part of the supreme law of the land. It is well settled that the same process that applies to the making of federal statutes (approval by a majority of both houses of Congress and presidential signature, or a supermajority congressional override of a presidential veto) also must be followed for the termination of federal statutes.[35]

Another structural consideration concerns Congress's well-accepted authority to override the domestic effect of a treaty by enacting a later-in-time inconsistent statute.[36] If that is all that Congress does, the international-law status of the treaty will continue, and the United States may end up in breach of its international obligations.[37] The fact that Congress has the authority to terminate the domestic effect of a treaty might suggest that it also can have a role in terminating the treaty's international-law effect, but the second power does not necessarily follow from the first. Conversely, even if the President has the unilateral authority to terminate a treaty internationally, it would not necessarily mean that he could (like Congress) terminate its domestic effect without having validly terminated its international-law effect. In fact, if treaties are part of the "Laws" that the President is obligated under Article II of the Constitution to take care to

34. *See, e.g.*, Free Enter. Fund v. Pub. Co. Accounting Oversight Bd., 130 S. Ct. 3138, 3146 (2010) ("Since 1789, the Constitution has been understood to empower the President to keep [executive] officers accountable—by removing them from office, if necessary.").

35. *See, e.g.*, Clinton v. City of New York, 524 U.S. 417, 438 (1998) ("There is no provision in the Constitution that authorizes the President to enact, to amend, or to repeal statutes."); INS v. Chadha, 462 U.S. 919, 954 (1983) ("[R]epeal of statutes, no less than enactment, must conform with Art. I.").

36. *See, e.g.*, Breard v. Greene, 523 U.S. 371, 376 (1998) (per curiam); The Chinese Exclusion Case, 130 U.S. 581, 600 (1889); Whitney v. Robertson, 124 U.S. 190, 193–95 (1888); *see also* La Abra Silver Mining Co. v. United States, 175 U.S. 423, 460 (1899) ("It has been adjudged that Congress by legislation, and so far as the people and authorities of the United States are concerned, could abrogate a treaty made between this country and another country which had been negotiated by the President and approved by the Senate.").

37. *See, e.g.*, Pigeon River Improvement, Slide & Boom Co. v. Charles W. Cox, Ltd., 291 U.S. 138, 160 (1934) (noting that although a federal statute that conflicted with a treaty provision "would control in our courts as the later expression of our municipal law, . . . the international obligation [would] remain[] unaffected"). Courts will attempt to construe statutes, however, to avoid a treaty violation if possible. *See, e.g.*, Cook v. United States, 288 U.S. 102, 120 (1933) ("A treaty will not be deemed to have been abrogated or modified by a later statute unless such purpose on the part of Congress has been clearly expressed."); *see also* BRADLEY, *supra* note 27, at 54–55 (explaining reasons for this canon of construction).

faithfully execute (as a number of scholars have concluded),[38] it may be constitutionally impermissible for the President to override the domestic effect of a treaty that is otherwise still in force.

Still another structural consideration is the role of the Executive Branch in communicating with foreign nations. The President has often been described as the "sole organ" of formal communications between the United States and the rest of the world, a role that is arguably implied from both the unitary nature of the Executive Branch as well as the President's constitutional authority to make treaties and appoint and receive ambassadors.[39] To be sure, the phrase "sole organ" is an overstatement, given that Congress often takes positions on matters of foreign policy and that members of Congress regularly interact with foreign officials.[40] But it has always been the case that formal diplomatic functions are handled by the Executive Branch.[41] Because a termination of a treaty needs to be communicated to the other treaty parties, the "sole organ" role of the President may mean that neither Congress nor the Senate can effectuate by themselves a treaty termination. This would not necessarily establish, of course, that the President has unilateral authority to terminate a treaty. After all, it is understood that no treaty can be ratified except through presidential action,[42] and yet the President is required to obtain the advice and consent of two-thirds of the Senate before engaging in such ratification. The President's "sole organ" authority might mean, however, that Congress cannot validly require the President to terminate a treaty.

---

38. *E.g.*, Derek Jinks & David Sloss, *Is the President Bound By the Geneva Conventions?*, 90 CORNELL L. REV. 97, 157–58 (2004); Michael D. Ramsey, *Torturing Executive Power*, 93 GEO. L.J. 1213, 1231–32 (2005); *see also* U.S. CONST. art. II, § 3 (stating that the President "shall take Care that the Laws be faithfully executed"). It is not clear whether treaties must be "self-executing" in order to qualify as "Laws" for this purpose.

39. *See, e.g.*, United States v. Curtiss-Wright Exp. Corp., 299 U.S. 304, 320 (1936) (referring to "the very delicate, plenary and exclusive power of the President as the sole organ of the federal government in the field of international relations"); *see also* 10 ANNALS OF CONG. 613 (1800) (describing a statement by John Marshall, made when serving as a Representative in Congress, that "[t]he President is the sole organ of the nation in its external relations, and its sole representative with foreign nations"); Alexander Hamilton, *Pacificus No. 1* (June 29, 1793), *reprinted in* 15 THE PAPERS OF ALEXANDER HAMILTON 33, 38 (Harold C. Syrett et al. eds., 1969) (describing the Executive Branch as "the *organ* of intercourse between the Nation and foreign Nations"); Letter from Thomas Jefferson to Edmond Charles Genet (Nov. 22, 1793), *in* 27 THE PAPERS OF THOMAS JEFFERSON 414, 414 (John Catanzariti et al. eds., 1997) (stating that the President is the "only channel of communication" between the United States and foreign nations).

40. *See, e.g.*, Ryan M. Scoville, *Legislative Diplomacy*, 112 MICH. L. REV. 331 (2013) (describing various forms of interactions between legislators and foreign nations and officials).

41. *See* LOUIS HENKIN, FOREIGN AFFAIRS AND THE UNITED STATES CONSTITUTION 41–45 (2d ed. 1996) (describing the longstanding view of the Executive Branch as having exclusive power to conduct diplomacy).

42. *See* CONG. RESEARCH SERV., 106TH CONG., TREATIES AND OTHER INTERNATIONAL AGREEMENTS: THE ROLE OF THE UNITED STATES SENATE 152 (Comm. Print 2001) [hereinafter CRS STUDY] (stating that a failure to ratify a treaty on the part of the President means that the treaty "cannot enter into force for the United States").

Finally, it is conceivable that the President might have some constitutional authority to *suspend* treaty obligations even if he or she did not have constitutional authority to *terminate* the obligations. As noted, the President has the obligation and authority to "take Care that the Laws be faithfully executed."[43] As part of his Take Care Clause responsibilities, the President necessarily makes judgments (at least within certain limits) about the levels of enforcement of federal law.[44] It is arguable that this authority encompasses the ability to direct a temporary suspension of U.S. compliance with a treaty while a dispute concerning the treaty is addressed.

C. *Importance of Historical Practice*

The historical practice of U.S. treaty termination is described in detail in Parts II and III. As will be seen, when there has been debate over how treaties can constitutionally be terminated, such as in Congress or the courts, the debate has often focused on historical practice.[45] Moreover, Executive Branch lawyers have focused heavily on historical practice in advising presidents and secretaries of state about their constitutional authority concerning treaty termination.[46] Scholars, too, have long accorded historical practice a prominent place in the legal analysis of this issue.[47]

Consider two modern controversies. In the 1970s, there was extensive debate over the issue of treaty termination in the wake of President Carter's announcement that he was terminating a mutual defense treaty with Taiwan in conjunction with his decision to recognize the People's Republic of China.[48] The congressional hearings, scholarly commentary, and judicial decisions relating to that controversy were all heavily focused on historical practice.[49] So was the Executive Branch's reasoning: In a memorandum to

---

43. U.S. CONST. art. II, § 3.

44. *See, e.g.*, Arizona v. United States, 132 S. Ct. 2492, 2499 (2012) ("The dynamic nature of relations with other countries requires the Executive Branch to ensure that enforcement policies [under the immigration laws] are consistent with this Nation's foreign policy with respect to these and other realities."); Heckler v. Chaney, 470 U.S. 821, 831–32 (1985) ("An agency generally cannot act against each technical violation of the statute it is charged with enforcing. The agency is far better equipped than the courts to deal with the many variables involved in the proper ordering of its priorities."); Wayte v. United States, 470 U.S. 598, 607 (1985) ("In our criminal justice system, the Government retains 'broad discretion' as to whom to prosecute."). There are presumably limits on this enforcement discretion. *See, e.g.*, Zachary S. Price, *Enforcement Discretion and Executive Duty*, 67 VAND. L. REV. (forthcoming 2014) (considering the scope of the Executive Branch's enforcement discretion).

45. *See infra* Parts II–III.

46. *See, e.g.*, *infra* subpart III(C) (describing how the State Department Legal Adviser relied heavily on historical practice when advising the President that he had authority to terminate a treaty with Taiwan).

47. *See infra* subparts II(E), III(E).

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

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

the Secretary of State advising him that the President had the constitutional authority to terminate the Taiwan treaty, for example, the Legal Adviser to the State Department cited twelve purported instances in which presidents had terminated treaties unilaterally and attached an appendix describing the "History of Treaty Termination by the United States."[50]

More recently, there was controversy in 2002 over President George W. Bush's announcement that he was unilaterally withdrawing the United States from the Anti-Ballistic Missile (ABM) Treaty with Russia.[51] Again, the Executive Branch relied extensively on historical practice. In concluding that President Bush had the unilateral authority to suspend or terminate the ABM Treaty, the Justice Department's Office of Legal Counsel argued that "[t]he executive branch has long held the view that the President has the constitutional authority to terminate treaties unilaterally, and the legislative branch seems for the most part to have acquiesced in it."[52]

There are a number of reasons why historical practice has played such a prominent role in discussions of this issue. As a general matter, arguments based on historical practice are a common feature of debates and decisions relating to the constitutional separation of powers.[53] This is especially true in debates and decisions relating to the scope of presidential power. Unlike the extensive list of powers granted to Congress, the text of the Constitution says relatively little about the scope of presidential authority.[54] Responding in part to this limited textual guidance, Justice Frankfurter emphasized the importance of historical practice to the interpretation of presidential power in his concurrence in the *Youngstown* steel seizure case. In his view, "[i]t is an inadmissibly narrow conception of American constitutional law to confine it to the words of the Constitution and to disregard the gloss which life has written upon them."[55]

---

50. *See infra* notes 223–24 and accompanying text.

51. *See infra* notes 252–58 and accompanying text.

52. Memorandum from John C. Yoo, Deputy Assistant Att'y Gen. & Robert J. Delahunty, Special Counsel, Office of Legal Counsel, U.S. Dep't of Justice, to John Bellinger, III, Senior Assoc. Counsel to the President & Legal Adviser to the Nat'l Sec. Council, Authority of the President to Suspend Certain Provisions of the ABM Treaty 15–16 (Nov. 15, 2001) [hereinafter Yoo & Delahunty Memorandum], *available at* http://www.justice.gov/olc/docs/memoabmtreaty11152001.pdf.

53. Bradley & Morrison, *supra* note 1, at 417–24. Practice-based arguments are also common in other areas of constitutional law, such as federalism. Invocations of practice in those areas raise issues that are potentially distinct from the issues considered here. *See id.* at 416–17.

54. *Id.* at 417–18.

55. Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 610 (1952) (Frankfurter, J., concurring); *see also, e.g.*, Mistretta v. United States, 488 U.S. 361, 401 (1989) (noting that "'traditional ways of conducting government . . . give meaning' to the Constitution" (omission in original) (quoting *Youngstown*, 343 U.S. at 610 (Frankfurter, J., concurring))); Dames & Moore v. Regan, 453 U.S. 654, 686 (1981) ("Past practice does not, by itself, create power, but 'long-continued practice, known to and acquiesced in by Congress, would raise a presumption that the

When constitutional controversies implicate foreign relations, invocations of historical practice are particularly common, in part because of the lower level of judicial review in that area.[56] For example, a frequent argument in support of the constitutionality of "executive agreements" (that is, binding international agreements concluded by the President without obtaining the advice and consent of two-thirds of the Senate) is the fact that presidents have long concluded such agreements.[57] Similarly, debates over the scope of the President's authority to initiate the use of military force in the absence of congressional authorization have been heavily informed by past uses of force.[58] Yet another example is the scope of the President's authority to determine which foreign governments are recognized by the United States.[59]

Nevertheless, appeals to historical practice are not confined to matters relating to foreign affairs. For example, the Supreme Court has emphasized longstanding presidential practice when considering when the President's "pocket veto" (that is, failure to sign a bill before Congress recesses) should be deemed to operate.[60] Similarly, in concluding that the President's pardon power extended to a contempt of court conviction, the Court reasoned that "long practice under the pardoning power and acquiescence in it strongly sustains the construction it is based on."[61] Moreover, as Trevor Morrison and I have noted elsewhere, "arguments about the scope of both the 'executive privilege' (concerning the ability to withhold internal executive branch communications from the other branches of government) and the

---

[action] had been [taken] in pursuance of its consent . . . .'" (alterations in original) (quoting United States v. Midwest Oil Co., 236 U.S. 459, 474 (1915))); WILLIAM HOWARD TAFT, OUR CHIEF MAGISTRATE AND HIS POWERS 135 (1916) ("Executive power is sometimes created by custom, and so strong is the influence of custom that it seems almost to amend the Constitution.").

56. Bradley & Morrison, *supra* note 1, at 420, 456.

57. *See, e.g.*, Whether Uru. Round Agreements Required Ratification as a Treaty, 18 Op. O.L.C. 232, 234 (1994) ("[P]ractice under the Constitution has established that the United States can assume major international trade obligations such as those found in the Uruguay Round Agreements when they are negotiated by the President and approved and implemented by Act of Congress . . . ."); *see also* Bradley & Morrison, *supra* note 1, at 468–76 (describing the role of historical practice in debates over the validity of congressional–executive agreements).

58. *See* Bradley & Morrison, *supra* note 1, at 461–68 (describing the role of historical practice in the war powers area). *Compare, e.g.*, Authority to Use Military Force in Libya, 35 Op. O.L.C. (Apr. 1, 2011), http://www.justice.gov/olc/2011/authority-military-use-in-libya.pdf (relying on historical practice in support of the argument that President Obama had the unilateral authority to initiate the use of military force in Libya), *with* Michael J. Glennon, *The Cost of "Empty Words": A Comment on the Justice Department's Libya Opinion*, HARV. NAT'L SECURITY J.F. 1, 3–4 (2011) (arguing that historical practice did not support the exercise of this authority).

59. For a recent example, see Zivotofsky *ex rel.* Zivotofsky v. Sec'y of State, 725 F.3d 197, 207 (D.C. Cir. 2013) ("We conclude that longstanding post-ratification practice supports the Secretary's position that the President exclusively holds the recognition power.").

60. *See* The Pocket Veto Case, 279 U.S. 655, 689 (1929) ("Long settled and established practice is a consideration of great weight in a proper interpretation of constitutional provisions of this character.").

61. *Ex parte* Grossman, 267 U.S. 87, 118–19 (1925).

'legislative privilege' (concerning, among other things, the internal powers of the two houses of Congress) are commonly informed by historical practice."[62]

On these and other separation of powers issues, lawyers and judges trained in the common law naturally look for precedent in evaluating legal claims, and when judicial precedent is lacking, it is not surprising that they turn to other forms of precedent.[63] Executive Branch agencies such as the Office of Legal Counsel also give weight to historical practice for reasons somewhat akin to the reasons that courts give weight to their own prior decisions under the doctrine of stare decisis, such as decisional efficiency and the protection of reliance interests.[64] Historical practice is particularly likely to be invoked for separation of powers issues not specifically addressed by the constitutional text,[65] as is the case for treaty termination. Among other things, when the implications of text are perceived to be unclear, appeals to past practice allow for a type of principled reasoning that might not otherwise be possible.[66]

To say that reliance on historical practice is unsurprising in this context is not to say that it is normatively attractive, and some of the tradeoffs associated with this sort of constitutional reasoning are explored in Part IV. One particular difficulty with a practice-based approach to the separation of powers is worth noting here: Most accounts of how historical practice can inform constitutional interpretation in this context require that the branch of government that is affected by a practice "acquiesce" in it before it is credited.[67] As Trevor Morrison and I have explained, however,

---

62. Bradley & Morrison, *supra* note 1, at 421; *see also* JOSH CHAFETZ, DEMOCRACY'S PRIVILEGED FEW 3–19 (2007); Archibald Cox, *Executive Privilege*, 122 U. PA. L. REV. 1383, 1384–405 (1974).

63. For a general consideration of the role of nonjudicial precedent in constitutional law, see Michael J. Gerhardt, *Non-Judicial Precedent*, 61 VAND. L. REV. 713 (2008).

64. *See, e.g.*, Trevor W. Morrison, *Stare Decisis in the Office of Legal Counsel*, 110 COLUM. L. REV. 1448 (2010).

65. *See* Bradley & Morrison, *supra* note 1, at 455–56.

66. *See id.*

67. *See, e.g.*, Glennon, *supra* note 1, at 134 ("[T]he branch placed on notice must have acquiesced in the custom."); Harold Hongju Koh, *Focus: Foreign Affairs Under the United States Constitution*, 13 YALE J. INT'L L. 1, 3 n.7 (1988) ("Under the heading of 'quasi-constitutional custom,' I would of course include executive practice of which Congress has approved or in which it has acquiesced."); Peter J. Spiro, *War Powers and the Sirens of Formalism*, 68 N.Y.U. L. REV. 1338, 1356 (1993) (reviewing JOHN HART ELY, WAR AND RESPONSIBILITY: CONSTITUTIONAL LESSONS OF VIETNAM AND ITS AFTERMATH (1993)) ("[T]he other branch must have accepted or acquiesced in the action."); Jane E. Stromseth, *Understanding Constitutional War Powers Today: Why Methodology Matters*, 106 YALE L.J. 845, 880 (1996) (reviewing LOUIS FISHER, PRESIDENTIAL WAR POWER (1995)) ("Congress . . . must not only be on notice of an executive practice and accompanying claim of authority to act; it also must accept or acquiesce in that practice and claim of authority."); *see also* Dames & Moore v. Regan, 453 U.S. 654, 680 (1981) ("Crucial to our decision today is the conclusion that Congress has implicitly approved the practice of claim settlement by executive agreement.").

acquiescence is a problematic concept, especially as applied to Congress.[68] Among other things, accounts of congressional acquiescence often assume a "Madisonian" model of interbranch rivalry that probably does not describe modern congressional–executive relations.[69] A number of factors contribute to the descriptive inaccuracy of this model, including the modern party system, which reduces the incentives of individual members of Congress to act systematically in constraining executive power or resisting executive aggrandizement.[70] If nothing else, the limitations with the acquiescence concept suggest that there should be a high bar for claims of congressional acquiescence and that greater attention should be paid to potential indications of congressional nonacquiescence that fall short of the enactment of contrary legislation, such as various forms of congressional "soft law."[71]

In theory, the courts could determine whether and to what extent the historical practice relating to treaty termination should be credited. A variety of justiciability limitations, however, make this unlikely. The Supreme Court declined to resolve the dispute over the termination of the Taiwan Treaty because of these limitations, with four Justices concluding that the case presented a political question and Justice Powell concluding that the case was not ripe for judicial review.[72] Since that decision, the Supreme Court has sharply limited the standing of members of Congress to challenge presidential action.[73] In 2002, a federal district court dismissed a suit brought by thirty-two members of Congress challenging President Bush's termination of the ABM treaty, based on both a lack of standing and the political question doctrine.[74] For these reasons, it can be expected that

---

68. *See* Bradley & Morrison, *supra* note 1, at 432–47.

69. *See id.* at 438–47; *see also* Daryl J. Levinson, *Parchment and Politics: The Positive Puzzle of Constitutional Commitment*, 124 HARV. L. REV. 657, 671 (2011) ("[A]ll indications are that political 'ambition counteracting ambition' has failed to serve as a self-enforcing safeguard for the constitutional structures of federalism and separation of powers in the way that Madison seems to have envisioned."); Eric A. Posner & Adrian Vermeule, *The Credible Executive*, 74 U. CHI. L. REV. 865, 884 (2007) ("Whether or not this [Madisonian] picture was ever realistic, it is no longer so today.").

70. *See generally* Daryl J. Levinson & Richard H. Pildes, *Separation of Parties, Not Powers*, 119 HARV. L. REV. 2311 (2006).

71. *See* Bradley & Morrison, *supra* note 1, at 446, 450. For discussion of congressional soft law, see generally Josh Chafetz, *Congress's Constitution*, 160 U. PA. L. REV. 715 (2012), and Jacob E. Gersen & Eric A. Posner, *Soft Law: Lessons from Congressional Practice*, 61 STAN. L. REV. 573 (2008).

72. Goldwater v. Carter, 444 U.S. 996, 997–1006 (1979).

73. *See* Raines v. Byrd, 521 U.S. 811, 823–24 (1997) (holding that members of Congress generally do not have standing to sue for injury to their institution absent a showing that their votes have been "completely nullified").

74. Kucinich v. Bush, 236 F. Supp. 2d 1, 2 (D.D.C. 2002); *see also* Beacon Prods. Co. v. Reagan, 633 F. Supp. 1191, 1198–99 (D. Mass. 1986), *aff'd*, 814 F.2d 1 (1st Cir. 1987) (relying on the political question doctrine to dismiss a challenge to a treaty termination by President Reagan).

historical practice will continue to develop relating to this issue and that disputes will continue to be resolved outside the courts.

* * *

The next two Parts of this Article consider the historical practice of U.S. treaty terminations.[75] Part II shows that, at least until the late nineteenth century, it was generally understood that presidents needed the agreement of Congress or the Senate in order to terminate a treaty. Part III describes the shift during the twentieth century towards unilateral presidential termination of treaties. As will be seen, the shift did not happen all at once but rather occurred over the course of decades as a result of repeated claims and actions by the Executive Branch in the face of congressional inaction. When a controversy finally did develop over this question of institutional authority—in connection with President Carter's termination of the Taiwan treaty—the President was able to plausibly maintain that his action was consistent with longstanding practice.

II. Founding Through the Early Twentieth Century

This Part reviews the instances, during the period from the constitutional founding through the early twentieth century, in which the United States announced that it was terminating or suspending treaty obligations. In doing so, it divides the practice into four categories:

- termination pursuant to *ex ante* congressional authorization or directive;
- termination pursuant to senatorial authorization;
- termination with post hoc congressional or senatorial approval; and
- unilateral presidential termination.

The historical practice reviewed here includes instances in which the United States ultimately decided not to terminate a treaty after announcing its intention of doing so, on the theory that such instances can shed light on the constitutional understandings of the President and Congress.[76]

---

75. For additional discussion of the historical practice, see DAVID GRAY ADLER, THE CONSTITUTION AND THE TERMINATION OF TREATIES 149–247 (1986); SAMUEL B. CRANDALL, TREATIES: THEIR MAKING AND ENFORCEMENT §§ 178–186 (2d ed. 1916); 5 GREEN HAYWOOD HACKWORTH, DIGEST OF INTERNATIONAL LAW § 509 (1943). There is also extensive discussion of the historical practice in the Senate hearings regarding President Carter's termination of the Taiwan treaty in subpart III(C) *infra*, as well as in the various State Department memoranda that are referred to throughout this Article.

76. If, for example, a President initiates a unilateral termination and Congress does not object, that would seem to be a relevant event even if the President decides to withdraw the termination for policy reasons. The approach of this Article therefore differs from that of David Adler, who suggests in his 1986 book on treaty termination that instances in which the termination was not fulfilled are not relevant in discerning the constitutional practice of treaty termination. *See* ADLER, *supra* note 75, at 164–65, 170, 184–85.

A. Congressional Authorization or Directive

In the first instance in which the United States purported to terminate treaties, Congress played an especially direct role. In 1798, on the eve of war with France, Congress passed (and President Adams signed) legislation stating that the four treaties the United States had at that time with France "shall not henceforth be regarded as legally obligatory on the government or citizens of the United States."[77] In the congressional debates over whether to enact the statute, there does not appear to have been any doubt about Congress's constitutional authority to terminate the treaties. One member of the House did observe that "[i]n most countries it is in the power of the Chief Magistrate to suspend a treaty whenever he thinks proper," but he noted that "here Congress only has that power."[78] Several years later, Thomas Jefferson referred to this episode in his *Manual of Parliamentary Practice*, observing that "[t]reaties being declared, equally with the laws of the U[nited] States, to be the supreme law of the land, it is understood that an act of the legislature alone can declare them infringed and rescinded. This was accordingly the process adopted in the case of France in 1798."[79]

Notwithstanding Jefferson's contention that legislative action was the exclusive method of terminating a treaty, the 1798 statute appears to be the only instance in U.S. history in which the full Congress purported to effectuate a termination directly. As noted in subpart I(B), it has generally been understood that formal communications between the United States and other nations are channeled through the Executive Branch. A possible exception to that "sole organ" role for the Executive, however, is Congress's authority to declare war. A state of war was understood as terminating certain types of treaty relationships, such as treaties of alliance.[80] So one way of understanding Congress's termination of the

---

77. Act of July 7, 1798, ch. 67, 1 Stat. 578, 578. Congress had already passed other war-related measures by that point. *See* ALEXANDER DECONDE, THE QUASI-WAR 102 (1966) (discussing a direct property tax enacted to pay for the expanded war program).

78. 8 ANNALS OF CONG. 2120 (1798). For additional discussion of the debate in Congress, see DAVID P. CURRIE, THE CONSTITUTION IN CONGRESS: THE FEDERALIST PERIOD, 1789–1801, at 250–53 (1997).

79. THOMAS JEFFERSON, A MANUAL OF PARLIAMENTARY PRACTICE § 52 (Wash., Samuel Harrison Smith 1801); *see also* Ware v. Hylton, 3 U.S. (3 Dall.) 199, 261 (1796) (Iredell, J.) (suggesting that only Congress has the authority to terminate a treaty based on a violation by the other party). Many years later, the U.S. Court of Claims held that the French treaties had been validly terminated by Congress. *See* Hooper v. United States, 22 Ct. Cl. 408, 418 (1887) ("The treaties therefore ceased to be a part of the supreme law of the land . . . ."); *see also* Chirac v. Chirac, 15 U.S. (2 Wheat.) 259, 272 (1817) (assuming that the treaties had been terminated in deciding a related property claim). During negotiations between the United States and France in 1800, however, France took the position that the U.S. treaty obligations had not been terminated (although not because of any claim that Congress was unable to terminate treaties). 5 JOHN BASSETT MOORE, A DIGEST OF INTERNATIONAL LAW § 774, at 357 (1906).

80. *See, e.g.*, 2 VATTEL, *supra* note 23, bk. 3, § 175 ("Conventions and treaties are broken or annulled when war breaks out between the contracting parties . . . .").

French treaties was as an exercise of its power to declare war (although Congress merely authorized naval warfare against France and did not formally declare war).[81]

In any event, without purporting directly to effectuate terminations, Congress has authorized or directed presidential termination of treaties in a number of other instances. In 1846, for example, Congress passed a joint resolution authorizing President Polk "at his discretion" to terminate a treaty with Great Britain relating to the two countries' joint occupation of the Oregon Territory.[82] This resolution was issued in response to a request from the President, in which he stated that a notice of termination would, in his judgment, "be proper to give, and I recommend that provision be made by law for giving it accordingly."[83] After Congress passed the resolution, the Secretary of State informed the U.S. Ambassador to Great Britain that "Congress have spoken their will upon the subject, in their joint resolution; and to this it is his (the President's) and your duty to conform."[84] This was apparently the first time that the United States attempted to terminate a treaty pursuant to a unilateral withdrawal provision. Before the expiration of the notice period, the United States and Great Britain negotiated a new treaty to supersede the one that the United States had acted to terminate.

Prior to the issuance of the 1846 resolution, there was substantial debate in Congress over whether it was proper for the House of Representatives to be involved in the issue. During that debate, a majority of those who spoke expressed the view that it was constitutionally proper for the full Congress to authorize termination.[85] Several members of the House issued a minority report, however, arguing that, except when a treaty is being terminated pursuant to a declaration of war, authorization of treaty termination properly should come from a supermajority of the Senate, not the full Congress.[86] No one argued for a unilateral presidential power to terminate.

---

81. *See* The Head Money Cases, 112 U.S. 580, 599 (1884) (noting that a declaration of war "must be made by Congress, and . . . when made, usually suspends or destroys existing treaties between the nations thus at war"); *cf.* Bas v. Tingy, 4 U.S. (4 Dall.) 37, 41 (1800) (Washington, J.) (concluding that France in 1799 qualified as an "enemy" within the meaning of a naval salvage statute and noting that "[C]ongress had raised an army; stopped all intercourse with *France*; dissolved our treaty; built and equipt ships of war; and commissioned private armed ships").

82. Joint Resolution of Apr. 27, 1846, 9 Stat. 109, 109–10.

83. James K. Polk, First Annual Message (Dec. 2, 1845), *in* 5 A COMPILATION OF THE MESSAGES AND PAPERS OF THE PRESIDENTS 2235, 2245 (James D. Richardson ed., 1897).

84. S. DOC. NO. 29-489, at 15 (1st Sess. 1846).

85. *See* CONG. GLOBE, 29th Cong., 1st Sess. 635 (1846).

86. H.R. REP. NO. 29-34, at 1–3 (1984). Some senators also expressed this view. *See, e.g.*, CONG. GLOBE, 29th Cong., 1st Sess. 635 (1846) (statement of Sen. Mangum) (contending that "[t]he power of treaty-making was one highly restricted by the Constitution—the Senate—two-thirds of it—and the Executive possessed the power," and, therefore, the Congress did not have the power to make or break a treaty). For additional discussion of the debate in Congress, see

Congress authorized additional treaty terminations in 1865 and 1874 without controversy.[87] In 1876, President Grant informed Congress that Great Britain was not complying with an extradition provision in a treaty, and he stated that "[i]t is for the wisdom of Congress to determine whether the article of the treaty relating to extradition is to be any longer regarded as obligatory on the Government of the United States or as forming part of the supreme law of the land."[88] In the meantime, he indicated that he would not comply with extradition requests from Great Britain under the treaty "without an expression of the wish of Congress that I should do so."[89] Extradition by the United States under the treaty was then suspended for six months until the dispute with Great Britain was resolved.[90]

Sometimes Congress went beyond authorizing the President to terminate treaties and affirmatively ordered him to do so. In 1883, for example, Congress directed President Arthur to terminate various articles in an 1871 treaty with Great Britain, and Arthur subsequently terminated the articles.[91] In 1915, Congress, in the Seaman's Act, "requested and directed" President Wilson to give notice of termination of various treaty obligations inconsistent with the Act,[92] and Wilson proceeded to do so.[93]

DAVID P. CURRIE, THE CONSTITUTION IN CONGRESS: DESCENT INTO THE MAELSTROM, 1829–1861, at 78–80 (2005).

87. In 1865, Congress directed the President to terminate an 1854 Reciprocity Treaty with Great Britain that concerned trade with Canada, and the Johnson Administration subsequently did so. *See* Joint Resolution of Jan. 18, 1865, 13 Stat. 566; *see also* Letter from Charles Francis Adams, Minister to the U.K., to William H. Seward, U.S. Sec'y of State (Mar. 23, 1865), *in* PAPERS RELATING TO FOREIGN AFFAIRS, pt. 1, at 258 (Wash., Gov't Printing Office 1866); Letter from William H. Seward, U.S. Sec'y of State, to Charles Francis Adams, Minister to the U.K. (Jan. 18, 1865), *in* PAPERS RELATING TO FOREIGN AFFAIRS, *supra*, at 93. In 1874, Congress authorized the President to terminate a Treaty of Commerce and Navigation with Belgium, and President Grant immediately did so. *See* Joint Resolution of June 17, 1874, 18 Stat. 287; Letter from Hamilton Fish, U.S. Sec'y of State, to J.R. Jones, Minister to Belgium (June 17, 1874), *in* PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES 64 (Wash., Gov't Printing Office 1874); *see also* Ulysses S. Grant, Sixth Annual Message (Dec. 7, 1874), *in* 10 A COMPILATION OF THE MESSAGES AND PAPERS OF THE PRESIDENTS, *supra* note 83, at 4238, 4242 (reporting that "[t]he notice directed by the resolution of Congress of June 17, 1874, to be given to terminate the convention of July 17, 1858, between the United States and Belgium has been given, and the treaty will accordingly terminate on the 1st day of July, 1875").

88. Letter from Ulysses S. Grant to the Senate and House of Representatives (June 20, 1876), *in* 10 A COMPILATION OF THE MESSAGES AND PAPERS OF THE PRESIDENTS, *supra* note 83, at 4324, 4327.

89. *Id.*

90. CRANDALL, *supra* note 75, § 185, at 464.

91. Joint Resolution of Mar. 3, 1883, 22 Stat. 641; Letter from Frederick T. Frelinghuysen, U.S. Sec'y of State, to J.R. Lowell, Minister to the U.K. (Apr. 5, 1883), *in* PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES 413, 413–14 (Wash., Gov't Printing Office 1884).

92. Seaman's Act, ch. 153, § 16, 38 Stat. 1164, 1184 (1915).

93. Circular from William Jennings Bryan, U.S. Sec'y of State, to Ambassador Page (May 29, 1915), *in* PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES 3 (1924). In *Van Der Weyde v. Ocean Transport Co.*, 297 U.S. 114 (1936), the Supreme Court upheld the

At times, however, presidents argued that Congress could not constitutionally compel them to take certain actions relating to a treaty. In 1879, for example, President Hayes vetoed an immigration bill on the ground that it was trying to get him to partially terminate a treaty. In the bill, Congress directed the President to terminate two provisions in a treaty with China relating to Chinese immigration.[94] In his veto message, President Hayes conceded that Congress had the authority to terminate a treaty, and in fact said that this was "free from controversy."[95] But he pointed out that the bill called for the abrogation only of *parts* of a treaty and argued that "the power of making new treaties or modifying existing treaties is not lodged by the Constitution in Congress, but in the President, by and with the advice and consent of the Senate."[96]

In 1920, President Wilson refused to implement a provision in the Merchant Marine Act (also known as the Jones Act)[97] that stated that he was "authorized and directed" to terminate within ninety days various treaty obligations that disallowed the United States from imposing discriminatory customs duties and tonnage dues.[98] The State Department issued a press release explaining that, while the Act was seeking to have the President partially terminate treaties, the treaties in question did not allow for such partial termination.[99] In explaining the proposed press release to the Undersecretary of State, the Solicitor for the State Department cited President Hayes's reasoning in his veto of the Chinese immigration bill and noted that although "Congress may pass an act violative of a treaty" and "may express its sense that a treaty should be terminated," it "cannot in effect undertake legally to modify a treaty."[100] Not surprisingly, Wilson's

---

termination of provisions in a treaty with Sweden and Norway pursuant to the directive in the Seaman's Act and noted that it was unnecessary in that case to address "the authority of the Executive in the absence of congressional action, or of action by the treaty-making power, to denounce a treaty of the United States." *Id.* at 117.

94. An Act to Restrict the Immigration of Chinese to the United States, H.R. 2423, 45th Cong. (1878).

95. Rutherford B. Hayes, Veto of the Chinese Immigration Bill, H.R. EXEC. DOC. NO. 45-102, at 5 (3d Sess. 1879).

96. *Id.*

97. *See* Statement by State Department (Sept. 24, 1920), *in* 17 A COMPILATION OF THE MESSAGES AND PAPERS OF THE PRESIDENTS 8871, 8871–72 (1927).

98. Merchant Marine (Jones) Act, ch. 250, § 34, 41 Stat. 988, 1007 (1920).

99. Press Release, U.S. Dep't of State 2–3 (Sept. 6, 1920) (on file with author); *see also President Won't Denounce Treaties; Defies Congress*, N.Y. TIMES, Sept. 25, 1920, http://query.nytimes.com/mem/archive-free/pdf?res=F00F10FC345511738DDDAC0A94D1405B808EF1D3 (reporting on the State Department's issuance of a press release explaining the President's refusal to implement the Act to the extent it would entail the illegal termination of treaty obligations).

100. Memorandum from the Solicitor of the Dep't of State to Norman H. Davis, U.S. Undersecretary of State 2–3 (Sept. 6, 1920) (on file with author). The Office of the Solicitor was the chief legal advisor to the State Department from 1891 to 1931 and was based within the Department of Justice. Harold Hongju Koh, Remarks, *The State Department Legal Adviser's*

refusal to implement this portion of the statute, after having signed the statute into law, generated controversy.[101]

The next year, Secretary of State Charles Evans Hughes advised President Harding that a partial termination like the one contemplated by Congress in the Jones Act was not permissible under international law. "As the existing treaties do not permit such partial termination by notice," explained Hughes, "it follows that Congress has failed to give a mandate on which the President can act."[102] There was no suggestion, however, that Congress could not direct the President to terminate a treaty in its entirety.

B. Senatorial Authorization

The President has occasionally terminated a treaty based on prior authorization solely from the Senate. In 1855, the Senate issued a resolution authorizing the President to terminate a Friendship, Commerce, and Navigation Treaty with Denmark, after President Pierce had indicated that he thought termination was warranted.[103] In announcing the U.S. termination, President Pierce noted that he was acting "[i]n pursuance of the authority conferred by a resolution of the Senate."[104]

The following year, the Senate debated whether it could properly act in this manner without the involvement of the House of Representatives. Senator Charles Sumner argued that, because a treaty is part of the supreme law of the land, it should only be repealed through action of the full legislature.[105] The Senate asked the Foreign Relations Committee to consider the issue, and the Committee prepared a report on the subject. It concluded that termination pursuant to senatorial authorization was constitutionally proper, at least where, as here, the treaty specifically

---

Office: Eight Decades in Peace and War, 100 GEO. L.J. 1747, 1750 (2012). It was replaced in 1931 by the Office of the Legal Adviser, which is based in the State Department. *Id.*

101. *See* Jesse S. Reeves, *The Jones Act and the Denunciation of Treaties*, 15 AM. J. INT'L L. 33, 33–34, 37–38 (1921).

102. Memorandum prepared by Charles E. Hughes, U.S. Sec'y of State, for President Harding 30 (Oct. 8, 1921) (on file with author).

103. *See* Franklin Pierce, Second Annual Message (Dec. 4, 1854), *in* 7 A COMPILATION OF THE MESSAGES AND PAPERS OF THE PRESIDENTS, *supra* note 83, at 2806, 2812 (stating that "I deem it expedient that the contemplated notice should be given to the Government of Denmark").

104. Franklin Pierce, Third Annual Message (Dec. 31, 1855), *in* 7 A COMPILATION OF THE MESSAGES AND PAPERS OF THE PRESIDENTS, *supra* note 83, at 2860, 2867. Some years earlier, Secretary of State James Buchanan had informed Denmark that, in order for the United States to withdraw from the treaty, "an Act must first pass Congress to enable the President to give the required notice." Letter from James Buchanan, Sec'y of State, to Robert P. Flenniken, Minister to Den. (Oct. 14, 1848), *in* 8 THE WORKS OF JAMES BUCHANAN, 220, 224 (John Bassett Moore ed., 1909).

105. CONG. GLOBE, 34th Cong., 1st Sess. 600 (1856) (statement of Sen. Sumner). Sumner, an ardent abolitionist, was apparently concerned that the pro-Southern Senate would seek to terminate a provision in the 1842 Webster–Ashburton Treaty that required patrols off the coast of Africa to suppress the slave trade. Reeves, *supra* note 101, at 35.

allowed for unilateral withdrawal. The Committee observed that, "so far as the 'practice of the government' is concerned, there is nothing to question the sufficiency of the notice that has been given to Denmark to terminate the treaty."[106] It appears that the only other instance of senatorial (as opposed to full congressional) involvement in a treaty termination occurred in 1921, when the Senate gave its advice and consent to U.S. termination of the International Sanitary Convention, based on a request from President Wilson.[107]

C. Ex Post *Congressional or Senatorial Approval*

Sometimes the President has acted to terminate a treaty and obtained subsequent approval from either the full Congress or the Senate. In 1864, for example, President Lincoln gave notice of termination of the Great Lakes Agreement with Great Britain (also known as the Rush–Bagot Agreement), which limited the naval military presence of the United States on the Lakes, pursuant to a six-months' notice provision in the Agreement.[108] Congress subsequently passed a joint resolution (which Lincoln signed) "adopt[ing] and ratif[ying]" the termination "as if the same had been authorized by [C]ongress."[109] In the debate on this resolution, Senator Davis objected that Congress was creating a "mischievous precedent. . . . which is to sanction and to give authority to an unauthorized act by the President."[110] Other senators agreed that Congress needed to approve the termination but thought that Congress could do so retroactively.[111] Despite the Senate's action, the President decided to

106. S. REP. NO. 34-97, at 7–8 (1st Sess. 1856). Senator Sumner (and other Senators) continued to dispute the point. CONG. GLOBE, 34th Cong., 1st Sess. 1147 (1856). For additional discussion of the debate in Congress, see CURRIE, *supra* note 86, at 80–84. A resolution was proposed in the Senate that would have confirmed the validity of the Senate's action, but it was never voted on. CONG. GLOBE, 34th Cong., 1st Sess. 826 (1856).

107. S. Res. of May 26, 1921, 67th Cong., 61 CONG. REC. 1793; *see also* 61 CONG. REC. 1793–94 (1921) (providing the text of both President Wilson's request for the Senate's advice and consent to terminate the treaty and the Senate resolution providing this authorization). For another reference to the idea of senatorial involvement in treaty termination, see *Techt v. Hughes*, 128 N.E. 185, 192 (N.Y. 1920) (Cardozo, J.), where the court found a treaty with the Austro-Hungarian Empire to still be in effect despite World War I and observed that the "President and senate may denounce the treaty, and thus terminate its life," a statement that was quoted by the Supreme Court in *Clark v. Allen*, 331 U.S. 503, 509 (1947).

108. *See* Abraham Lincoln, Fourth Annual Message (Dec. 6, 1864), *in* 8 A COMPILATION OF THE MESSAGES AND PAPERS OF THE PRESIDENTS, *supra* note 83, at 3444, 3447. Lincoln was responding to Confederate raids from Canada. *See id.*

109. Joint Resolution of Feb. 9, 1865, 13 Stat. 568.

110. CONG. GLOBE, 38th Cong., 2d Sess. 312 (1865); *see also id.* at 313 (Sen. Davis) ("[I]t is indispensably incumbent and necessary, in order to secure the termination of this treaty, that it shall be terminated not by the action of the President, but by the action of Congress.").

111. *See, e.g.*, *id.* at 313 (Sen. Sumner); *id.* at 314–15 (Sen. Johnson).

rescind the notice of termination after further negotiations with Great Britain, so it never took effect.[112]

Another example is President Taft's action in 1911, when he gave notice to Russia of a termination of a commercial treaty. In response to Russia's mistreatment of American Jews, the House of Representatives had passed a strongly worded resolution demanding termination of the treaty (on a vote of 301 to 1),[113] and the resolution was thought likely to pass in the Senate.[114] Taft, who had been reluctant to terminate the treaty at all, was concerned that the harsh tone of the House resolution would needlessly offend Russia.[115] He therefore quickly communicated his own statement of termination to Russia and submitted that statement to the Senate "with a view to its ratification and approval."[116] The Senate Foreign Relations Committee proposed a joint resolution stating that the notice of termination was "adopted and ratified," and this resolution was subsequently passed by both houses of Congress (with the vote in the Senate being unanimous) and was signed by the President.[117] The discussion in Congress primarily concerned whether the Senate or the full Congress should be involved in approving the termination, not whether the President had a unilateral power of termination.[118]

---

112. ADLER, *supra* note 75, at 164–65.

113. *See* 48 CONG. REC. 353 (1911). In the deliberations on this resolution in the House Committee on Foreign Affairs, the prominent constitutional lawyer Louis Marshall testified that the proper procedure for terminating a treaty was by joint resolution of Congress. He noted that he initially "had an idea that the executive department had ample power to deal with the matter," but, after studying the historical practice, he had reached the conclusion "that the power rests in Congress." *Termination of the Treaty of 1832 Between the United States and Russia: Hearing Before the H. Comm. on Foreign Affairs*, 62d Cong. 42 (1911) (statement of Louis Marshall). In his testimony, Marshall presented Congress with a memorandum (prepared by Herbert Friedenwald, Secretary of the American Jewish Committee) describing the past practice of treaty terminations, which was reprinted as an appendix to the committee hearings. *See id.* at 49, app. III at 295.

114. ADLER, *supra* note 75, at 181.

115. *Id.* at 182.

116. 48 CONG. REC. 453 (1911); s*ee also Taft Himself May End Treaty*, N.Y. TIMES, Dec. 18, 1911, http://query.nytimes.com/mem/archive-free/pdf?res=F70D11FA395517738DDDA10994DA415B818DF1D3 (discussing the likelihood of Taft denouncing the Russian treaty and asking only for the Senate's approval, thereby avoiding presidential approbation of the harsh statement in the House).

117. *See* Joint Resolution of Dec. 21, 1911, 37 Stat. 627 (1911); 48 CONG. REC. 507 (1911) (recording the Senate vote); *id.* at 600 (documenting the fact that the President had signed the resolution).

118. *See* 48 CONG. REC. 484 (statement of Senator Stone noting that the issue was whether the termination should be accomplished "with the joint sanction of the two Houses of Congress or whether it should be taken by the President with the approval of the Senate alone"). *Compare, e.g.*, *id.* at 473 (statement of Senator Rayner that "[a] treaty is the supreme law of the land under the language of the Constitution, and the supreme law of the land ought not to be set aside except by legislative action of both Houses"), *with id.* at 479 (statement of Senator Lodge that "in cases where treaties have involved no legislation the power of the Senate and the President to terminate a treaty by notice, or to arrest its operation . . . is absolute, because in making such a treaty the

Both of these episodes are obviously closer to presidential unilateralism than situations in which the President obtains advance authorization for a treaty termination. Lincoln's action was controversial for that reason.[119] The Taft episode was less controversial because it was obvious when Taft acted that Congress supported termination of the treaty and was in fact the driving force behind the decision to terminate, and the only issue was over how the message would be conveyed to Russia. It is also worth noting that, when writing some years later about presidential power, Taft made clear that he thought that "[t]he President may not annul or abrogate a treaty without the consent of the Senate unless he is given that specific authority by the terms of the treaty."[120]

D. Unilateral Presidential Termination

In modern debates, the Executive Branch has sometimes claimed that the first unilateral presidential termination of a treaty occurred in 1815,[121] but that is erroneous. The Madison administration observed that year that a treaty with The Netherlands, which had been concluded in 1782, had been "annulled" in light of the fact that The Netherlands had in the meantime been assimilated into the French Empire of Napoleon and then reconstructed in the Congress of Vienna.[122] The observation occurred in response to a suggestion by The Netherlands that the two countries conclude a new treaty based on the terms of the old one, a suggestion that itself assumed that the old treaty was no longer in force.[123] Under

---

Senate and the President represent the high contracting party"). Some members of the House of Representatives cited historical practice in support of the proposition that the full Congress could terminate a treaty. *See, e.g.*, *id.* at 319 (statement of Rep. Legare) (citing treaties that Congress had terminated in the past); *id.* at 331 (statement of Rep. Peters) (referring to past treaties abrogated by Congress).

119. The Rush–Bagot Agreement that Lincoln had proposed terminating was originally concluded by President Monroe unilaterally based on his Commander in Chief authority, although it eventually received senatorial advice and consent. *See* BRADLEY, *supra* note 27, at 90. As a result, it may have been viewed as occupying an uncertain place between sole executive agreements (which indisputably can be terminated unilaterally by the President) and Article II treaties.

120. TAFT, *supra* note 55, at 115–16.

121. *See, e.g.*, Memorandum from Herbert J. Hansell, Legal Adviser, U.S. Dep't of State, to Cyrus R. Vance, U.S. Sec'y of State, President's Power to Give Notice of Termination of U.S.–ROC Mutual Defense Treaty (Dec. 15, 1978) [hereinafter Hansel Memorandum], *in* S. COMM. ON FOREIGN RELATIONS, 95TH CONG., TERMINATION OF TREATIES: THE CONSTITUTIONAL ALLOCATION OF POWER 395, 397 (Comm. Print 1978) ("In 1815, President Madison exchanged correspondence with the Netherlands which has been construed by the United States as establishing that the 1782 Treaty of Amity and Commerce between the two countries had been annulled.").

122. *See* CRANDALL, *supra* note 75, § 179, at 429 ("The state thus formed, although in general considered the successor to, differed in name, territory, and form of government from, the state which had entered into the treaty of October 8, 1782 with the United States.").

123. *Id.*

international law at that time, a treaty imposing reciprocal obligations became void if one of the parties ceased to exist—for example, if it was conquered by another nation.[124] The United States, therefore, did not terminate this treaty.

There is an even earlier episode that, although it did not involve any announced treaty termination, is sometimes cited in support of a unilateral presidential authority to suspend or terminate treaties.[125] In 1793, there was a debate within George Washington's cabinet over whether to receive an ambassador from revolutionary France with, or without, qualifications.[126] Receiving him without qualifications might signal that the United States accepted the continuing effect of the treaties it had with France, including a treaty of alliance, notwithstanding the changes in France's government.[127] Receiving him with qualifications, by contrast, might allow the United States the option of suspending or terminating the treaties.[128] Secretary of the Treasury Alexander Hamilton and Secretary of War Henry Knox thought the ambassador should be received with qualifications, whereas Secretary of State Thomas Jefferson and Attorney General Edmund Randolph thought he should be received without qualifications.[129] The cabinet members prepared memoranda focused on whether international law allowed for suspension or termination of the treaties under these circumstances.[130]

Ultimately, Washington decided to receive the ambassador without qualifications, so there was no effort to reserve the option of suspending or

---

124. *See, e.g.*, VATTEL, *supra* note 23, § 203 (noting that a treaty comes to an end "if, for any cause whatever, the Nation should lose its character as an independent political society"); *see also* HENRY WHEATON, ELEMENTS OF INTERNATIONAL LAW 191 (London, Cary, Lea & Blanchard 1836) ("*Treaties* . . . expire of course:—1. In case either of the contracting parties loses its existence as an independent State.").

125. *See* Prakash & Ramsey, *supra* note 30, at 324–26 (arguing that President Washington's belief that he could renounce the treaties with France suggests that people during this period believed that the President had the power to terminate or suspend treaties); Yoo & Delahunty Memorandum, *supra* note 52, at 15–16 (citing Hamilton and Knox's recommendation to Washington that he consider suspending the French treaty as evidence of a general understanding that the President had unilateral authority to suspend treaties).

126. *See generally* WILLIAM R. CASTO, FOREIGN AFFAIRS AND THE CONSTITUTION IN THE AGE OF FIGHTING SAIL 32–33 (2006) (describing the debate); STANLEY ELKINS & ERIC MCKITRICK, THE AGE OF FEDERALISM: THE EARLY AMERICAN REPUBLIC, 1788-1800, at 339–41 (1993) (same).

127. Bradley & Flaherty, *supra* note 32, at 667.

128. *Id.*

129. *See id.* at 667–68.

130. *See, e.g.*, Letter from Alexander Hamilton and Henry Knox to George Washington (May 2, 1793), *in* 14 THE PAPERS OF ALEXANDER HAMILTON 367, 367–96 (Harold C. Syrett et al. eds., 1969) (arguing that the United States could choose to suspend or even renounce the treaties with France); Letter from Thomas Jefferson to George Washington (Apr. 28, 1793), *in* 25 THE PAPERS OF THOMAS JEFFERSON 607, 607–18 (John Catanzariti et al. eds., 1992) (arguing that the United States should not renounce the French treaties).

terminating the treaties.[131] The cabinet members did not discuss in their memoranda whether it was proper as a matter of U.S. constitutional law for the President to suspend or terminate treaties unilaterally. Their silence might suggest that they assumed that the President had this authority, especially since they decided not to call Congress into special session, but this is reading a lot into mere silence.[132] In a related context, Alexander Hamilton did take the position that the President had the authority to suspend a treaty in response to a revolutionary change in the government of the other treaty party, but James Madison (whose views in this period were similar to Jefferson's) sharply disputed Hamilton's claim.[133] In any event, the Executive Branch never made any public claim of a unilateral suspension or termination authority, so there was no opportunity to find out Congress's views on the matter, and certainly no circumstance for crediting any sort of congressional acquiescence. Finally, when the United States did take action five years later to terminate the French treaties, it did so, as noted above, by congressional resolution, not unilateral executive action.[134]

The first instance in which a President actually proceeded to terminate treaty provisions without even after-the-fact congressional or senatorial approval appears to have been in 1899, when the McKinley Administration terminated certain clauses in an 1850 commercial treaty with Switzerland.[135] McKinley did not terminate the entire treaty, and in fact some provisions in the treaty remain in effect even today.[136] In addition, McKinley's action need not be viewed as purely unilateral, given that he

131. *See* Bradley & Flaherty, *supra* note 32, at 669.

132. *Cf.* CURRIE, *supra* note 78, at 182 n.63 (noting that Washington's decision "avoid[ed] the difficult constitutional question whether the President alone could terminate a treaty").

133. *Compare* Hamilton, *supra* note 39, at 42 ("Hence in the case stated, though treaties can only be made by the President and Senate, their activity may be continued or suspended by the President alone."), *with* James Madison, *"Helvidius" Number 3* (Sept. 7, 1793), *reprinted in* 15 THE PAPERS OF JAMES MADISON 95, 99 (Thomas A. Mason et al. eds., 1985) ("Nor can [the President] have any more right to suspend the operation of a treaty in force as a law, than to suspend the operation of any other law."). More than two years earlier, Madison had suggested in a letter that the termination of a treaty in response to a breach by the other party required either congressional or senatorial approval. *See* Letter from James Madison to Edmund Pendleton (Jan. 2, 1791), *in* 13 THE PAPERS OF JAMES MADISON 342, 344 (Charles F. Hobson et al. eds., 1981) (stating that the Constitution requires that only the Legislature can terminate a "Treaty of peace" (emphasis omitted)).

134. *See supra* note 77 and accompanying text.

135. *See* Letter from John Hay, U.S. Sec'y of State, to Ambassador Leishman (Mar. 8, 1899), *in* PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES 753, 753–54 (1901); *see also* Stefan A. Riesenfeld, *The Power of Congress and the President in International Relations: Three Recent Supreme Court Decisions*, 25 CALIF. L. REV. 643, 661 (1937) (observing that "there seems to be at least one instance where the President alone without cooperation of Senate or Congress has terminated certain treaty provisions, *i.e.*, in the case of a treaty with Switzerland").

136. U.S. DEP'T OF STATE, TREATIES IN FORCE: A LIST OF TREATIES AND OTHER INTERNATIONAL AGREEMENTS OF THE UNITED STATES IN FORCE ON JANUARY 1, 2012, at 266, *available at* http://www.state.gov/documents/organization/202293.pdf.

was responding to a potential conflict between the treaty and a federal statute. The Tariff Act of 1897 had authorized the President to negotiate reciprocal trade agreements,[137] and, pursuant to the Act, the United States had concluded such an agreement with France.[138] Switzerland contended that it was automatically entitled to the benefit of the concessions granted to France because of most-favored-nation provisions in the 1850 treaty.[139] But granting it such concessions, without obtaining in return concessions similar to the ones given by the French, would have been contrary to longstanding U.S. trade policy, including the policy of Congress reflected in the Tariff Act.[140]

E. Early Scholarly Commentary

The only treaties that the United States terminated in the early years of its history were the French treaties, and that termination was related to the imminent state of hostilities between the two countries. Moreover, early U.S. treaties did not contain clauses allowing for discretionary withdrawal,[141] so that scenario likely would not have been considered. Perhaps for these reasons, constitutional law treatises in the early part of the nineteenth century have little if any discussion of treaty termination. Thomas Sergeant's 1822 treatise on constitutional law did note, however, that "[i]t seems, the authority to declare a treaty to have been violated, and to be therefore void, belongs only to Congress; the judiciary cannot exercise it."[142] And William Rawle's constitutional law treatise, published in 1825, tied a congressional power to terminate treaties to Congress's power to declare war.[143] Similarly, Joseph Story stated in his 1833 *Commentaries on the Constitution* that "it will not be disputed, that [treaties] are subject to the legislative power, and may be repealed, like other laws, at its pleasure."[144] As discussed in the next Part, it appears that the first scholar to suggest a unilateral presidential authority to terminate treaties was Westel

137. Tariff Act of July 24, 1897, ch. 11, § 3, 30 Stat. 151, 203 (repealed 1909).

138. ADLER, *supra* note 75, at 165.

139. *Id.*

140. For the exchange of correspondence between Switzerland and the United States about this issue, see PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES, *supra* note 135, at 740–57.

141. *See supra* note 25 and accompanying text.

142. THOMAS SERGEANT, CONSTITUTIONAL LAW 403 (Phila., Abraham Small 1822).

143. *See* WILLIAM RAWLE, A VIEW OF THE CONSTITUTION OF THE UNITED STATES OF AMERICA 68 (Phila., Philip H. Nicklin 2d ed. 1829) ("Congress alone possesses the right to declare war; and the right to qualify, alter, or annul a treaty being of a tendency to produce war, is an incident to the right of declaring war."); *see also* WILLIAM ALEXANDER DUER, A COURSE OF LECTURES ON THE CONSTITUTIONAL JURISPRUDENCE OF THE UNITED STATES 184 (N.Y.C., Harper & Bros. 1843) ("[T]he power in question may be regarded as an incident to that of declaring war.").

144. 3 JOSEPH STORY, COMMENTARIES ON THE CONSTITUTION OF THE UNITED STATES § 1832, at 695 (Bos., Hilliard, Gray & Co. 1833).

Willoughby, a political science professor at Johns Hopkins, in his 1910 treatise on U.S. constitutional law.[145]

* * *

Historical practice through at least the late nineteenth century suggests an understanding that congressional or senatorial approval was constitutionally required for the termination of U.S. treaties. Not only was Congress or the Senate almost always involved in treaty terminations, but presidents generally acted as if they needed such involvement. The chief debate was simply over whether the full Congress or merely the Senate should be involved in treaty terminations, and historical practice was viewed as relevant to that debate. Lincoln's initially unilateral action in 1864 was potentially contrary to this understanding, but it was an outlier and generated constitutional criticism in Congress rather than acquiescence. Grant's action in 1876 might have suggested some unilateral authority to *suspend* a treaty obligation, but this action was embedded within an acknowledgment of the need for congressional approval of termination. It was not until McKinley's action with respect to the Swiss treaty in 1899 that there was anything resembling a clear precedent for a unilateral presidential termination authority, and that action involved only a partial termination and was arguably part of an effort to implement congressional policy. Moreover, at least before the 1899 termination, the Executive Branch made no claim of a unilateral termination authority. For example, in the digests of international practice prepared by the Executive Branch in the late nineteenth century, the materials quoted relating to treaty termination referred only to termination by Congress.[146]

This historical account presents difficulties for scholars who have attempted to defend a presidential power over treaty termination on originalist grounds, such as under the Vesting Clause Thesis (which hypothesizes that the vesting clause of Article II of the Constitution implicitly conveys to the President authority not otherwise listed in Article II).[147] There is no direct evidence that the Founders understood that the Constitution was granting the President a unilateral power of treaty termination. Moreover, to the extent that originalists credit historical practice, they typically place much more emphasis on early practice than

---

145. *See infra* subpart III(E).

146. *See, e.g.*, JOHN L. CADWALADER, DIGEST OF THE PUBLISHED OPINIONS OF THE ATTORNEYS-GENERAL, AND OF THE LEADING DECISIONS OF THE FEDERAL COURTS WITH REFERENCE TO INTERNATIONAL LAW, TREATIES, AND KINDRED SUBJECTS §§ 48–50, at 234 (rev. ed. 1877) (discussing the principles of treaty abrogation but making no mention of abrogation by the Executive Branch); 2 FRANCIS WHARTON, A DIGEST OF THE INTERNATIONAL LAW OF THE UNITED STATES § 137(a), at 58–65 (2d ed. 1887) (same); *see also* David A. Schnitzer, Note, *Into Justice Jackson's Twilight: A Constitutional and Historical Analysis of Treaty Termination*, 101 GEO. L.J. 243, 265–66 (2012) (surveying period digests' treatment of treaty abrogation).

147. *See supra* notes 30–33 and accompanying text.

modern practice, on the theory that it is closer in time to the founding and, thus, either is more likely to reflect founding intent or is a "liquidation" of issues unsettled at the founding.[148] Yet the first century of U.S. practice weighs strongly against a unilateral presidential power of treaty termination. If the Article II Vesting Clause conveyed to presidents the unilateral authority to terminate treaties, it is surprising that no one (with the possible exception of Alexander Hamilton) seemed to be aware of it for a hundred years.

III. Twentieth-Century Shift to Presidential Unilateralism

This Part describes the shift in U.S. practice during the twentieth century towards unilateral presidential termination of treaties. The accretion of claims and practice relating to this issue occurred over a long period, running from Congress's protectionist trade policy of the early twentieth century, to the U.S. rejection of the Versailles Treaty after World War I, to the onset of World War II and the related rise of the United States as a superpower. Although there was significant controversy surrounding the issue in connection with President Carter's announcement in 1978 that he was unilaterally terminating a mutual defense treaty with Taiwan, the Executive Branch prevailed in that dispute, and unilateral presidential termination of treaties has since become the norm. In addition to considering publicly available materials such as congressional hearings, official correspondence, and presidential proclamations, the description in this Part takes account of a number of internal memoranda prepared by the legal office for the State Department during the first half of the twentieth century, which have been retrieved from the State Department archives.[149]

A. Seeds of Change

The stirrings of a shift to presidential unilateralism can be seen in the early years of the twentieth century. In 1909, at the outset of the Taft Administration, the Solicitor for the State Department wrote an internal memorandum suggesting that it was constitutionally permissible for the President to act unilaterally in terminating a treaty. The memorandum stated that, although presidential action pursuant to a congressional

148. Some originalists accept that the Founders allowed certain unresolved constitutional issues to be worked out, or "liquidated," by early practice and decisions. *See, e.g.*, Caleb Nelson, *Originalism and Interpretive Conventions*, 70 U. CHI. L. REV. 519, 547–53 (2003). This liquidation idea was famously articulated by James Madison in *The Federalist Papers*. *See* THE FEDERALIST NO. 37, at 225 (James Madison) (Clinton Rossiter ed., 1961) (stating that the meaning of the Constitution, like that of all laws, would be "liquidated and ascertained by a series of particular discussions and adjudications" (emphasis omitted)).

149. Some of these memoranda have been partially excerpted in digests of practice published by the U.S. State Department. *See, e.g.*, HACKWORTH, *supra* note 75, § 509, at 319 (containing such an excerpt).

directive might be the "most effective and unquestionable method" for terminating a treaty, the President also had the option under U.S. law either of acting in conjunction with the Senate or through "notice given by the President upon his own initiative without either a resolution of the Senate or the joint resolution of the Congress."[150] In support of the last option, the memorandum noted that there had been one instance of unilateral presidential termination of a treaty, namely the 1899 termination of the provisions in the Swiss treaty.[151] The memorandum concluded that the choice of which method to use for terminating a treaty "would seem to depend either upon the importance of the international question or upon the preference of the Executive."[152] As discussed above, two years later the Taft Administration moved to terminate a treaty with Russia as a result of Russia's mistreatment of American Jews,[153] and it seems likely that this memorandum was prepared in connection with the Administration's initial consideration of that issue.[154]

A few years later, the Supreme Court seemed to suggest in dicta that the Executive Branch could decide whether to stop complying with a bilateral treaty in response to a breach by the other party. In *Charlton v. Kelly*,[155] the Court concluded that it was not improper for the Executive Branch to extradite a U.S. citizen to Italy pursuant to an extradition treaty between the two countries, notwithstanding the fact that Italy had declined to extradite its own citizens to the United States.[156] "The executive department having thus elected to waive any right to free itself from the obligation to deliver up its own citizens," the Court reasoned, "it is the plain duty of this court to recognize the obligation to surrender the appellant as one imposed by the treaty as the supreme law of the land and as affording authority for the warrant of extradition."[157] It is not clear how much should be read into such dicta, but it is worth recalling that there was the

150. Memorandum from James Brown Scott, Solicitor, U.S. Dep't of State, to President Wilson 1–2 (June 12, 1909) (on file with author).

151. *See id.* at 2.

152. *Id.* at 3.

153. *See supra* text accompanying notes 113–18.

154. The memorandum was prepared in June 1909. The American Jewish Committee was in communication with the Administration about the Russian issue during this time period, and members of the Committee met with Taft during the summer of 1909. *See* Naomi W. Cohen, *The Abrogation of the Russo-American Treaty of 1832*, 25 JEWISH SOC. STUD. 3, 9–10 (1963) (describing how Committee members Judge Sulzberger and Dr. Cyrus Adler met with Taft, the Secretary of State, and the American ambassador to Russia in the summer of 1909 and advocated for abrogation); Clifford L. Egan, *Pressure Groups, the Department of State, and the Abrogation of the Russian–American Treaty of 1832*, 115 PROC. AM. PHIL. SOC'Y 328, 329–30 (1971) (describing Taft's interaction with the American Jewish Committee in 1909 and 1910).

155. 229 U.S. 447 (1913).

156. *Id.* at 475–76.

157. *Id.* at 476.

nineteenth-century precedent of President Grant in effect suspending extradition until an issue of treaty compliance could be worked out.[158]

The topic of treaty termination arose again in 1919, during the debate in the Senate over whether to give its advice and consent to the Versailles Treaty, which, among other things, established the League of Nations. The League of Nations Covenant had a provision allowing any member to withdraw after two years "provided that all its international obligations and all its obligations under this Covenant shall have been fulfilled at the time of its withdrawal."[159] Senator Henry Cabot Lodge, the Republican leader in the Senate and Chairman of the Foreign Relations Committee, proposed attaching a reservation to the Senate's advice and consent providing that the United States could withdraw from the Covenant through enactment of a concurrent resolution by Congress.[160] This proposal generated substantial discussion.

Noting that a concurrent resolution does not require the agreement of the President, Senator Thomas moved to delete the concurrent resolution clause from the reservation on the ground that withdrawal from a treaty was "an executive and not a legislative function."[161] In the debate on the motion, Senator Jones asked "whether or not the President could give such notice [of termination] without authorization from Congress."[162] Senator Walsh replied, "I think not; clearly not. I cannot believe that anybody could entertain any serious doubt as to that."[163] In arguing in favor of the concurrent resolution clause, however, Senator Spencer contended that the President could unilaterally withdraw the United States from the treaty and thus the concurrent resolution clause was simply adding another option for U.S. withdrawal.[164] Numerous senators, however, either expressly disagreed with Spencer's premise or expressed skepticism about it.[165]

---

158. *See supra* notes 88–90 and accompanying text.

159. League of Nations Covenant art. 1, para. 3.

160. 58 CONG. REC. 8074 (1919). Lodge had significant concerns about the League of Nations Covenant, especially Article Ten, which involved a precommitment by the members to use military force in response to aggression. He led a group of Republicans that insisted that the Senate include a package of reservations with its advice and consent to the Covenant. Although there was majority support in the Senate for Lodge's proposed approach, neither his proposal nor a Democratic proposal to have the Senate give its advice and consent without the reservations was able to garner the required two-thirds vote. *See* JOHN MILTON COOPER, JR., BREAKING THE HEART OF THE WORLD: WOODROW WILSON AND THE FIGHT FOR THE LEAGUE OF NATIONS 234–375 (2001).

161. 58 CONG. REC. 8074 (1919).

162. *Id.* at 8076.

163. *Id.*

164. *Id.* at 8121–22.

165. *See, e.g.*, *id.* (Sen. Brandegee) (challenging Spencer's premise that the President could withdraw the United States from the League of Nations unilaterally without the consent of Congress); *id.* at 8122 (Sen. Poindexter) (asking Spencer how the President can unilaterally repeal treaties if they are the supreme law of the land); *id.* (Sen. Thomas) (expressing skepticism of

In one of the more extensive analyses of the issue, Senator Robinson explained:

> While the authorities on the subject are somewhat confusing, and while the Senate precedents, as in almost every disputed case, are somewhat conflicting, I believe that I can successfully maintain that the proper, the constitutional, the customary method of giving such notice as is contemplated in the reservation is through some action which contemplates the concurrence of the Executive and the two Houses of Congress.[166]

Robinson subsequently noted, however, that "[t]here may be cases . . . where the Executive alone[] has the authority to terminate a treaty, but these cases are exceptional."[167]

Nevertheless, the views on this question were mixed. Senator Lenroot later pointed out that Westel Willoughby's 1910 constitutional law treatise stated that the President had a unilateral withdrawal power.[168] This seemed to cause Senator Walsh to retreat to some extent from his earlier statement to the contrary, while noting that he would "want to examine the question with very great care before [he] could accept any such doctrine [as argued by Willoughby]."[169] Senator King then asked Senator Lenroot whether, if Willoughby were correct, there was any way that the Senate could protect itself against a President unilaterally terminating a treaty that the Senate had agreed to, and Lenroot responded that the courts would likely treat the matter as a political question, so the principal tool of Congress would probably be impeachment.[170] Lenroot further noted, in response to another question from King, that the Senate could prospectively limit the President's termination authority by including a provision to that effect in its advice and consent to a treaty.[171] There was also some discussion of whether Taft's termination in 1911 of the treaty with Russia was precedent for a unilateral termination authority, and Lenroot expressed the view that it was.[172] Ultimately, the proposed amendment to the reservation was

---

Spencer's interpretation of the League of Nations article requiring "member" to be equivalent to "President" and that the President has unilateral authority to repeal treaties as supreme laws of the land under the treaty-making power); *id.* at 8124 (Sen. Robinson) (expressing doubt that the Executive Branch can terminate a treaty without involving Congress unless perhaps the treaty relates to functions exclusively within the Executive's power).

166. *Id.* at 8124.

167. *Id.* at 8125.

168. *Id.* at 8129, 8132.

169. *Id.* at 8131; *see also id.* at 8130 (responding that he did "not undertake to say . . . whether the actual concurrence" of the President and Congress for withdrawal is "essential").

170. *Id.* at 8132.

171. *Id.*

172. *Id.*

rejected. Shortly thereafter, the Senate voted against giving its advice and consent to the treaty even with the reservations.[173]

The termination issue does not appear to have been settled. In the next session of Congress, the Senate reconsidered its rejection of the treaty.[174] During that discussion, Senator Lodge moved to amend his proposed reservation to make clear that the United States could withdraw either by two houses of Congress or by presidential action.[175] Lodge explained that the usual method of terminating treaties had been pursuant to congressional direction or concurrence, but he noted that there were two instances in which presidents had acted unilaterally—McKinley in 1889 and Taft in 1911.[176] When asked whether the President would have the authority to withdraw the United States from the Versailles Treaty even if this authority were not specified in the reservation, Lodge replied (somewhat awkwardly) that, "I think it is at least doubtful whether the President has not the power to do that."[177] The ensuing debate on his motion, however, concerned whether the Senate could delegate termination authority to the President, not whether he had such authority independently.[178] In any event, Lodge's amendment was rejected,[179] and the original language of his proposed reservation was retained.[180] The Senate then proceeded to reject the Versailles Treaty a second time.[181]

Despite these various discussions of treaty termination, no President actually terminated a treaty unilaterally during the twentieth century until 1927. In that year, the Coolidge Administration withdrew the United States from a smuggling convention with Mexico without authorization or subsequent approval from Congress or the Senate.[182] The administration explained that the United States had no commercial treaty with Mexico and that

> it is not deemed advisable to continue in effect an arrangement which might in certain contingencies bind the United States to cooperation for the enforcement of laws or decrees relating to the importation of commodities of all sorts into another country with which this

173. *See id.* at 8803.
174. 59 CONG. REC. 3229 (1920).
175. *Id.* at 3229–30.
176. *Id.*
177. *Id.* at 3230.
178. *Id.* at 3230–32.
179. *Id.* at 3242.
180. *Id.*
181. *Id.* at 4599.
182. ADLER, *supra* note 75, at 183–84.

> Government has no arrangement, by treaty or otherwise, safeguarding American commerce against possible discrimination.[183]

This action was taken after extensive concerns had been raised in Congress about Mexico's confiscation of American property.[184]

Unilateral presidential terminations subsequently became more common in the administration of Franklin D. Roosevelt, although some of these terminations, like McKinley's 1899 termination of provisions in the Swiss treaty, were because of potential conflicts with trade legislation. In 1933, the Executive Branch withdrew the United States from a convention abolishing import and export restrictions, without authorization or subsequent approval from Congress or the Senate, because of (among other things) alleged conflicts between the convention and the new National Industrial Recovery Act.[185] Also in 1933, Roosevelt unilaterally announced termination of an extradition treaty with Greece because of its purported breach of the treaty after Greece had refused to extradite Samuel Insull, a billionaire tycoon who was accused of financial misdealings.[186] After Greece forced Insull to leave the country and a protocol to the extradition

---

183. Telegram from Frank B. Kellogg, U.S. Sec'y of State, to Ambassador Sheffield (Mar. 21, 1927), *in* 3 PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES, 1927, at 230, 230–31 (1942). The Executive Branch has sometimes claimed that President Wilson unilaterally terminated a treaty with Belgium in 1920, but this is incorrect. Pursuant to Congress's directive in the Seaman's Act, *see supra* note 93 and accompanying text, Wilson had given notice to Belgium in 1916 that the United States was terminating certain provisions in a treaty concerning the Congo. *See* Letter from Robert Lansing, U.S. Sec'y of State, to Baron Beyens, Belg. Minister of Foreign Affairs (Nov. 11, 1916), *in* PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES 34, 34–35 (1925) (notifying the Government of Belgium that pursuant to an Act of Congress, the United States government was terminating certain portions of a previously-agreed-upon treaty). Belgium responded by saying that it preferred simply to terminate the entire treaty, and it asked the United States to formally acknowledge this denunciation. Letter from Baron Beyens, Belg. Minister of Foreign Affairs, to Robert Lansing, U.S. Sec'y of State (Dec. 31, 1916), *in* PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES, *supra*, at 35, 36. Eventually, in 1920, the United States "acknowledge[d]" Belgium's notice of termination. Letter from Norman H. Davis, U.S. Undersecretary of State, to Brand Whitlock, U.S. Ambassador to Belg. (Nov. 19, 1920), *in* 1 PAPERS RELATING TO THE FOREIGN RELATIONS OF THE UNITED STATES, 1920, at 207, 207–09 (1935). Noting that the treaty did not contain any clause specifying the amount of notice required for withdrawal, the United States said that it would assume that the Belgian government wished the treaty to have terminated one year from the time of its notice of termination, since that was a customary period of notice. *Id.* at 209. The treaty was therefore terminated by Belgium, not the United States, and the U.S. action that prompted Belgium to terminate the treaty was directed by Congress.

184. *See, e.g.*, 68 CONG. REC. 4591 (1927) (presenting a letter from an unknown source in Mexico explaining the theft of American property by a newly radical Mexican government).

185. *See* Press Release, U.S. Dep't of State, Withdrawal of United States from International Convention for the Abolition of Import and Export Prohibitions and Restrictions (July 5, 1933), *reprinted in* DEP'T OF STATE, PRESS RELEASES, JULY 1–DECEMBER 30, 1933 18, 18; *see also* 1 FOREIGN RELATIONS OF THE UNITED STATES: DIPLOMATIC PAPERS, 1933, at 783–86 (1950) (collecting telegram exchanges between U.S. officials that document the considerations surrounding the decision to withdraw from the convention).

186. *See* 2 FOREIGN RELATIONS OF THE UNITED STATES: DIPLOMATIC PAPERS, 1933, at 552–69 (1949) (collecting telegram exchanges between U.S. officials that document the Insull affair).

treaty was subsequently negotiated, however, Roosevelt withdrew the notice.[187]

In 1936, the Executive Branch withdrew the United States from a commercial treaty with Italy.[188] The State Department wrote a memorandum advising President Roosevelt that this unilateral action was constitutional.[189] While acknowledging that "[t]he question as to the authority of the Executive to terminate treaties independently of the Congress or of the Senate is in a somewhat confused state," the memorandum maintained that "[n]o settled rule or procedure has been followed."[190] It also noted that there was a potential conflict between the treaty with Italy and a 1934 trade statute and that, if the treaty were not terminated, the President could "be placed in the position of having to choose between the execution of the act and observance of the treaty."[191] The memorandum observed that this situation was "closely analogous" to the termination of provisions in the Swiss treaty in 1899, and it said that the 1899 "precedent" was confirmed by the U.S. withdrawal from the import–export treaty in 1933.[192]

B. Establishing a Pattern

Because many of the early-twentieth-century presidential terminations were based on potential conflicts with statutes, these actions would not necessarily have been understood as fully unilateral in nature. By the late 1930s, however, the Executive Branch was increasingly asserting a purely unilateral authority. In 1939, the Roosevelt Administration announced that the United States was terminating a commercial treaty with Japan, after resolutions had been introduced in both houses of Congress supporting withdrawal.[193] In connection with this decision, the State Department argued that the President had unilateral termination authority, this time relying on the "general spirit" of the Supreme Court's 1936 decision in

187. *See* ADLER, *supra* note 75, at 184–85.

188. *See* 2 FOREIGN RELATIONS OF THE UNITED STATES: DIPLOMATIC PAPERS, 1936, at 356–59 (1954) (collecting exchanges between U.S. officials that document this withdrawal).

189. Memorandum from R. Walton Moore, Acting U.S. Sec'y of State, to President Roosevelt 5 (Nov. 9, 1936) (on file with author).

190. *Id.* at 2–3.

191. *Id.* at 3–4.

192. *Id.* at 4–5. The State Department Legal Adviser had prepared a memorandum earlier that year on abrogation of treaties. That memorandum contends that, regardless of whether the President has a general power to terminate treaties unilaterally, it seems "that little doubt could arise when, as in the case of the Seaman's Act, he is called upon to terminate provisions of treaties inconsistent with an Act of Congress and when failure to do so would place this Government in the position of failing to observe its treaty obligations." HACKWORTH, *supra* note 75, at 327–28 (citing the Legal Adviser's memorandum).

193. H.R. Res. 264, 76th Cong. (1939); S. Res. 166, 76th Cong. (1939); Letter from Cordell Hull, U.S. Sec'y of State, to Ambassador Horinouchi (July 26, 1939), *in* 3 FOREIGN RELATIONS OF THE UNITED STATES: DIPLOMATIC PAPERS, 1939, at 558, 558–59 (1955).

United States v. Curtiss-Wright Export Corp.,[194] which had referred to the "delicate, plenary and exclusive power of the President as the sole organ of the federal government in the field of international relations."[195] The State Department reasoned that "the power to denounce a treaty inheres in the President of the United States in his capacity as Chief Executive of a sovereign state," and it further contended that the President had "full control over the foreign relations of the nation, except as specifically limited by the Constitution."[196]

The 1930s also saw a political transformation in the United States, with Roosevelt having landslide victories in the presidential elections of 1932 and 1936 and the Democrats coming to dominate both houses of Congress.[197] In addition, the national security environment was changing significantly in this period, with increasing aggression by Adolf Hitler in Germany, the invasion of China by Japan, and eventually the start of World War II. This environment was conducive to broader claims of executive authority. A year after *Curtiss-Wright*, the Supreme Court decided *United States v. Belmont*,[198] in which it held that a sole executive agreement entered into by President Roosevelt as part of his recognition of the Soviet Union preempted state law.[199] Shortly thereafter, the Court began giving absolute deference to Executive Branch determinations relating to foreign sovereign immunity.[200]

National security soon became directly relevant to the issue of treaty termination and suspension. In 1939, for example, President Roosevelt suspended the London Naval Treaty (which limited naval armaments) because of the changed circumstances created by the war in Europe.[201] Two

194. 299 U.S. 304 (1936).

195. *Id.* at 320. The Court in *Curtiss-Wright*, in an opinion authored by Justice Sutherland, upheld a delegation of authority from Congress to the President to criminalize arms sales to countries involved in a conflict in Latin America. *Id.* at 329–33.

196. HACKWORTH, *supra* note 75, at 331–32 (excerpting from a State Department memorandum).

197. 2 BRUCE ACKERMAN, WE THE PEOPLE: TRANSFORMATIONS 286, 311 (1998).

198. 301 U.S. 324 (1937).

199. *Id.* at 327. The opinion in *Belmont*, like the opinion in *Curtiss-Wright*, was authored by Justice Sutherland. *Id.* at 325; *supra* note 195.

200. *See* Republic of Mex. v. Hoffman, 324 U.S. 30, 35–36 (1945) (deferring to the State Department in deciding foreign-government immunity); *Ex parte* Republic of Peru, 318 U.S. 578, 586–87 (1943) (holding that a ship owned by Peru, seized in the course of private litigation, should be released because the State Department declared Peru immune from suit); Compania Espanola de Navegacion Maritima v. The Navemar, 303 U.S. 68, 74 (1938) (deferring to the Executive Branch in determining whether foreign governments are immune from suit); *see also* ARTHUR M. SCHLESINGER, JR., THE IMPERIAL PRESIDENCY 110–26 (Mariner Books 2004) (1973) (describing Roosevelt's increasingly aggressive approach to the exercise of foreign-affairs authority); G. Edward White, *The Transformation of the Constitutional Regime of Foreign Relations*, 85 VA. L. REV. 1, 8 (1999) (referring to "the triumph of executive discretion in the constitutional regime of foreign relations between 1933 and the close of the Second World War").

201. *See Armament Reduction*, 1 DEP'T ST. BULL. 354, 354 (1939).

years later, he suspended, for the duration of the war, the International Load Lines Convention (which regulated ocean shipping) after his Attorney General, Francis Biddle, advised him that "[t]he convention may be declared inoperative or suspended by the President."[202] Biddle also noted, however, that, since "[i]t is not proposed that the United States denounce the convention under [the unilateral withdrawal clause in the treaty], nor that it be otherwise abrogated. . . . [A]ction by the Senate or by the Congress is not required."[203] The opinion thus seemed to suggest that a full termination of a treaty, as opposed to a suspension, would require legislative action. Nevertheless, the Roosevelt Administration terminated another treaty unilaterally in 1944—a protocol relating to a Latin American trademark treaty—citing the treaty's general ineffectiveness.[204]

The 1950s saw several additional unilateral presidential terminations, usually in low-profile situations that did not generate much attention, such as the Truman Administration's withdrawal of the United States from a whaling convention[205] and the Eisenhower Administration's termination of both a Convention on Uniformity of Nomenclature for the Classification of Merchandise and a Treaty of Friendship, Commerce, and Navigation with El Salvador.[206] Although not solely a U.S. termination, the Eisenhower Administration also entered into a sole executive agreement in 1958 with Morocco to end a treaty relating to the management of a lighthouse in that country.[207]

A 1958 memorandum from the State Department's Deputy Assistant Legal Adviser for Treaty Affairs, William Whittington, noted that, although "matters of policy or special circumstances may make it appear to be advisable or necessary to obtain the concurrence or support of the Congress or the Senate," in practice treaties have been terminated in a variety of ways, including through unilateral presidential action.[208] The memorandum also asserted that, at least for a self-executing treaty containing a unilateral withdrawal clause, "it is now generally considered that . . . it is proper for the Executive acting alone to take the action necessary to terminate or denounce the treaty."[209] Attached to the memorandum were appendices

202. Int'l Load Line Convention, 40 Op. Att'y Gen. 119, 123 (1941).

203. *Id.*

204. *Treaty Information*, 11 DEP'T ST. BULL. 442, 442 (1944).

205. *Treaty Termination: Hearings Before the S. Comm. on Foreign Relations*, 96th Cong. 83 (1979) [hereinafter *Treaty Termination Hearings*].

206. *See Treaty Information*, 32 DEP'T ST. BULL. 906, 906 (1955) (noting the withdrawal from the nomenclature convention); *Treaty Information*, 38 DEP'T ST. BULL. 238, 238 (1958) (noting the termination of the treaty with El Salvador).

207. Cape Spartel Light: Transfer of Management to Morocco; Termination of Convention of May 31, 1865, Mar. 31, 1958, 9 U.S.T. 527, 532.

208. *See* Whittington Memorandum, *supra* note 25, at 5–6.

209. *Id.* at 5.

listing the various treaty terminations in U.S. history and how they were carried out.[210]

The practice of unilateral terminations continued during the 1960s. In 1962, the Kennedy Administration terminated a commercial treaty with Cuba as part of the United States' embargo policy following the Cuban revolution.[211] In 1965, the Johnson Administration gave notice that the United States was withdrawing from the Warsaw Convention that governs liability for international air carriers,[212] but retracted the withdrawal shortly before the notice period expired.[213]

There were still occasions in this period, however, in which the United States terminated treaties pursuant to congressional directive. In 1951, for example, President Truman terminated commercial treaties with the Soviet Union and various Eastern European countries pursuant to a directive in the Trade Agreements Extension Act of 1951.[214] In 1976, the Ford Administration withdrew the United States from several treaties relating to fishing pursuant to a directive in the 1976 Fishery Conservation and Management Act.[215]

C. *Termination of the Taiwan Treaty*

During the 1970s, the United States began to pursue closer relations with the People's Republic of China (PRC). As one of the conditions to a normalization of relations between the two countries, the PRC insisted that the United States terminate its 1954 mutual defense treaty with Taiwan. Anticipating a change in Executive Branch policy concerning Taiwan,

---

210. *Id.* at 7.

211. *See* Convention for Commercial Relations with Cuba, U.S.-Cuba, Dec. 11, 1902, 33 Stat. 2136 (establishing good commercial relations between the United States and Cuba); Proclamation No. 3447, Embargo on All Trade with Cuba, 27 Fed. Reg. 1085 (Feb. 7, 1962), *reprinted in* 76 Stat. 1446 (1962) (terminating commercial relations with Cuba).

212. *See Treaty Information*, 53 DEP'T ST. BULL. 923, 924 (1965) (relating the United States' denunciation of the Warsaw Convention and its attempt to negotiate revised terms).

213. *See* Press Release, Dep't of State, United States Government Action Concerning the Warsaw Convention (May 5, 1966), *reprinted in The Warsaw Convention—Recent Developments and the Withdrawal of the United States Denunciation*, 32 J. AIR L. & COM. 243, 245–46 (1966) (discussing the United States' notification of termination and its subsequent withdrawal of that notice). For commentary suggesting that the proposed unilateral withdrawal would have been unconstitutional, see John H. Riggs, Jr., *Termination of Treaties by the Executive Without Congressional Approval: The Case of the Warsaw Convention*, 32 J. AIR L. & COM. 526, 527–28 (1966), and Comment, *Presidential Amendment and Termination of Treaties: The Case of the Warsaw Convention*, 34 U. CHI. L. REV. 580, 581–82 (1967).

214. *See* Trade Agreements Extension Act of 1951, ch. 141, 65 Stat. 72 (codified as amended at 19 U.S.C. § 1351 (2012)) (giving the President authority to take action in order to bring trade agreements "into conformity" with the Act); Proclamation No. 2949, 3 C.F.R. § 134 (1949–1953), *reprinted in* 65 Stat. c44 (1951).

215. *See* Fishery Conservation and Management Act of 1976, Pub. L. No. 94-265, § 202(b), 90 Stat. 331, 340–41 (giving the Executive Branch power to renegotiate international fishing treaties).

Congress in 1978 enacted (and the President signed) the International Security Assistance Act, which, among other things, expressed "the sense of the Congress that there should be prior consultation between the Congress and the executive branch on any proposed policy changes affecting the continuation in force of the Mutual Defense Treaty [with Taiwan] of 1954."[216] In December 1978, President Carter announced that the United States would recognize the PRC as the sole government of China and would terminate the Taiwan treaty pursuant to the unilateral withdrawal clause in the treaty (which required one year's notice).[217]

In a memorandum advising the President that he had the constitutional authority to terminate the treaty, the State Department Legal Adviser relied heavily on historical practice.[218] The memorandum cited twelve instances in which presidents had purportedly terminated treaties unilaterally, and it included an extensive appendix entitled "History of Treaty Terminations by the United States."[219] The memorandum concluded that "[w]hile treaty termination may be and sometimes has been undertaken by the President following Congressional or Senate action, such action is not legally necessary."[220]

Carter's action prompted substantial debate in the Senate. Several resolutions were introduced in early January 1979, including a resolution sponsored by Senator Harry Byrd, Jr., that provided that it was "the sense of the Senate that approval of the U.S. Senate is required to terminate any mutual defense treaty between the United States and another nation."[221] The Foreign Relations Committee held three days of hearings on this

---

216. International Security Assistance Act of 1978, Pub. L. No. 95-384, § 26(b), 92 Stat. 730, 746; *see also* Jimmy Carter, International Security Assistance Act of 1978: Statement on Signing S. 3075 into Law (Sept. 26, 1978), *in* 2 PUBLIC PAPERS OF THE PRESIDENTS OF THE UNITED STATES: JIMMY CARTER, 1978, at 1636, 1636 (1978) [hereinafter PUBLIC PAPERS OF JIMMY CARTER].

217. *See* President Jimmy Carter, Address to the Nation: Diplomatic Relations Between the United States and the People's Republic of China (Dec. 15, 1978), *in* PUBLIC PAPERS OF JIMMY CARTER, *supra* note 216, at 2264, 2264–65 (recognizing the PRC as the sole government of China and announcing his intention to maintain relations with Taiwan "through nongovernmental means").

218. *See* Hansell Memorandum, *supra* note 121, at 397–99 (listing past presidential treaty terminations). For criticism of the State Department's description of the historical practice, see, for example, J. Terry Emerson, *The Legislative Role in Treaty Abrogation*, 5 J. LEGIS. 46, 77–78 (1978); David J. Scheffer, Comment, *The Law of Treaty Termination as Applied to the United States De-Recognition of the Republic of China*, 19 HARV. INT'L L.J. 931, 979 (1978); and Jonathan York Thomas, Article, *The Abuse of History: A Refutation of the State Department Analysis of Alleged Instances of Independent Presidential Treaty Termination*, 6 YALE STUD. WORLD PUB. ORD. 27, 30 (1979).

219. Hansell Memorandum, *supra* note 121, at 397–98, 400.

220. *Id.* at 395.

221. S. Res. 15, 96th Cong., 125 CONG. REC. 475 (1979).

resolution.[222] The hearings included testimony and prepared statements from a variety of witnesses, including a number of scholars. Scholars such as Arthur Bestor, Thomas Franck, and Michael Reisman testified or submitted statements in favor of senatorial or congressional participation in treaty termination.[223] Other scholars, such as Abram Chayes, Andreas Lowenfeld, and John Norton Moore, testified in favor of a unilateral presidential power of termination.[224]

The Foreign Relations Committee rejected the approach of the Byrd Resolution and reported out instead a resolution that would have recognized fourteen grounds for justifying unilateral presidential action to terminate treaty obligations, including the existence of a termination clause.[225] After it reached the Senate floor, however, the Senate (on a vote of 59–35) substituted for its consideration the original Byrd Resolution, after Byrd's motion for substitution was supported by a number of Senators who expressed the view that the President should not have unilateral power over treaty termination.[226] But the Senate never actually voted on this resolution.[227]

In the meantime, former Senator Barry Goldwater, along with a group of eight current senators and sixteen current members of the House of Representatives, filed a lawsuit in the federal district court in D.C. seeking declaratory and injunctive relief to prevent the termination of the Taiwan treaty.[228] The district court initially dismissed the suit without prejudice, reasoning that the legislators would not have standing until there was action taken on the resolutions pending in the Senate.[229] After the substitution of the Byrd Resolution in the Senate, the plaintiffs argued that they now had standing, and the court agreed, noting that the action on the Resolution was "evidence [of] at least some congressional determination to participate in the process whereby a mutual defense treaty is terminated, and clearly falls short of approving the President's termination effort."[230] The court proceeded to reach the merits and concluded that "the President's notice of termination must receive the approval of two-thirds of the United States Senate or a majority of both houses of Congress for it to be effective under

222. *Treaty Termination Hearings*, *supra* note 205, at iii (indicating that the hearings were held from April 9 to 11, 1979).

223. *Id.* at 25–32, 223–74, 387–96.

224. *Id.* at 306–12, 396–425, 426–43.

225. 125 CONG. REC. 13,685 (1979). For the discussion in the Senate of this issue, see *id.* at 13,672–710.

226. *Id.* at 13,695–96.

227. *Id.* at 13,710.

228. Goldwater v. Carter, No. 78-2412, 1979 U.S. Dist. LEXIS 11893, at *1 (D.D.C. June 6, 1979).

229. *Id.* at *16–17.

230. Goldwater v. Carter, 481 F. Supp. 949, 954 (D.D.C. 1979).

our Constitution to terminate the [Taiwan treaty]."[231] In addition to textual and structural considerations, the court relied on historical practice, reasoning that "[t]he predominate United States' practice in terminating treaties, including those containing notice provisions, has involved mutual action by the executive and legislative branches."[232]

The D.C. Circuit reversed, holding that President Carter's termination of the treaty was constitutional.[233] In addition to emphasizing the President's role as "sole organ" in foreign relations, the court noted that the historical practice was varied and that there was no past instance in which "a treaty [has] been continued in force over the opposition of the President."[234] The court also emphasized that Carter had acted pursuant to a unilateral withdrawal clause in the treaty and reasoned that "the President's authority . . . is at its zenith when the Senate has consented to a treaty that expressly provides for termination on one year's notice, and the President's action is the giving of notice of termination."[235] In other words, the court was claiming that Carter was acting within the highest category of presidential authority laid out by Justice Jackson in his concurrence in *Youngstown*.[236]

Judge MacKinnon issued a lengthy dissent, focused especially on the history of treaty terminations. He contended that "[c]ongressional participation in termination has been the overwhelming historical practice."[237] As for the instances of unilateral presidential termination, MacKinnon reasoned:

> It is almost farcical for appellant to contend that the President, acting alone, has absolute power to terminate a major United States defense treaty, and by the same token hereafter any defense treaty, because a few earlier Presidents withdrew financial support of a treaty bureau because of non-filing of trademarks by El Salvador, Honduras, Paraguay, et al., and terminated several violated treaties, or terminated treaties relating to a light house museum in Morocco, nomenclature in economic reports, smuggling with a country with

231. *Id.* at 965.

232. *Id.* at 960–64.

233. Goldwater v. Carter, 617 F.2d 697, 699 (D.C. Cir. 1979) (per curiam).

234. *Id.* at 706–07.

235. *Id.* at 708.

236. *See* Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 635 (1952) (Jackson, J., concurring) ("When the President acts pursuant to an express or implied authorization of Congress, his authority is at its maximum, for it includes all that he possesses in his own right plus all that Congress can delegate.").

237. *Goldwater*, 617 F.2d at 723 (MacKinnon, J., dissenting).

> whom we had no commercial treaty, or with respect to which notices of termination had been given and then withdrawn.[238]

MacKinnon also argued that the majority's suggestion that the treaty itself authorized Carter to engage in unilateral termination was a "deceptive misstatement" since "the President is *not* named *in* the Treaty to give notice of termination" and "[t]he sole issue in this case is who can act for the United States; that issue is not determined by the Treaty but by the Constitution of the United States."[239]

Without hearing oral argument, the Supreme Court vacated the D.C. Circuit's decision and remanded with instructions to dismiss the case.[240] Four Justices reasoned that the case presented a nonjusticiable political question.[241] Providing a fifth vote for nonjusticiability, Justice Powell reasoned that the case was not politically ripe, given that the Senate had never voted on a resolution to disapprove the termination. "If the Congress chooses not to confront the President," said Powell, "it is not our task to do so."[242] The controversy effectively ended with this dismissal.[243]

D. Subsequent Treaty Terminations

In the years since the controversy over the termination of the Taiwan treaty, the United States has terminated dozens of treaties, and almost all of these terminations have been accomplished by unilateral presidential action. To take one example, the Reagan Administration gave notice in 1985 of its termination of a Treaty of Friendship, Commerce, and Navigation with Nicaragua, and the treaty terminated the following year.[244] In 2002, the

---

238. *Id.* at 733–34 (emphasis omitted). For additional criticism of the D.C. Circuit's reasoning, see generally Raoul Berger, *The President's Unilateral Termination of the Taiwan Treaty*, 75 NW. U. L. REV. 577 (1980).

239. *Goldwater*, 617 F.2d at 737 (MacKinnon, J., dissenting).

240. Goldwater v. Carter, 444 U.S. 996, 996 (1979).

241. *Id.* at 1002 (Rehnquist, J, concurring) (opinion joined by Justices Stewart and Stevens and Chief Justice Burger).

242. *Id.* at 998 (Powell, J., concurring). Justice Brennan reasoned that the termination was lawful because it was "a necessary incident" to President Carter's recognition of mainland China, which fell within his constitutional authority. *Id.* at 1007 (Brennan, J., dissenting). Justices Blackmun and White wanted to hold oral argument before making a decision. *Id.* at 1006 (Blackmun & White, JJ., dissenting in part).

243. For additional discussion of the termination controversy, see generally VICTORIA MARIE KRAFT, THE U.S. CONSTITUTION AND FOREIGN POLICY: TERMINATING THE TAIWAN TREATY (1991). In the mid-1980s, Goldwater introduced a resolution that would have provided that it was the sense of the Senate that, unless otherwise provided in a treaty, termination required either the advice and consent of the Senate or congressional approval. S. Res. 40, 99th Cong., 131 CONG. REC. 678 (1985). But the Senate never voted on the resolution. *See id.* at 679–80.

244. *Economic Sanctions Against Nicaragua*, 85 DEP'T ST. BULL. 74, 74–75 (1985). A federal district court subsequently applied the political question doctrine to dismiss a challenge to this termination. *See* Beacon Prods. Corp. v. Reagan, 633 F. Supp. 1191, 1198–99 (D. Mass. 1986) ("[A] challenge to the President's power vis-a-vis treaty termination raise[s] a nonjusticiable political question."). Not all the terminations were unilateral. In 1986, Congress

State Department Legal Adviser's Office listed twenty-three bilateral treaties and seven multilateral treaties that had been terminated by presidential action since termination of the Taiwan treaty.[245] Since then, the Bush Administration terminated two treaties: a protocol to a consular convention in 2005[246] and a tax treaty with Sweden in 2007.[247]

Most of these terminations do not appear to have generated controversy. An exception is President George W. Bush's announcement in 2002 that he was withdrawing the United States from the Anti-Ballistic Missile (ABM) Treaty with Russia. In an Op-Ed article, the prominent constitutional law scholar Bruce Ackerman contended that Bush was acting unconstitutionally and asked rhetorically, "If President Bush is allowed to terminate the ABM treaty, what is to stop future presidents from unilaterally taking America out of NATO or the United Nations?"[248] As noted earlier, thirty-two members of Congress brought suit challenging the constitutionality of the termination of the ABM Treaty, but the suit was dismissed for lack of standing and under the political question doctrine.[249] The Justice Department's Office of Legal Counsel (OLC) issued a memorandum concluding that the President had the authority to suspend or terminate the treaty.[250] The memorandum relies on textual and structural

---

directed President Reagan to terminate a tax treaty and an air services treaty with South Africa as part of the Anti-Apartheid Act (which was enacted over Reagan's veto), and he did so. *See* Comprehensive Anti-Apartheid Act of 1986, Pub. L. No. 99-440, 100 Stat. 1086, 1100, 1104 (directing the President to terminate the treaties); *Current Actions*, 86 DEP'T ST. BULL. 84, 87 (1986) (indicating that the treaties had been terminated).

245. OFFICE OF LEGAL ADVISER, U.S. DEP'T OF STATE, 2002 DIGEST OF UNITED STATES PRACTICE IN INTERNATIONAL LAW 2002, at 202–06 (Sally J. Cummins & David P. Stewart eds., 2002).

246. Frederic L. Kirgis, *President Bush's Determination Regarding Mexican Nationals and Consular Convention Rights*, ASIL INSIGHTS add. (Mar. 2005), http://web.archive.org/web/20120716203621/http://www.asil.org/insights050309a.cfm.

247. Press Release, U.S. Dep't of the Treasury, United States Terminates Estate and Gift Tax Treaty with Sweden (June 15, 2007), *available at* http://www.treasury.gov/press-center/press-releases/Pages/hp463.aspx.

248. Bruce Ackerman, Op-Ed., *Treaties Don't Belong to Presidents Alone*, N.Y. TIMES, Aug. 29, 2001, http://www.nytimes.com/2001/08/29/opinion/treaties-don-t-belong-to-presidents-alone.html. In a very partial description of the historical practice, Ackerman mentioned Congress's 1798 termination of the French treaties and President Polk's solicitation of congressional authorization to terminate the Oregon Territory Treaty in 1846, and then asserted that "[t]he big change occurred in 1978, when Jimmy Carter unilaterally terminated our mutual defense treaty with Taiwan." *Id.*

249. Kucinich v. Bush, 236 F. Supp. 2d 1, 2 (D.D.C. 2002). Some members of the House of Representatives proposed a resolution that would have opposed termination of the ABM Treaty on policy grounds, but it did not take a position on the constitutionality of the termination, and the House never voted on it. *See* H.R. Res. 313, 107th Cong. (2001) (asserting that termination of the ABM Treaty could, among other things, "be perceived by other nations as a threat" and "weaken ties with traditional allies"); *see also* 147 CONG. REC. 25,917 (2001) (introducing H.R. Res. 313 but not voting on it).

250. Yoo & Delahunty Memorandum, *supra* note 52, at 9. OLC later disavowed this and another opinion relating to the suspension of treaty obligations, in part because it found

arguments such as the Vesting Clause Thesis and the President's role as the "sole organ" in foreign relations, as well as on historical practice.[251] Invoking the historical gloss concept, the memorandum reasons that "[t]he executive branch has long held the view that the President has the constitutional authority to terminate treaties unilaterally, and the legislative branch seems for the most part to have acquiesced in it."[252] While acknowledging that Congress and the Senate have sometimes been involved in treaty terminations, the memorandum contends that "[t]hese examples represent the workings of practical politics, rather than acquiescence in a constitutional régime."[253] Despite complaints by select members of Congress, there was no formal effort by Congress as a body to oppose the termination of the ABM Treaty,[254] and Congress ultimately approved funding for Bush's missile defense plan.[255]

E. Shift in Scholarly Commentary

As late as the early twentieth century, most commentators took the position that the President needed either senatorial or congressional approval to terminate a treaty. Charles Butler's highly regarded treatise on the U.S. treaty-making power, published in 1902, noted that treaties could be abrogated "by Congressional action in several different methods" and did not seem to contemplate termination by unilateral presidential action.[256] Similarly, the prominent constitutional law scholar Edward Corwin, in his 1917 book, *The President's Control of Foreign Relations*, stated: "All in

---

unconvincing the reasoning in the opinions suggesting that the President could suspend a treaty even when such suspension was not permissible under international law. *See* Memorandum of Steven G. Bradbury, Principal Deputy Assistant Att'y Gen., for the Files, Status of Certain OLC Opinions Issued in the Aftermath of the Terrorist Attacks of September 11, 2001, at 8–9 (Jan. 15, 2009), *available at* http://www.gwu.edu/~nsarchiv/torturingdemocracy/documents/20090115.pdf (indicating that the two opinions should not be relied upon "to the extent they suggest[] that the President has unlimited authority to suspend a treaty beyond the circumstances traditionally recognized"). OLC noted, however, that a 2007 opinion, which it was not disavowing, had observed that presidents have traditionally exercised the power to suspend treaties unilaterally "where suspension was authorized by the terms of the treaty or under recognized principles of international law." *Id.* at 9.

251. Yoo & Delahunty Memorandum, *supra* note 52, at 3–5, 13.

252. *Id.* at 9.

253. *Id.* at 14.

254. *See* DAVID M. ACKERMAN, CONG. RESEARCH SERV., RS21088, WITHDRAWAL FROM THE ABM TREATY: LEGAL CONSIDERATIONS 6 (2002) (noting that "the ABM Treaty has been terminated by President Bush in accordance with the terms of the treaty, and neither Congress nor the courts have acted to forestall or overturn that action"). Senator Kyl spoke on the floor of the Senate *in favor* of President Bush's authority to terminate the ABM Treaty. *See* 148 CONG. REC. 4536 (2002) (Sen. Kyl) (arguing that the text and structure of the Constitution, the intent of the Framers, and Supreme Court precedent all established executive authority to terminate treaties).

255. *See* Paul Richter, *Senate GOP Wins Funding Battle for Missile Defense*, L.A. TIMES, June 27, 2002, http://articles.latimes.com/2002/jun/27/nation/na-missile27.

256. 2 CHARLES HENRY BUTLER, THE TREATY-MAKING POWER OF THE UNITED STATES § 384, at 129 (1902).

all, it appears that legislative precedent, which moreover is generally supported by the attitude of the Executive, sanctions the proposition that the power of terminating the international compacts to which the United States is party belongs, as a prerogative of sovereignty, to Congress alone."[257]

Quincy Wright similarly supported legislative involvement in treaty termination in his important 1922 treatise on foreign relations law, although his discussion is somewhat more equivocal than Corwin's, stating that the President "ought not to [terminate] without consent either of Congress or of the Senate, except in extraordinary circumstances."[258] Also writing in 1922 (shortly before becoming Solicitor of the State Department), Charles Cheney Hyde noted in his treatise on international law that "[i]n behalf of the United States, notice of termination is given by the President, commonly in pursuance of a joint resolution of the Congress; and it has followed the unanimous resolution of the Senate."[259]

As noted earlier, an important exception to this early-twentieth-century consensus was the view of the constitutional law scholar Westel Willoughby, who stated without discussion in his 1910 constitutional law treatise that "[t]hough the Senate participates in the ratification of treaties, the President has the authority, without asking for senatorial advice and consent, to denounce an existing treaty and to declare it no longer binding upon the United States."[260] The second edition of Willoughby's treatise, published in 1929, contains a much more extensive discussion of the issue of treaty termination, but it argues only that the President is not obligated to submit his treaty terminations to the full Congress and does not specifically address whether he must obtain the consent of the Senate.[261]

In any event, by the 1920s there were additional commentators who defended a unilateral presidential authority to terminate treaties. For example, John Mabry Mathews, in his 1922 treatise on foreign relations law, argued that

---

257. EDWARD S. CORWIN, THE PRESIDENT'S CONTROL OF FOREIGN RELATIONS 115 (1917).

258. QUINCY WRIGHT, THE CONTROL OF AMERICAN FOREIGN RELATIONS 260 (1922). The year before, however, Wright had stated that "[p]ractice seems to sanction independent initial negotiation *and denunciation* of treaties by the President." Quincy Wright, *The Control of Foreign Relations*, 15 AM. POL. SCI. REV. 1, 11 (1921) (emphasis added).

259. 2 CHARLES CHENEY HYDE, INTERNATIONAL LAW: CHIEFLY AS INTERPRETED AND APPLIED BY THE UNITED STATES § 539, at 80 (1922).

260. 1 WESTEL WOODBURY WILLOUGHBY, THE CONSTITUTIONAL LAW OF THE UNITED STATES § 223 (1910).

261. *See* 1 WESTEL WOODBURY WILLOUGHBY, THE CONSTITUTIONAL LAW OF THE UNITED STATES § 324, at 585 (2d ed. 1929) ("[T]here is no constitutional obligation upon the part of the Executive to submit his treaty denunciations to the Congress for its approval . . . ."). On the issue of whether the President needed congressional approval, Willoughby expressly disagreed with Corwin. *See id.* at 585 n.59 (stating that "[t]he author cannot, therefore, accept the conclusion of Corwin" that the power of treaty termination rests with Congress).

> since the Senate has already, in its treaty-making capacity, acted upon a treaty providing for its termination upon notice, no further Senatorial action is necessary in effecting such termination, and that the President alone, as the mouthpiece of the nation in its international relations, may denounce the treaty by giving notice of its termination.[262]

Similarly, Jesse Reeves (a political science professor at the University of Michigan) expressed the view in 1921 that "[i]t seems to be within the power of the President to terminate treaties by giving notice on his own motion without previous Congressional or Senatorial action."[263]

Nevertheless, scholarly views continued to be mixed, and there did not appear to be any settled understanding that the President possessed a unilateral power of termination. Berkeley law professor Stefan Riesenfeld, writing in 1937, argued that

> [t]he most logical view is that the power to denounce a treaty is vested in the President by and with the advice and consent of the Senate, so that the department of the government which makes the treaty can terminate it, regardless of whether the termination is by unilateral, but lawful, denunciation or by a new treaty.[264]

In his history of the Senate, published in 1938, George Haynes observed that there was uncertainty about whether the President could unilaterally terminate a treaty and that "[d]enunciation of treaties has usually been by joint resolution, originating sometimes in the House, sometimes in the Senate."[265]

By the 1940s, however, scholarly commentary was increasingly supportive of unilateral presidential authority. For example, the second edition of Hyde's treatise on international law (published in 1945 after Hyde had served as Solicitor for the State Department) added to what it had stated in 1922 as follows:

> The President is not believed, however, to lack authority to denounce, in pursuance of its terms, a treaty to which the United States is a party, without legislative approval. In taking such action, he is merely exercising in behalf of the nation a privilege already conferred upon it by the agreement, and which involves no necessary modification thereof. Denunciation in such case may be regarded as

---

262. JOHN MABRY MATHEWS, THE CONDUCT OF AMERICAN FOREIGN RELATIONS 252 (1922).

263. Reeves, *supra* note 101, at 38.

264. Riesenfeld, *supra* note 135, at 660.

265. 2 GEORGE H. HAYNES, THE SENATE OF THE UNITED STATES: ITS HISTORY AND PRACTICE 670 (1938).

> a mere normal incident in the conduct of foreign relations as they are confided to the Executive.[266]

This was also a time of significant discussion of the President's power to conclude executive agreements, and commentators who favored broad presidential authority to conclude such agreements also tended to favor unilateral presidential authority to terminate treaties.[267]

There were additional scholarly endorsements of unilateral presidential termination authority in the 1950s and 1960s.[268] The American Law Institute's *Restatement (Second) of the Foreign Relations Law of the United States*, published in 1965, continued this trend. It contended that the President had the authority to terminate a treaty pursuant to the terms of the treaty or based on the grounds for termination allowed under international law.[269] The *Restatement* explained that this power stemmed from "the authority of the President to conduct the foreign relations of the United States as part of the executive power vested in him by Article II, Section 1 of the Constitution."[270]

---

266. 2 CHARLES CHENEY HYDE: INTERNATIONAL LAW, CHIEFLY AS INTERPRETED AND APPLIED BY THE UNITED STATES § 539, at 1519–20 (2d rev. ed. 1945) (footnote omitted).

267. *See, e.g.*, WALLACE MCCLURE, INTERNATIONAL EXECUTIVE AGREEMENTS: DEMOCRATIC PROCEDURE UNDER THE CONSTITUTION OF THE UNITED STATES 306 (1941) (claiming that "[i]n treaty making[,] . . . negative action, not being feared by the constitution makers, was left to the repository of general executive power"); Myres S. McDougal & Asher Lans, *Treaties and Congressional-Executive or Presidential Agreements: Interchangeable Instruments of National Policy: I*, 54 YALE L.J. 181, 336–37 (1945) (asserting that termination of both executive agreements and treaties can be "effected by executive denunciation, with or without prior Congressional authorization"). Many years later, in the context of President Carter's termination of the Taiwan Treaty, McDougal appeared to have changed his mind. *See Treaty Termination Hearings*, *supra* note 205, at 387, 391 (statement of Michael Reisman) (averring on behalf of himself and McDougal that "the constitutional system, if we are going back to this fundamental dynamic, seems to be based on a notion of sharing of power, rather than shifting it all to one branch"). McDougal also joined an amicus brief in the Taiwan case on behalf of the plaintiffs, Brief of Myres S. McDougal & W. Michael Reisman as Amici Curiae in Support of Petition for Certiorari, Goldwater v. Carter, 444 U.S. 996 (1979) (No. 79-856), and co-authored an article in the *National Law Journal* arguing against a unilateral presidential power of termination. *See* Michael Reisman & Myres S. McDougal, *Who Can Terminate Mutual Defense Treaties?*, NAT'L L.J., May 21, 1979, at 19, 19 ("[I]n the absence of material breach or *rebus sic stantibus* and, arguably, in the absence of an overwhelming external crisis to the body politic, the presumption must be that the president requires congressional authorization to terminate any agreement, other than a presidential agreement.").

268. *See, e.g.*, 2 BERNARD SCHWARTZ, A COMMENTARY ON THE CONSTITUTION OF THE UNITED STATES: THE POWERS OF GOVERNMENT § 215, at 132 (1963) (noting that the unilateral termination of a treaty by the President "appears justified by the constitutional position of the President as the nation's sole organ of foreign intercourse"); Randall H. Nelson, *The Termination of Treaties and Executive Agreements by the United States: Theory and Practice*, 42 MINN. L. REV. 879, 887 (1958) (expounding that, because the "conduct of foreign relations" is a "plenary executive power" and no limitation is placed on treaty termination under the Constitution, the President has the power to unilaterally terminate treaties).

269. RESTATEMENT (SECOND) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 163 (1965).

270. *Id.* § 163 cmt. a.

In his influential foreign relations law treatise, published in 1972, Louis Henkin suggested that the answer to the constitutional question was unclear but that "since the President acts for the United States internationally he can effectively terminate or violate treaties, and the Senate has not established its authority to join or veto him."[271] He also noted that "[i]f issues as to who has power to terminate treaties arise again, . . . it seems unlikely that Congress will successfully assert the power."[272] Here, Henkin appears to have been making a political science observation as much as a legal observation: whatever one may think about the correct distribution of constitutional authority on this issue, Henkin was suggesting that the President's assertion of unilateral authority was likely to prevail as a practical matter in congressional–executive relations.[273] Historical practice since 1972 tends to support this assessment.

The controversy over President Carter's termination of the Taiwan treaty revealed that the issue was still not settled, and, as noted, a number of scholars at that time took the position that congressional or senatorial approval was required for treaty termination.[274] Since that termination, however, the controversy seems to have receded. Like the earlier *Restatement (Second)*, the *Restatement (Third) of the Foreign Relations Law of the United States*, published in 1987, contends that the President has the authority to terminate a treaty as long as the treaty allows for unilateral withdrawal or there is an international-law ground for termination.[275] A number of scholars, including some who do not always favor expansive readings of presidential authority, have agreed with this proposition.[276] As a result, it is probably fair to describe this as the prevailing, although certainly not unanimous, view.[277]

---

271. LOUIS HENKIN, FOREIGN AFFAIRS AND THE CONSTITUTION 169 (1972).

272. *Id.* at 170.

273. *See also* FRANCIS D. WORMUTH & EDWIN B. FIRMAGE, TO CHAIN THE DOG OF WAR: THE WAR POWER OF CONGRESS IN HISTORY AND LAW 194 (2d ed. 1989) ("[T]he President has demonstrated an effective power to terminate treaties, and the Senate has not successfully challenged that right to do so.").

274. *See supra* text accompanying note 223.

275. RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 339 (1987).

276. *See, e.g.*, MICHAEL J. GLENNON, CONSTITUTIONAL DIPLOMACY 153–55 (1990) (pronouncing the *Restatement*'s position "sound"); Jinks & Sloss, *supra* note 38, at 156 (agreeing with the functionalist rationale of presidential power to terminate treaties).

277. In its comprehensive 2001 study on treaties, prepared for the Senate Foreign Relations Committee, the Congressional Research Service noted that "[t]he constitutional requirements that attend the termination of treaties remain a matter of some controversy," and it described the issue of whether the President has a unilateral termination power to be "a live issue." CRS STUDY, *supra* note 42, at 198–99. Nevertheless, it also noted that, "[a]s a practical matter . . . the President may exercise this power since the courts have held that they are conclusively bound by an executive determination with regard to whether a treaty is still in effect." *Id.* at 201; *cf.* H. Jefferson Powell, Essay, *The President's Authority over Foreign Affairs: An Executive Branch Perspective*, 67 GEO. WASH. L. REV. 527, 562 (1999) ("Despite its obvious importance and the

IV. Implications for Law, Theory, and Politics

This Part considers the implications of the historical practice of U.S. treaty terminations, both for the specific issue of whether the President has a unilateral termination authority and for the more general historical gloss method of constitutional interpretation. It also reflects on the extent to which a practice-based account of institutional authority, such as the account given here, constitutes a description of constitutional law as opposed to a description of mere politics.

A. Current Law of Treaty Termination

As we have seen, as a matter of practice, presidents today exercise a unilateral power of treaty termination. The precedent for this practice can be traced back to the end of the nineteenth century, and the practice has been especially robust since the 1930s.[278] Moreover, with the important exception of the debate over the termination of the Taiwan Treaty, Congress has not seriously opposed exercises of this presidential authority.[279] Even during the Taiwan Treaty debate, the Senate Foreign Relations Committee took the position that the President had the authority to terminate a treaty when, as was true there, termination was permissible under international law. To be sure, a majority of the Senate appeared to disagree with the Committee, but it is also the case that the full Senate never voted on any resolution to contest the President's authority.

As discussed in Part I, most accounts of how historical practice can inform the separation of powers would require "acquiescence" by the affected branch of government. There are a number of conceptual difficulties with this concept, however, especially as applied to Congress, and these difficulties argue for caution before treating mere inaction by Congress as acquiescence.[280] Nevertheless, the congressional inaction surrounding the issue of treaty termination is noteworthy. First, it has been longstanding, involving numerous congresses and presidential administrations, during times of both unified and divided government. With the exception of the debate over the termination of the Taiwan Treaty, there has been a century of congressional passivity in the face of presidential treaty terminations. Second, Congress has failed to protest presidential terminations even with "soft law" measures such as one-house resolutions or statements by congressional leadership, even when presidential treaty terminations have received significant public attention (as they did, for example, in both the Taiwan termination debate and the

---

substantial history surrounding the issue, the question of which political branch has the power to withdraw from or terminate treaties remains unsettled.").

278. *See supra* subpart III(B).

279. *See supra* subpart III(C).

280. *See supra* text accompanying notes 67–71.

debate over the termination of the ABM Treaty). Third, even though it has approved numerous treaties containing withdrawal clauses, the Senate has failed to address the question of which U.S. actor can invoke these clauses, even though it could easily do so in its resolutions of advice and consent.[281]

There are, in any event, reasons for crediting historical practice in the separation of powers area that do not turn on institutional acquiescence.[282] One such reason is the general desirability, for legitimacy and other reasons, of having an account of constitutional law that bears a reasonable resemblance to actual constitutional practice, both now and in the foreseeable future.[283] In addition, if in fact government actors look to past practice to inform their own understanding of—and to shape their claims about—the law, legal philosophers working in the tradition of H.L.A. Hart would treat that second-order practice as itself a fundamental feature of the legal order.[284] These considerations have particular salience for the issue of treaty termination. Unilateral presidential termination of treaties is an established and longstanding practice, and it seems unlikely that Congress will do anything in the coming years to destabilize that practice. Moreover, the courts have shown little inclination to resolve the issue, and the longer they wait the more entrenched the practice becomes. As a result, an account of modern U.S. constitutional law that denied a presidential authority to terminate treaties (at least as a general matter) would face serious descriptive limitations.[285]

---

281. *Cf.* Curtis A. Bradley & Jack L. Goldsmith, *Treaties, Human Rights, and Conditional Consent*, 149 U. PA. L. REV. 399, 399–405 (2000) (discussing conditions imposed by the Senate in connection with its ratification of human rights treaties).

282. *See* Bradley & Morrison, *supra* note 1, at 456–60 (describing reasons for crediting historical practice concerning separation of powers).

283. *See id.* at 456 (arguing that the legitimacy of a law is partially tied to actual behavior and practice related to it); *cf.* RONALD DWORKIN, LAW'S EMPIRE 66 (1986) ("The justification need not fit every aspect or feature of the practice, but it must fit enough for the interpreter to be able to see himself as interpreting that practice, not inventing a new one."); LON L. FULLER, THE MORALITY OF LAW 81 (rev. ed. 1969) (discussing the importance of "congruence between official action and the law").

284. *See* H.L.A. HART, THE CONCEPT OF LAW 94–99 (3d ed. 2012) (discussing secondary "rules of recognition"); 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) ("The ultimate rule of recognition is a matter of social fact, and so determining it is for empirical investigation rather than legal analysis."); *see also* Matthew D. Adler & Kenneth Einar Himma, *Introduction* to THE RULE OF RECOGNITION AND THE U.S. CONSTITUTION xiii, xv (Matthew D. Adler & Kenneth Einar Himma eds., 2009) ("[T]he U.S. rule of recognition may be substantially longer and more complicated than a simple reference to the 1787 Constitution (or the Amendment Clause thereof), in part because it may give independent effect to extraconstitutional sources of law, such as judicial precedent or official custom.").

285. Because the precise contours of constitutional custom are contestable, it is still possible to argue as a descriptive matter that certain types of treaties are not subject to unilateral presidential termination. It might be argued, for example, that in light of Congress's power to declare war, a president may not unilaterally terminate a peace treaty.

The absence of judicial review may itself be related to the longstanding nature of the practice. In abstaining on this issue, courts may reasonably perceive that the durability of a practice over numerous presidential administrations is evidence that the practice is functionally desirable, or at least not too functionally problematic.[286] It is easy to imagine that there are advantages to the United States of being able to make credible threats of exit from treaty regimes as part of negotiations to reform international institutions or induce better compliance by its treaty partners—advantages that could be facilitated by allowing for unilateral presidential action.[287] Moreover, it is possible that ease of exit as a matter of U.S. constitutional procedure makes it easier to persuade the Senate to agree to such treaties in the first place.[288] While such ease of exit could also in theory be destabilizing to foreign relations, it is not obvious from the historical record that there is any presidential tendency to devalue international commitments more than Congress.

For all these reasons, the best description of the current U.S. constitutional law governing treaty termination is probably as described by the *Restatement (Third) of the Foreign Relations Law*: the President has the unilateral authority to terminate treaties when such termination is permitted under international law and is not disallowed either by the Senate in its advice and consent to the treaty or by Congress in a statute.[289] Unlike the *Restatement (Third)*, however, which chiefly relies on a purported implication of the President's role as the "sole organ" in foreign affairs,[290] the account presented here is grounded chiefly in the longstanding accretion of Executive Branch practice and claims in the face of congressional inaction and judicial abstention.

Some scholars (and, for a time, the Executive Branch during the Bush Administration) have gone even further, suggesting that the President can

---

286. *Cf.* United States v. Curtiss-Wright Exp. Corp., 299 U.S. 304, 329 (1936) (explaining, in declining to invalidate a congressional delegation of foreign-affairs authority to the President, that "[t]he uniform, long-continued and undisputed legislative practice" of making broad delegations to the President in foreign affairs "rests upon an admissible view of the Constitution which, even if the practice found far less support in principle than we think it does, we should not feel at liberty at this late day to disturb").

287. *Cf.* Matthew C. Waxman, *The Constitutional Power to Threaten War*, 123 YALE L.J. (forthcoming 2014) (arguing that, in thinking about the scope of the President's war authority, it is important to consider the President's ability to threaten war).

288. Recall that an argument along these lines was made, albeit unsuccessfully, in an effort to broker a compromise on the Versailles Treaty. *See supra* notes 159–61 and accompanying text.

289. *See supra* note 275 and accompanying text.

290. *See* RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 339 reporters' note 1 (1987) (arguing that "[a] power so characterized would seem to include the authority to decide on behalf of the United States to terminate a treaty that no longer serves the national interest, or is out of date, or which has been breached by the other side" while also stating that the power to terminate treaties "is implied in [the President's] office *as it has developed over almost two centuries*" (emphasis added)).

(like Congress) terminate or override a treaty's domestic effect even when there is no basis in international law for terminating the treaty.[291] That proposition is highly contested, however, and there is little historical practice in support of it. Moreover, the unusual decision in 2009 by the Office of Legal Counsel to withdraw an earlier claim of this authority renders it even more suspect.[292]

To say that the President has a unilateral authority to terminate treaties is not to say that this is an *exclusive* presidential power. If it is merely a concurrent power shared with either the full Congress or the Senate, then either Congress or the Senate could potentially place limitations on it. The termination authority, in other words, would fall within what Justice Jackson described in *Youngstown* as an intermediate "zone of twilight" in which the President and Congress might have overlapping authority.[293] If Congress or the Senate took action to prohibit presidential termination—for example, if the Senate made senatorial approval of termination a condition of its advice and consent to the treaty—then a unilateral presidential termination in violation of such a condition would cause the President's action to fall within what Jackson referred to as the "lowest ebb" of presidential authority.[294]

During the debates over the termination of the Taiwan Treaty, the Executive Branch suggested that it viewed the presidential power of

---

291. *See, e.g.*, HENKIN, *supra* note 41, at 214 (arguing that when the President terminates a treaty, it ceases to exist in international and domestic law); John C. Yoo, Rejoinder, *Treaties and Public Lawmaking: A Textual and Structural Defense of Non-Self-Execution*, 99 COLUM. L. REV. 2218, 2242 (1999) (arguing that because the President, rather than Congress, has full policymaking control in treaty formation, the President may terminate a treaty unilaterally at will).

292. *See supra* note 250. This is an example of how international law might at least indirectly limit presidential authority: if international law causes a treaty to remain in force, then the U.S. Constitution may give the treaty a domestic-law status that cannot be terminated unilaterally by the President. A slight potential counterexample occurred in 2005, when the Bush Administration purported to withdraw the United States from a protocol to a consular convention. The Administration seemed to suggest that the withdrawal was effective immediately, whereas it was arguable that international law required a year's notice. *See* Kirgis, *supra* note 246 (discussing the legal ramifications of withdrawing from the consular convention). For additional consideration of potential interactions between international law and the separation of powers, see Jean Galbraith, *International Law and the Domestic Separation of Powers*, 99 VA. L. REV. 987 (2013).

293. *See* Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 637 (1952) (Jackson, J., concurring) ("When the President acts in absence of either a congressional grant or denial of authority, he can only rely upon his own independent powers, but there is a zone of twilight in which he and Congress may have concurrent authority, or in which its distribution is uncertain."); *see also* GLENNON, *supra* note 276, at 152 (arguing that "*in the face of congressional silence*, treaty termination by the President does not impinge upon the constitutional prerogatives of the Senate or Congress").

294. *See Youngstown*, 343 U.S. at 637 (Jackson, J., concurring); *see also* Hamdan v. Rumsfeld, 548 U.S. 557, 593 n.23 (2006) ("Whether or not the President has independent power, absent congressional authorization, to convene military commissions, he may not disregard limitations that Congress has, in proper exercise of its own war powers, placed on his powers.").

termination as exclusive.[295] Importantly, though, the Senate Foreign Relations Committee made clear during the debate that it did not accept that proposition, despite otherwise favoring robust presidential authority with respect to treaty terminations.[296] Moreover, there is no significant historical practice to support the Executive Branch's claim. Perhaps for this reason, the *Restatement (Third)* contends that if the Senate gave its advice and consent to a treaty on the condition that any termination occur only with its consent, and the President proceeded to conclude the treaty, "he would be bound by the condition."[297] A number of scholars have expressed agreement with this proposition.[298] In its 2001 study on treaties prepared for the Senate Foreign Relations Committee, the Congressional Research Service correctly noted that "the assertion of an exclusive Presidential power in the context of a treaty is controversial and flies in the face of a substantial number of precedents in which the Senate or Congress have been participants."[299]

B. Constitutional Interpretation and Change

The account given in subpart IV(A) of the current constitutional law of treaty termination has potential implications for theories of constitutional interpretation and change. Under that account, a unilateral presidential termination authority does not exist today because of an assessment of founding intent or understanding. Nor does it follow clearly from constitutional text or structure, or from judicial decisions, although those aspects of constitutional interpretation are of course relevant. Rather, the President's constitutional authority for this issue exists in part because *some aspects of U.S. constitutional law are made by the participants in the*

---

295. *See Treaty Termination Hearings*, *supra* note 205, at 218 (statement of Larry A. Hammond, Deputy Assistant Att'y Gen., Office of Legal Counsel) ("[W]e do not believe that the Senate may expand that advice and consent power by attaching reservations with respect to termination.").

296. *See* S. REP. NO. 96-119, at 11 (1979) (expressing the view that it was "clear beyond question" that the Senate could validly limit the President's authority to terminate a treaty by placing a condition on such termination in the Senate's advice and consent to the treaty).

297. RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 339 reporters' note 3 (1987).

298. *See, e.g.*, GLENNON, *supra* note 276, at 156 (arguing that the Constitution compels the President to follow any termination procedure prescribed by the Senate); Kristen E. Eichensehr, *Treaty Termination and the Separation of Powers*, 53 VA. J. INT'L L. 247, 279–86 (2013) (arguing that "for cause" limitations imposed by the Senate on the President's treaty-termination power are constitutional); *see also* Powell, *supra* note 277, at 563 (concluding, based largely on historical practice, that the power of treaty termination is not exclusive to the President). Presumably, Congress could similarly limit presidential withdrawal from "congressional–executive agreements"—that is, international agreements approved or authorized by a majority of both houses of Congress rather than two-thirds of the Senate. *See* Hathaway, *supra* note 27, at 1332–33 (discussing possible limitations that Congress can place on the President in such agreements).

299. CRS STUDY, *supra* note 42, at 199.

system over time. Treaty termination is, in another words, an instance of what some scholars have termed "constitutional construction"—the fleshing out of constitutional meaning in ways that go beyond merely interpreting constitutional text.[300]

The best description of this constitutional law today is also different from the description that most constitutional observers would have been given, say, in 1900. Treaty termination thus provides a vivid illustration of how constitutional understandings can change even when the courts are not involved. This change did not occur at one particular moment in time but rather developed over the course of decades. While the dispute over President Carter's termination of the Taiwan Treaty in the late 1970s was important in leading to a consolidation of presidential authority over this issue, that consolidation was facilitated by the accretion of claims and practice that had already occurred. The dynamic described here thus differs from accounts of constitutional change that focus primarily on dramatic moments and episodes.[301] There are reasons to believe, moreover, that something like this pattern of constitutional change can be identified for other issues as well, especially in the area of separation of powers.[302]

---

300. *See* JACK M. BALKIN, LIVING ORIGINALISM 5 (2011) (noting that the actions of all three branches of government can contribute to constitutional construction); KEITH E. WHITTINGTON, CONSTITUTIONAL CONSTRUCTION: DIVIDED POWERS AND CONSTITUTIONAL MEANING 228 (1999) (discussing constitutional construction, the realm in which "the Constitution adapts and evolves to accommodate and to cause external change"); *see also* STEPHEN M. GRIFFIN, LONG WARS AND THE CONSTITUTION (2013) (considering how modern war powers authority has been constitutionally constructed); Alan M. Wachman, *Carter's Constitutional Conundrum: An Examination of the President's Unilateral Termination of a Treaty*, 8 FLETCHER F. WORLD AFF. 427, 456 (1984) (contending that "a decision [about treaty termination authority] would be a constitutional construction of our own making, not one found in the document [of the Constitution]"). For additional discussion of the concept of constitutional construction and the relationship of this concept to originalist theories of constitutional interpretation, see Lawrence B. Solum, *Originalism and Constitutional Construction*, 82 FORDHAM L. REV. 453 (2013) (discussing originalism and constitutional construction).

301. *See generally, e.g.*, 1 BRUCE ACKERMAN, WE THE PEOPLE: FOUNDATIONS (1991) (arguing that there are rare instances in American politics of "higher lawmaking" sufficient to change the Constitution despite the absence of a formal amendment); Eric A. Posner & Adrian Vermeule, *Constitutional Showdowns*, 156 U. PA. L. REV. 991 (2008) (exploring the effect of "constitutional showdowns," which involve interbranch confrontations that can produce precedent about the meaning of the Constitution).

302. For preliminary case studies on war powers and congressional–executive agreements that describe somewhat comparable patterns, see Bradley & Morrison, *supra* note 1, at 461–76. *See also* Peter J. Spiro, *Treaties, Executive Agreements, and Constitutional Method*, 79 TEXAS L. REV. 961, 1009 (2001) (arguing, in addressing the debate over the constitutionality of congressional–executive agreements, that constitutional change can and does occur through "increments" rather than dramatic points in time); Peter J. Spiro, *War Powers and the Sirens of Formalism*, 68 N.Y.U. L. REV. 1338, 1355 (1993) (reviewing JOHN HART ELY, WAR AND RESPONSIBILITY: CONSTITUTIONAL LESSONS OF VIETNAM AND ITS AFTERMATH (1993)) (suggesting, in a consideration of the distribution of war authority between Congress and the President, that the relevant constitutional law stems from "an accretion of interactions among the branches" that "gives rise to basic norms governing the branches' behavior in the area").

Very likely the change in treaty-termination practice was driven in part by other changes—such as the increased role of the United States in the world—that were contributing to the enhancement of Executive authority across a wide range of issues.[303] The growth in both treaty-making in general, and the increasingly widespread inclusion of unilateral withdrawal clauses in treaties, probably also were factors. But lawyers, including lawyers within the State Department as well as legal scholars, also appear to have played an active role in assessing and influencing the relationship between the constitutional practice and constitutional understandings.[304] While not playing a direct role, the Supreme Court also may have helped facilitate the shift, through its increasingly deferential posture towards the Executive Branch starting in the 1930s.[305]

That constitutional change occurs in the United States in this way does not necessarily mean, of course, that it is desirable. The lack of modern resistance by Congress to presidential unilateralism on treaty termination could be for normatively attractive reasons, such as a recognition that the President is likely to have better information about the costs and benefits of such action and will have more negotiating power if he can make threats that are not dependent on legislative ratification. But this lack of resistance could be for other reasons, such as a disinterest by members of Congress in issues that are unlikely to be of concern to constituents, a phenomenon that may apply to a broad range of foreign-affairs issues, including treaty termination.[306] If so, crediting such inaction might produce socially undesirable outcomes.

The accretion dynamic described here also implicates tradeoffs associated more generally with the idea of "common law constitutionalism," an approach usually associated with judicial decision making but which in theory might also apply to constitutional reasoning by nonjudicial actors.[307] On the one hand, having the law develop through the accretion of precedents can lead to path dependency and, relatedly, a lack of

303. *See supra* notes 198–204 and accompanying text.

304. *See supra* notes 146, 189, 192, 202, 208 and accompanying text.

305. *See supra* text accompanying notes 198–200. Although not explored here, the social science literature on "historical institutionalism" might offer additional insights for assessing this sort of change in institutional practice. Recent scholarship in that area has focused on how institutions change, sometimes dramatically, through incremental shifts. *See generally, e.g.*, PAUL PIERSON, POLITICS IN TIME: HISTORY, INSTITUTIONS, AND SOCIAL ANALYSIS (2004); KATHLEEN THELEN, HOW INSTITUTIONS EVOLVE: THE POLITICAL ECONOMY OF SKILLS IN GERMANY, BRITAIN, THE UNITED STATES, AND JAPAN (2004).

306. *See* Bradley & Morrison, *supra* note 1, at 442 (describing the focus on reelection as a primary motivator for the actions of members of Congress).

307. *See, e.g.*, David A. Strauss, *Common Law Constitutional Interpretation*, 63 U. CHI. L. REV. 877, 925 (1996) [hereinafter Strauss, *Constitutional Interpretation*] (describing common law constitutionalism); *see also* DAVID A. STRAUSS, THE LIVING CONSTITUTION 33–42, 46–49 (2010) (same).

concentrated deliberation. On the other hand, it can also help ensure that the law is shaped to address specific, real-world contexts rather than abstract speculations about the future. This benefit might have particular salience for foreign relations law issues, such as treaty termination, in light of the ever-changing nature of the international environment and the United States' role within it. The wide variety of situations that might trigger a decision to suspend or terminate treaty obligations, or to threaten to do so, also supports an inductive, evolutionary approach to the issue rather than one based on a general theory or abstract reasoning.

Like any precedent-based approach, the historical gloss method of discerning the separation of powers also presents interpretive challenges. As an initial matter, there can be difficult questions about what counts as relevant practice. For example, it might be unclear how to weight claims of authority made by institutional actors that are not carried out (such as treaty terminations that are threatened but then rescinded). In addition, customary practice is not self-liquidating; it requires interpretation and description, which inevitably involves an element of judgment and subjectivity.[308] Of course, the same is probably true of other sources of constitutional interpretation, but the lack of a canonical text may exacerbate the difficulty. Moreover, if the relevant law is tied to practice, then the law can potentially change over time, as in fact appears to have happened with respect to the authority over treaty termination. Although this might be perceived as a virtue in that it allows the law to adapt to changing conditions, it might also pose challenges for stability and predictability in the law.[309] Again, though, this is not a problem unique to this interpretive source; constitutional law can and does change, for example, through Supreme Court interpretations. At least with Supreme Court opinions, however, there is an understood public text that serves as a point of reference and potentially also as a stare decisis break on deviations.

There is another potential problem that relates specifically to the reliance on historical practice in the area of separation of powers. For a

---

308. *See, e.g.*, Richard Craswell, *Do Trade Customs Exist?*, *in* THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW 118, 122 (Jody S. Kraus & Steven D. Walt eds., 2000) ("[A]ny history of prior decisions will always underdetermine the possible patterns that might be ascribed to that history."); Martin S. Flaherty, *Post-Originalism*, 68 U. CHI. L. REV. 1089, 1105 (2001) (reviewing DAVID P. CURRIE, THE CONSTITUTION IN CONGRESS: THE JEFFERSONIANS, 1801–1829 (2001)) ("As a theoretical matter, custom has its own problems. Not least among these are the questions of what counts as the relevant custom, at what level of generality, and for how long."). *But cf.* Michael D. Ramsey, *The Limits of Custom in Constitutional and International Law*, 51 SAN DIEGO L. REV. (forthcoming 2014), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2249133 (distinguishing between applications of custom that do not involve contested value judgments and those that do).

309. *See, e.g.*, Keith E. Whittington, *The Status of Unwritten Constitutional Conventions in the United States*, 2013 U. ILL. L. REV. 1847, 1869 (noting that constitutional conventions "are under constant pressure of erosion").

variety of reasons, the Executive Branch probably has a greater ability than Congress to generate both institutional practice and instances of nonacquiescence.[310] If so, then there is an obvious danger that a practice-based approach will favor Executive authority over the long term, which may contribute to an imbalance of authority between the branches. Indeed, it is generally thought that presidential authority has expanded in the modern era relative to congressional authority.[311] This phenomenon might be exacerbated by a tendency of Executive Branch lawyers to over-claim about past practice, something that appears to have been the case at various times with respect to the issue of treaty termination.[312]

Many commentators have suggested that the solution to the potential imbalance between the ability of Congress and the President to take direct action is greater judicial review.[313] It may well be that some additional amount of judicial review is needed in the separation of powers area, especially if judicial abstention is premised on the idea that Congress has sufficient capacity and incentives to sufficiently guard its institutional interests.[314] At the same time, courts are themselves part of the separation of powers structure, and thus there is no guarantee that they will be less acquiescent than Congress when faced with Executive unilateralism.[315]

---

310. *See* Bradley & Morrison, *supra* note 1, at 439–45 (discussing structural impediments, political asymmetries, and issues of congressional–executive relations as explaining why Congress and the President "are not equally situated in their ability to take action"); *see also* Terry M. Moe & William G. Howell, *The Presidential Power of Unilateral Action*, 15 J.L. ECON. & ORG. 132, 133–34 (1999) (describing a variety of ways in which presidents can take actions that have legal effect without the participation of Congress).

311. *See* Richard H. Pildes, *Law and the President*, 125 HARV. L. REV. 1381, 1381 (2012) (reviewing ERIC A. POSNER & ADRIAN VERMEULE, THE EXECUTIVE UNBOUND: AFTER THE MADISONIAN REPUBLIC (2010)) ("It is widely recognized that the expansion of presidential power from the start of the twentieth century onward has been among the central features of American political development.").

312. *See supra* notes 50–52, 218 and accompanying text.

313. *See, e.g.*, ELY, *supra* note 302, at 54 (discussing how the courts' "relative insulation from the democratic process . . . situate[s] them uniquely well to police malfunctions in that process"); THOMAS M. FRANCK, POLITICAL QUESTIONS/JUDICIAL ANSWERS: DOES THE RULE OF LAW APPLY TO FOREIGN AFFAIRS? 7 (1992) (arguing that judges should stop abdicating in favor of the other branches of government in foreign-affairs cases because they "are much better suited than is sometimes alleged to make decisions incidentally affecting foreign relations and national security"); HAROLD HONGJU KOH, THE NATIONAL SECURITY CONSTITUTION: SHARING POWER AFTER THE IRAN-CONTRA AFFAIR 181–84 (1990) (noting that "the role of judges is to define the rule of law by drawing the line between illegitimate exercises of political power and legitimate exercises of legal authority," in part by moving away from doctrines of abstention in certain types of cases).

314. *See* Bradley & Morrison, *supra* note 1, at 451–52 (questioning "Madisonian assumptions about congressional capacity and motivation" and arguing that "courts should be more circumspect about invoking congressional acquiescence as a basis for deferring to executive practice").

315. *See* Eric A. Posner & Adrian Vermeule, *Inside or Outside the System?*, 80 U. CHI. L. REV. 1743, 1752 (2013) ("[A]rguments for 'Madisonian' judging go wrong by assuming that judges stand outside the Madisonian system."); *cf.* Jide Nzelibe, *Our Partisan Foreign Affairs*

Moreover, for a variety of reasons, courts often give weight to established patterns of governmental practice.[316] If so, they might actually reduce Congress's ability to resist assertions of presidential authority rather than enhance it, by instantiating Executive practice into judicial doctrine.

In any event, it is worth noting that the shift to a new understanding of presidential authority on treaty termination cannot be attributed simply to Executive aggrandizement. It is striking how actively involved Congress and the Senate were in these issues in the nineteenth and early twentieth centuries, even in instances in which presidents sought to act unilaterally. That sort of congressional focus on treaty termination dissipated, however, by the 1930s. Although the issue would resurface in select instances of policy debate, most notably in the debate over the termination of the Taiwan Treaty, Congress and the Senate no longer sought to protect institutional prerogatives relating to treaty termination in any systematic way. Moreover, Congress and the Senate seem largely to have given up on the issue since the Taiwan debate, mounting only token resistance at the time of the termination of the ABM Treaty and no resistance at all to dozens of other presidential terminations.

Whether normatively attractive or not, the influence of historical practice on the separation of powers is likely to vary depending on the issue. Treaty termination is an especially good candidate for it, given the lack of any specific constitutional text relating to the issue.[317] The overlay

---

Constitution, 97 MINN. L. REV. 838, 899–900 (2013) ("Even though judges and academic commentators may not necessarily be susceptible to the same instrumental motivations as elected officials, they may very well be plagued by both the kinds of cognitive biases and motivated reasoning that largely track partisan judgments in the electoral arena.").

316. *See* Bradley & Morrison, *supra* note 1, at 418–22 (discussing the Supreme Court's recognition of "the significance of . . . practice-based 'gloss'" when textual or other forms of guidance are absent or ambiguous).

317. When there is constitutional text that is perceived to be clear, it is likely to serve as a focal point for the practice of government actors. *See, e.g.*, Daryl J. Levinson, *Parchment and Politics: The Positive Puzzle of Constitutional Commitment*, 124 HARV. L. REV. 657, 708 (2011) (noting that "it is an indisputable feature of constitutional practice that the text is taken to be authoritative within its domain"); John O. McGinnis, *Constitutional Review by the Executive in Foreign Affairs and War Powers: A Consequence of Rational Choice in the Separation of Powers*, 56 LAW & CONTEMP. PROBS. 293, 300–01 (1993) ("Relatively clear [constitutional] provisions in the separation of powers area may be enforced because they are natural focal points of bargains."); Strauss, *Constitutional Interpretation*, *supra* note 307, at 911 (describing "conventionalism" as "a way of avoiding costly and risky disputes and of expressing respect for fellow citizens" through "allegiance to the text of the Constitution"). Relatedly, clear text may have a tendency to "crowd out" norms based on practice. *See* Michael C. Dorf, *How the Written Constitution Crowds Out the Extraconstitutional Rule of Recognition*, *in* THE RULE OF RECOGNITION, *supra* note 284, at 69, 76. That said, whether text is perceived as being clear might itself be affected by practice. *See* Bradley & Morrison, *supra* note 1, at 431; *see also* Curtis A. Bradley & Neil S. Siegel, Constructed Constraint and the Constitutional Text (Feb. 19, 2014) (unpublished manuscript) (on file with author) (developing this point). The perception of textual clarity might also be affected by what is at stake, *see* Levinson, *supra*, at 709–10 (asserting that the Constitution is perceived as being clear on many low-stakes issues but unclear on many high-stakes ones), and by one's

of a mix of international-law rules governing treaty termination, as well as potential distinctions between suspension and termination and between partial and complete termination, have also made presidential unilateralism relating to the issue a more complicated target to assess and criticize. In addition, judicial review has been especially limited for this issue, which means that the political branches have had to work the issue out themselves, without even much of a shadow of judicial supervision. On issues for which there is more textual guidance, a less complicated legal landscape, or a greater likelihood of judicial intervention, practice is likely to play a lesser role. Certainly decisions like *INS v. Chadha*[318] confirm that the Supreme Court will not inevitably give effect to even longstanding political-branch practice.[319]

C. *Is It Law?*

Another challenge to the practice-based approach to constitutional authority described in this Article would be to dismiss it as merely an account of politics rather than law. The argument would be that, without any dispositive judicial resolution, the practice will simply be the result of the push and pull of the political process. The constitutional "law" of treaty termination, on this account, would merely be a pattern of behavior without normative significance.[320] If so, it might not be entitled to any particular weight in debates about constitutional interpretation.

As an initial matter, it is not clear why judicial review is so central to this purported distinction between politics and law. Presidential compliance with judicial decisions is itself a practice-based norm of U.S. constitutional law. It is largely taken for granted today, but this has not always been the case. As Daryl Levinson has noted, "[c]asting courts as constitutional enforcers merely pushes the question back to why powerful political actors are willing to pay attention to what judges say; why 'people with money and guns ever submit to people armed only with gavels.'"[321]

---

constitutional methodology, *see* Alison L. LaCroix, *Historical Gloss: A Primer*, 126 HARV. L. REV. F. 75, 81 (2013) ("Whether a text is ambiguous is itself determined by one's chosen interpretive method.").

318. 462 U.S. 919 (1983).

319. *Id.* at 944–45, 959 (holding that a "legislative veto" provision enacted by Congress was unconstitutional even though Congress had enacted hundreds of legislative veto provisions since the 1930s).

320. *Cf.* Jack Goldsmith & Daryl Levinson, *Law for States: International Law, Constitutional Law, Public Law*, 122 HARV. L. REV. 1791, 1836 (2009) ("We might also understand the settlement of non-textual constitutional issues as instances of successful coordination."); Posner & Vermeule, *supra* note 301, at 1002 ("Precedents may just be patterns of behavior that parties recognize as providing focal points that permit cooperation or coordination.").

321. Levinson, *supra* note 317, at 661 (quoting Matthew C. Stephenson, *"When the Devil Turns . . .": The Political Foundations of Independent Judicial Review*, 32 J. LEGAL STUD. 59, 60 (2003)).

To be sure, there is a strand of British (and, more generally, Commonwealth) constitutional thinking that would limit the term "constitutional law" to norms that are enforceable by the judiciary. Under this view, associated most notably with the writings of A.V. Dicey, norms of constitutional practice that are not judicially enforceable are termed instead "constitutional conventions."[322] This distinction, however, does not map well onto U.S. constitutional understandings. For example, there are a variety of nonjusticiability doctrines in U.S. law, such as the political question doctrine, that hypothesize that there can be constitutional law that might not be judicially enforceable.[323] In addition, there has been a significant emphasis in U.S. scholarship in recent years on the importance of "constitutional law outside the courts,"[324] an approach that implicitly declines to equate constitutional law simply with what is enforced by the judiciary. The longstanding idea of "underenforced constitutional norms" similarly is based on the idea that constitutional law is broader than what is judicially enforceable.[325]

In any event, the likelihood of judicial review for the issue of treaty termination is not zero, and in fact the lower federal courts did address the issue in the controversy over the termination of the Taiwan Treaty.[326] Thus, even if a shadow of possible judicial review were needed in order for a norm to have a legal character, such a shadow does exist for this issue, although it may be faint. Moreover, we know that courts often take account of longstanding practices when interpreting the separation of powers.[327]

---

322. *See* A.V. DICEY, INTRODUCTION TO THE STUDY OF THE LAW OF THE CONSTITUTION 361, 366 (6th ed. 1902) ("'[C]onventions of the constitution' . . . [are] customs, practices, maxims, or precepts which are not enforced or recognised by the Courts" and "cannot be enforced by any Court of law [and so] have no claim to be considered laws"); *see also* Adrian Vermeule, *Conventions of Agency Independence*, 113 COLUM. L. REV. 1163, 1182 (2013) (noting this point).

323. *See, e.g.*, Vieth v. Jubelirer, 541 U.S. 267, 292 (2004) (plurality opinion) ("The issue [before the Court] is not whether severe partisan gerrymanders violate the Constitution, but whether it is for the courts to say when a violation has occurred, and to design a remedy.").

324. *See, e.g.*, LARRY D. KRAMER, THE PEOPLE THEMSELVES (2004); MARK TUSHNET, TAKING THE CONSTITUTION AWAY FROM THE COURTS (1999).

325. *See, e.g.*, Lawrence Gene Sager, *Fair Measure: The Legal Status of Underenforced Constitutional Norms*, 91 HARV. L. REV. 1212, 1221 (1978) (stating that "constitutional norms which are underenforced by the federal judiciary should be understood to be legally valid to their full conceptual limits, and federal judicial decisions which stop short of these limits should be understood as delineating only the boundaries of the federal courts' role in enforcing the norm" rather than the boundaries of the norms themselves); *see also* Richard H. Fallon, Jr., *Judicially Manageable Standards and Constitutional Meaning*, 119 HARV. L. REV. 1274, 1299 (2006) (crediting Sager's argument that "it would be a mistake to equate judicial enforcement, and thus the tests applied by courts, with the meaning of constitutional guarantees").

326. The Supreme Court has also recently signaled a narrow view of the political question doctrine, even in the area of foreign affairs. Zivotofsky v. Clinton, 132 S. Ct. 1421, 1427 (2012) (describing the political question doctrine as a "narrow exception" to the judiciary's obligation to decide cases).

327. *See supra* notes 53–62 and accompanying text.

Another potential answer to the "it's all politics" critique is the simple observation that participants in the legal system generally view the issue of treaty termination as governed by legal norms. As noted earlier, at least according to legal philosophers working in the tradition of H.L.A. Hart, whether something is "law" depends on social facts—that is, it depends on whether the relevant community treats it as law.[328] Under that conception, it is significant that Congress, the Executive Branch, and legal scholars have long treated the issue of treaty termination as one of constitutional law and that they have viewed historical practice as relevant to determining the content of this law. The issue of treaty termination can therefore be distinguished from other customary conventions of U.S. constitutional practice that are not viewed as legal in character, such as (to take one example) the convention of senatorial courtesy for judicial appointments.[329]

To say that the issue of treaty termination is one of constitutional law does not mean that the law on this issue is fully settled. It is conceivable that the Senate or Congress at some point could assert itself on this issue, especially in a situation in which there was significant policy disagreement with the President's decision to terminate a particular treaty. It is even conceivable that the Senate or Congress could successfully force a President to back down, or at least to seek formal legislative approval for a termination. But the description of the constitutional law set forth above in subpart (IV)(A) is probably both the best prediction of likely future practice and also the best prediction of the position of the courts if they were at some point to intervene in this area. In any event, many issues of constitutional law are not entirely settled even after being resolved by the Supreme Court, especially if the Court is closely divided, and this fact is not viewed by itself as making constitutional law merely epiphenomenal.

Notwithstanding these points, the dynamic between Congress and the Executive Branch with respect to treaty termination is obviously intertwined with political, and not just legal, considerations. Political realities, such as the President's first-mover advantage over Congress and the tendency of members of Congress to support the President if he is of the same party, are likely to play a role in how the legal norms governing this issue develop. It is no coincidence that the most significant controversy over presidential termination of treaties occurred in connection with the Taiwan Treaty and associated recognition of mainland China, surrounding

328. *See supra* note 284 and accompanying text.

329. Scholarship on U.S. constitutional conventions has tended to mix together legal and nonlegal practices. *See, e.g.*, AKHIL REED AMAR, AMERICA'S UNWRITTEN CONSTITUTION 333–87 (2012); HERBERT W. HORWILL, THE USAGES OF THE AMERICAN CONSTITUTION 23–43 (1925). For a useful effort to distinguish between conventions based on whether they impose "thin" or "thick" obligations, see Vermeule, *supra* note 322, at 1186–91. *See also* Dorf, *supra* note 317, at 89 (distinguishing between entrenched practices and constitutionally normative practices).

which there was substantial policy disagreement. Legally normative conventions in this context are therefore affected, perhaps heavily, by politics.[330] But this does not mean that these conventions entirely collapse into politics.[331] Moreover, this blending of law and politics almost certainly describes constitutional law in other contexts as well.[332]

There are also likely still some legal constraints on presidential action in this area. It can reasonably be predicted, for example, that if the Senate conditioned its advice and consent to a treaty on senatorial approval of any termination, and a president later attempted to ignore that condition, there would be significant resistance, even by senators of the President's own party.[333] Moreover, this resistance would likely be framed and debated in legal terms. It is also likely that, when deciding whether to take such action, the President would be advised by lawyers who would consider past governmental practice in assessing the state of the law. None of this is to suggest that these considerations would be dispositive in presidential decision making, just that they would likely be a factor.

Despite these points, a focus on the role of historical practice in discerning the separation of powers almost inevitably mixes together internal and external perspectives on the law.[334] As noted, invocations of such practice have long been part of the internal legal argumentation in debates over treaty termination. At the same time, there are a variety of

---

330. In answer to a question, noted above, that was posed by Bruce Ackerman after President Bush announced that he was terminating the ABM Treaty, *see supra* note 248 and accompanying text, politics (both domestic and international) would likely operate as a significant constraint on unilateral presidential termination of something like the NATO pact or the UN Charter.

331. *See* 1 HOWARD GILLMAN, MARK A. GRABER & KEITH E. WHITTINGTON, AMERICAN CONSTITUTIONALISM: STRUCTURES OF GOVERNMENT 5 (2013) ("Rather than obsess about whether constitutionalism is pure law or pure politics, we should study the distinctive ways American constitutionalism blends legal and political considerations."). *See generally* Curtis A. Bradley & Trevor W. Morrison, Essay, *Presidential Power, Historical Practice, and Legal Constraint*, 113 COLUM. L. REV. 1097 (2013) (explaining why a connection between practice-based understandings of constitutional authority and political considerations does not make the understandings nonlegal).

332. For example, efforts to reconcile the political and legal aspects of the Supreme Court's exercise of constitutional judicial review are longstanding and include perhaps most famously ALEXANDER M. BICKEL, THE LEAST DANGEROUS BRANCH: THE SUPREME COURT AT THE BAR OF POLITICS (1962). *See also* Frank B. Cross, *The Ideology of Supreme Court Opinions and Citations*, 97 IOWA L. REV. 693, 695 (2012) ("Scholars today widely recognize that Supreme Court opinions are not purely legal but, to some degree, reflect the ideology of the Justices.").

333. An analogous issue concerns executive agreements: Despite the general rise of congressional–executive agreements in lieu of Article II treaties, the Senate has made clear at various times that it believes that significant arms-control agreements must be concluded as Article II treaties, and there has been successful bipartisan resistance in the Senate—framed in legal terms—to presidential efforts to do otherwise. *See* Bradley & Morrison, *supra* note 1, at 473–75 (describing this resistance).

334. *See* HART, *supra* note 284, at 89 (distinguishing between the "external" perspective of someone who is merely an observer of the rules of a social group and the "internal" perspective of someone who is a member of the group and "accepts and uses [the rules] as guides to conduct").

reasons to think that such practice also has an external effect on the development of the law relating to this issue, whether such law is interpreted by the courts or by nonjudicial actors. There is tension between these two accounts since the more that the account is external, the more that the law will seem epiphenomenal. It is at least plausible to think, however, that the internal and external accounts are interrelated, such that historical practice not only affects legal understandings but is also itself affected by such understandings.

Conclusion

Termination of treaties by the United States provides an important illustration of how historical practice can inform and even define the separation of powers. The constitutional text does not specifically address the issue, so practice has by necessity long played a central role in the legal analysis. Particularly in the nineteenth and early twentieth centuries—and then again in the controversy in the 1970s over the termination of the Taiwan Treaty—debates in Congress repeatedly focused on practice as relevant evidence of constitutional meaning. Legal advisers in the Executive Branch have also long emphasized the importance of practice in assessing the Constitution's distribution of authority over this issue. In addition to showing how practices can inform constitutional interpretation, the issue of treaty termination enriches our understanding of constitutional change. The twentieth-century shift towards a unilateral presidential power of termination was not the result of one particular controversy or period of deliberation, and it was not primarily driven by judicial decisions. Instead, the shift involved a gradual accretion of actions and claims by the Executive Branch combined with long periods of inaction by Congress. This account sheds light on some of the interpretive and normative challenges associated with a practice-based approach to the separation of powers.

* * *

Tax, Command . . . or Nudge?: Evaluating the New Regulation

Brian Galle*

This Article compares for the first time the relative economic efficiency of "nudges" and other forms of behaviorally inspired regulation against more common policy alternatives, such as taxes, subsidies, or traditional quantity regulation. Environmental economists and some legal commentators have dismissed nudge-type interventions out of hand for their failure to match the revenues and informational benefits taxes can provide. Similarly, writers in the law and economics tradition argue that fines are generally superior to nonpecuniary punishments.

Drawing on prior work in the choice-of-instruments literature, and contrary to this popular wisdom, I show that nudges may out-perform fines, other Pigouvian taxes, or subsidies in some contexts. These same arguments may also imply the superiority of some traditional "command and control" regulations over their tax or subsidy alternatives. I then apply these lessons to a set of contemporary policy controversies, such as New York City's cap on beverage portion sizes, climate change, retirement savings, and charitable contributions.

---

* Associate Professor, Boston College Law School. I am grateful for helpful suggestions and comments from Ben Alarie, Reuven Avi-Yorah, Ian Ayres, Michael Barr, Jake Brooks, Gerrit de Geest, Mirit Eyal-Cohen, John Friedman, Bill Gale, David Gamage, Ed Glaeser, Calvin Johnson, Louis Kaplow, Kyle Logue, Eric Muehlegger, Jacob Nussim, J.J. Prescott, Jim Repetti, Diane Ring, Dan Shaviro, Joel Slemrod, Michael Smart, Lauren Willis, Larry Zelenak, and attendees of presentations at Boston College, Chicago-Kent, Duke, Harvard, Michigan, and NYU Law Schools and at the Annual Meeting of the National Tax Association.

Introduction

It wasn't anyone's first choice. Diabetes, hypertension, and heart attacks were all on the rise in New York, and with them the City's costs of care.[1] The mayor's office explored a "sin tax" on soda and fatty foods, but food and beverage industry lobbyists went to Albany and blocked the tax in the state legislature.[2] So the City leaders searched for other ways to confront its citizens with the true costs of unhealthy lifestyles. They came up with the cap: No covered establishment could sell sugary beverages over 16 ounces in volume.[3] New York would become the City of Refills.

Critics were legion. Some complained that the city was setting up a "nanny state" to protect New Yorkers from themselves.[4] Others, perhaps

---

1. *Diabetes Among New York City Adults*, NYC VITAL SIGNS (N.Y.C. Dep't of Health & Mental Hygiene), Nov. 2009, at 1, 3; Notice of Public Hearing, Opportunity to Comment on the Proposed Amendment of Article 81 (Food Preparation and Food Establishments) of the New York City Health Code, found in Title 24 of the Rules of the City of New York (June 5, 2012), *available at* http://www.nyc.gov/html/doh/downloads/pdf/notice/2012/amend-food-establishments.pdf.

2. Michael M. Grynbaum, *New York Plans to Ban Sale of Big Sizes of Sugary Drinks*, N.Y. TIMES, May 30, 2012, http://www.nytimes.com/2012/05/31/nyregion/bloomberg-plans-a-ban-on-large-sugared-drinks.html?pagewanted=all.

3. *Id.*

4. Nick Gillespie, *3 Cheers for Coercive Paternalism-Or, Why Rich, Elected Officials Really Are Better than You*, REASON (Mar. 25, 2013, 10:41 AM), http://reason.com/blog/2013/03/25/3-cheers-for-coercive-paternalism-or-why; Katrina Trinko, *Soda Ban? What About Personal*

unaware of the legal maneuverings that preceded the cap, argued something of the opposite: if the City wanted to make beverages scarcer, it should have just imposed a tax.[5] Yet others doubted the cap would have any effect at all.[6] Despite the many skeptics, and as of this writing a set-back in the New York courts, the idea has proven popular in other municipalities, several of which are reportedly studying versions of their own.[7]

The beverage cap arrives after a decade of debate over "nudges" and other forms of behaviorally informed regulation. As Richard Thaler and Cass Sunstein, Ian Ayres, and others have ably summarized, evidence shows us that innocuous little speed bumps, like the nuisance of getting back up to fetch another cup of cola, or of filling out a form to start saving for retirement, can have surprising impact on individual behavior.[8] Choice architecture, the timing and context in which options are presented, matters.[9] That ice-cold Coke is a lot more tempting when we can see it fizzing sweetly beneath our thirsty lips than when it's stowed around the corner.[10] Time will tell, but there are now many good reasons to think the cap will work better than some have predicted.

Choice?, USA TODAY (Mar. 10, 2013, 5:40 PM), http://www.usatoday.com/story/opinion/2013/03/10/soda-ban-what-about-personal-choice-column/1977091/.

5. Sarah Kliff, *Why Ban Soda When You Can Tax It?*, WONKBLOG, WASH. POST (June 1, 2012, 1:16 PM), http://www.washingtonpost.com/blogs/wonkblog/post/why-ban-soda-when-you-can-tax-it/2012/06/01/gJQAT27E7U_blog.html; Nathan Sadeghi-Nejad, *NYC's Soda Ban Is a Good Idea, But a Tax Would Be Better*, FORBES, Sept. 13, 2012; Matthew Yglesias, *A Soda Tax Would Be Smart, Banning Big Cups Is Dumb*, SLATE (June 1, 2012, 10:49 AM), http://www.slate.com/blogs/moneybox/2012/06/01/a_soda_tax_would_be_smart_banning_big_cups_is_dumb.html; *see also* Robert H. Lustig et al., *The Toxic Truth About Sugar*, 482 NATURE 27, 28 (2012) (describing taxes as "the most . . . effective" policy for curbing excess sugar consumption).

6. Jacob Sullum, *The Benefit of Bloomberg's Big Beverage Ban*, REASON (June 20, 2012), reason.com/archives/2012/06/20/the-benefit-of-bloombergs-big-beverage-b.

7. Brock Parker, *Cambridge Mayor Proposes Limits on Soda Sizes: Idea Surprises City Council*, BOSTON GLOBE, June 19, 2012, http://www.bostonglobe.com/metro/2012/06/18/cambridge-consider-limiting-soda-sizes/j9PEQXqVRkbPWJBXALVYDP/discuss.html; Mark Segraves, *Some on D.C. Council Favor Restricting Sugary Drinks*, WTOP (Oct. 23, 2012, 7:22 AM), http://www.wtop.com/41/3088930/DC-Council-considers-restricting-sugary-drinks. In addition, many jurisdictions have already imposed some kind of tax on unhealthy food or drink. Alberto Alemanno & Ignacio Carreño, *Fat Taxes in the EU Between Fiscal Austerity and the Fight Against Obesity*, 2011 EUR. J. RISK REG. 571, 571–72; Michael F. Jacobson & Kelly D. Brownell, *Small Taxes on Soft Drinks and Snack Foods to Promote Health*, 90 AM. J. PUB. HEALTH 854, 855 tbl.1 (2000).

8. *See, e.g.*, IAN AYRES, CARROTS AND STICKS 3–44 (2010); RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 74–102 (rev. ed. 2009).

9. THALER & SUNSTEIN, *supra* note 8, at 83–102.

10. For overviews of the evidence that portion sizes affect consumption, see BRIAN WANSINK, MINDLESS EATING 17–19, 47–52 (2006), and Pierre Chandon, *How Package Design and Packaged-Based Marketing Claims Lead to Overeating*, 35 APPLIED ECON. PERSPECTIVES & POL'Y 7, 13–18 (2013).

Many other policy designers have taken those lessons to heart. Both the United States and the United Kingdom have recently launched government offices to expand the use of behaviorally informed regulation.[11] Efforts are already under way to cue families about their energy usage, to display healthy cafeteria foods in ways that are more appealing to kids, to make organ donations psychologically easier, and to make abortions more "informed" but emotionally more difficult.[12] Some noted economists have hinted recently at replacing the entire $125 billion in U.S. tax incentives[13] for retirement savings with a system in which individuals will have to opt out of saving rather than the most common current default, which is to opt in.[14] Proposals to rely on nudges now span the globe and virtually every regulatory domain.[15]

Despite the rapid policy evolution of nudges, debate over whether they *should* be used is less developed.[16] To be sure, there has been much debate over whether nudges escape the standard "paternalism" critique of government regulation. Proponents argue that nudges represent "libertarian paternalism" or are otherwise not coercive in the sense of traditional government regulation: People always retain the freedom to defy the government's preferences, and in many cases the costs of defiance are quite small.[17] Yet these arguments seem not to have assuaged the many anti-paternalism complaints about the nudge framework.[18]

---

11. RHYS JONES ET AL., CHANGING BEHAVIOURS: ON THE RISE OF THE PSYCHOLOGICAL STATE xii (2013); Cheryl K. Chumley, *White House Presses for Team of 'Nudge' Experts to Sway American Behavior*, WASH. TIMES, July 30, 2013, http://www.washingtontimes.com/news/2013/jul/30/white-house-presses-team-nudge-experts-sway-americ/.

12. *See, e.g.*, THALER & SUNSTEIN, *supra* note 8, at 1–2, 177–84, 231–39; Hunt Allcott et al., *Energy Policy with Externalities and Internalities* 33–34 (Nat'l Bureau of Econ. Research, Working Paper No. 17977, 2012), *available at* http://www.nber.org/papers/w17977; *A Nudge on a Hot Button Issue*, NUDGE (May 1, 2008), http://nudges.org/a-nudge-on-a-hot-button-issue-abortion/.

13. *See infra* note 247 and accompanying text.

14. Raj Chetty et al., *Active vs. Passive Decisions and Crowdout in Retirement Savings Accounts: Evidence from Denmark*, Q.J. ECON. (forthcoming) (manuscript at 43), *available at* http://www.rajchetty.com/index.php/papers-and-data/papers-and-data-listed-chronologically.

15. JONES ET AL., *supra* note 11, at vii–xii; *see also* On Amir & Orly Lobel, *Liberalism and Lifestyle: Informing Regulatory Governance with Behavioural Research*, 2012 EUR. J. RISK REG. 17, 17–18.

16. *See* Pierre Schlag, *Nudge, Choice Architecture, and Libertarian Paternalism*, 108 MICH. L. REV. 913, 919 (2010) (offering this critique of the nudge framework).

17. Colin Camerer et al., *Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism,"* 151 U. PA. L. REV. 1211, 1212 (2003); Cass R. Sunstein & Richard H. Thaler, *Libertarian Paternalism is Not an Oxymoron*, 70 U. CHI. L. REV. 1159, 1160–62 (2003) [hereinafter Sunstein & Thaler, *Libertarian Paternalism*].

18. *See* Claire A. Hill, *Anti-Anti-Anti-Paternalism*, 2 N.Y.U. J. L. & LIBERTY 444, 445–48 (2007); Jeffrey J. Rachlinski, *The Uncertain Psychological Case for Paternalism*, 97 NW. U. L. REV. 1165, 1219–25 (2003); Mario J. Rizzo & Douglas Glen Whitman, *The Knowledge Problem of New Paternalism*, 2009 BYU L. REV. 905, 909; Joshua D. Wright & Douglas H. Ginsburg, *Behavioral Law & Economics: Its Origins, Fatal Flaws, and Implications for Liberty*, 106 NW. U.

These debates are unlikely to end soon. Libertarians reject the claim that nudges do not reduce human freedom.[19] Nor do they see nudges as distinctive on that front: as the Harvard economist Ed Glaeser argues, taxes too permit individuals to defy the government at a price that sometimes is modest.[20] A soda tax can readily be avoided by skipping the sipping, and abstinence is easy for those without a sweet tooth.

Nudge proponents have for some reason left aside another strong potential argument. Even if nudges and their ilk are no less coercive than other forms of regulation, might they be preferable because they are economically more efficient?[21] If so, in what settings? Assuming policy makers can choose freely among both new and old regulatory instruments, what factors should they consider in deciding which to use?

This Article takes up these questions. I argue that nudges and other novel regulatory instruments can be evaluated using tools that are mostly already familiar in the economics of regulation. For example, at least since Gary Becker's seminal 1968 article, punishment theorists have argued over whether fines are a better enforcement tool than prison, with "shaming" and other collateral sanctions more recently joining the mix.[22] Environmental economists similarly debate the regulatory choice between taxes and other regulatory options, such as "command and control" regulation.[23]

Commentators overwhelmingly prefer taxes and other "price instruments" to regulation, and this would seem to be bad news for nudge defenders.[24] Both taxes and regulation distort private behavior. Taxes also bring in revenues, though, which can be used to improve the lives of those

---

L. REV. 1033, 1067–79 (2012); M. Ryan Calo, *Code, Nudge, or Notice?* 10–13 (Univ. of Wash., Legal Studies Research Paper No. 2013-04, 2013), *available at* http://ssrn.com/abstract=2217013.

19. *Cf.* On Amir & Orly Lobel, *Stumble, Predict, Nudge: How Behavioral Economics Informs Law and Policy*, 108 COLUM. L. REV. 2098, 2120–22 (2008) (suggesting that nudges are not value neutral); Gregory Mitchell, Review Essay, *Libertarian Paternalism is an Oxymoron*, 99 NW. U. L. REV. 1245, 1260–61 (2005) (arguing that libertarians would not be persuaded by Sunstein and Thaler's welfarist claims).

20. *See* Edward L. Glaeser, *Paternalism and Psychology*, 73 U. CHI. L. REV. 133, 135 (2006) (comparing a nudge to a psychic tax).

21. *See* Chetty et al., *supra* note 14 (noting that normative comparison of defaults and price instruments would be "a natural next step" for the literature).

22. Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 196–99 (1968); A. Mitchell Polinsky & Steven Shavell, *The Theory of Public Enforcement of Law*, *in* 1 HANDBOOK OF LAW & ECONOMICS 403, 407–08, 409 & n.10 (A. Mitchell Polinsky & Steven Shavell eds., 2007); Dan M. Kahan & Eric A. Posner, *Shaming White-Collar Criminals: A Proposal for Reform of the Federal Sentencing Guidelines*, 42 J.L. & ECON. 365, 366–68 (1999); Eric Rasmusen, *Stigma and Self-Fulfilling Expectations of Criminality*, 39 J.L. & ECON. 519, 520 (1996).

23. Gloria E. Helfand et al., *The Theory of Pollution Policy*, *in* HANDBOOK OF ENVIRONMENTAL ECONOMICS 249, 251, 287 (Karl-Göran Mäler & Jeffrey R. Vincent eds., 2003); Ian W.H. Parry & Wallace E. Oates, *Policy Analysis in the Presence of Distorting Taxes*, 19 J. POL'Y ANALYSIS & MGMT. 603, 608–10 (2000).

24. For elaboration of the points in this paragraph, *see infra* subparts II(A) and II(B).

who are inconvenienced by the regulatory policy. Glaeser claims that this same factor makes taxes more economically efficient than nudges.[25] Moreover, commentators argue that prices can reveal private information that other unpriced regulation—such as, perhaps, New York's beverage limits—cannot.[26]

As I will attempt to show, these claims about the superiority of taxes and tax-like instruments rest on overly simplifying assumptions, neglect possible innovations in governance structures, and do not apply fully for some nudges. For example, I argue that the supposedly unique informational benefits of price instruments can be captured through small-scale experiments, and this information used to run a larger regulatory scheme. I also argue that the revenue benefits of price instruments are considerably smaller than commentators assume in many settings; among other reasons, I show that nudges may have lesser negative impact on labor supply than their tax-like alternatives.

Further, by their nature prices usually require us to transfer resources from one party to another. Prior authors, including this one, have debated government's choice between two kinds of prices.[27] On the one hand are sticks, which can include taxes and other kinds of subjective changes for the worse.[28] On the other are carrots, which can include subsidies, or perhaps just relief from a currently expected cost. Although these instruments usually have very similar marginal effects, they also can differ importantly from one another in their impact on actors' preferences, in their incentives for future behavior, in their distributive consequences, and in their politics. Choosing between the two often requires balancing between these considerations.

Nudges and other transferless regulation, I'll argue, represent a hybrid or middle ground between sticks and carrots, and thus offer yet a third set of possible trade-offs.[29] For example, it is true that the beverage cap brings in no revenue for New York. But at the same time, it may also have better distributive consequences than a soda tax and avoid unwanted effects on

25. Glaeser, *supra* note 20, at 150.

26. *See infra* subpart II(A).

27. Giuseppe Dari-Mattiacci & Gerrit De Geest, *Carrots, Sticks, and the Multiplication Effect*, 26 J.L. ECON. & ORG. 365, 365–66 (2009); Brian Galle, *The Tragedy of the Carrots: Economics and Politics in the Choice of Price Instruments*, 64 STAN. L. REV. 797, 813–40 (2012); Jonathan Baert Wiener, *Global Environmental Regulation: Instrument Choice in Legal Context*, 108 YALE L.J. 677, 755–60 (1999).

28. For discussion of the points in this paragraph, see *infra* subpart II(C).

29. I don't mean to suggest that the three are mutually exclusive. *See* Michael P. Vandenbergh et al., *Regulation in the Behavioral Era*, 95 MINN. L. REV. 715, 719 (2011) (proposing "[p]airing price-[based] . . . approaches with behavioral approaches").

preferences and incentives.[30] Whether the cap is an attractive policy depends on the weights attached to these alternative consequences.

Similarly, as I show through additional examples, many other nudges can be compared directly to tort liability, taxes, or subsidies. Nudges might represent an important way forward for preventing climate change, where politics has stymied the best choices but the remaining traditional alternatives are mostly subsidies with crippling side-effects. At the same time, nudge enthusiasts may want to do some additional calculations before rushing to scrap the U.S. retirement-incentive system, as some noted authors have recently proposed.[31]

Part I of the Article sketches some background for readers new to these concepts. Part II lays the groundwork for later analysis by refining existing tools for comparing policy instruments. Part III employs my framework to compare nudges and other novel regulation to traditional alternatives. Part III also argues that the prevailing view of the superiority of corrective taxation over regulation may fail to consider some important factors. Part IV then applies these general principles to a series of (hopefully) illuminating examples, including soda and climate change as well as retirement savings, charity, and others.

I. Regulating Externalities: An Introduction

Modern economic theories of government regulation begin with the premise that markets sometimes fail.[32] Externalities are a classic example.[33] An externality, simply put, is a harm (negative externality) or benefit (positive externality) that affects someone other than the actor making an economic decision.[34]

---

30. Prior legal analyses of "sin taxes" have tended to emphasize instead philosophical questions about the government's role in regulation, Gary Lucas, Jr., *Saving Smokers from Themselves: The Paternalistic Use of Cigarette Taxes*, 80 U. CIN. L. REV. 693, 698–742 (2012), or questioned whether government should share in the profits from bad deeds, Andrew J. Haile, *Sin Taxes: When the State Becomes the Sinner*, 82 TEMP. L. REV. 1041, 1053–63 (2009). The substantial literature on the regulation of obesity does already raise some questions about the best methods for regulating but has not yet attempted to compare nudges to more traditional alternatives. *See* E. Katherine Battle & Kelly D. Brownell, *Confronting a Rising Tide of Eating Disorders and Obesity: Treatment vs. Prevention and Policy*, 21 ADDICTIVE BEHAVIORS 755, 762 (1996); Tom Marshall, *Exploring a Fiscal Food Policy: The Case of Diet and Ischaemic Heart Disease*, 320 BRIT. MED. J. 301, 301 (2000); Katherine Pratt, *A Constructive Critique of Public Health Arguments for Antiobesity Soda Taxes and Food Taxes*, 87 TUL. L. REV. 73, 114–39 (2012); Jeff Strnad, *Conceptualizing the "Fat Tax": The Role of Food Taxes in Developed Economies*, 78 S. CALIF. L. REV. 1221, 1294–1322 (2005); Stephen D. Sugarman & Nirit Sandman, *Fighting Childhood Obesity Through Performance-Based Regulation of the Food Industry*, 56 DUKE L.J. 1403, 1429–90 (2007).

31. Chetty et al., *supra* note 14.

32. JONATHAN GRUBER, PUBLIC FINANCE AND PUBLIC POLICY 3 (3d ed. 2011).

33. *Id.* at 123.

34. *Id.* at 124–28.

In general, the goal of regulation is neither to eliminate negative nor to produce boundless quantities of positive externalities but rather to achieve what might be called the optimal level of externality.[35] Eliminating even the worst pollutants is costly.[36] Should government bankrupt coal producers, or is there a way to balance clean air against the costs of achieving it? On the positive externality side, everyone might agree that charity is beneficial. But should government spend millions to clothe or educate one more child?

Economists typically answer these kind of balancing questions using marginal analysis.[37] Under this approach, the policy maker asks herself, "on the margin—that is, for the very next unit of good or bad produced—what is the harm or benefit of that one unit for *everyone in society*?" We might therefore call this the "marginal social damage," (MSD) in the case of a negative externality and "marginal social benefit" (MSB) for a positive one. She then compares this harm or benefit against the marginal costs to the producer. If the producer's private marginal cost is greater than the marginal social damage, it doesn't pay, on net, to prevent the damage: counting the producer's losses, society would lose by forcing the producer to avoid the externality.[38]

To see this graphically, consider Figure 1.

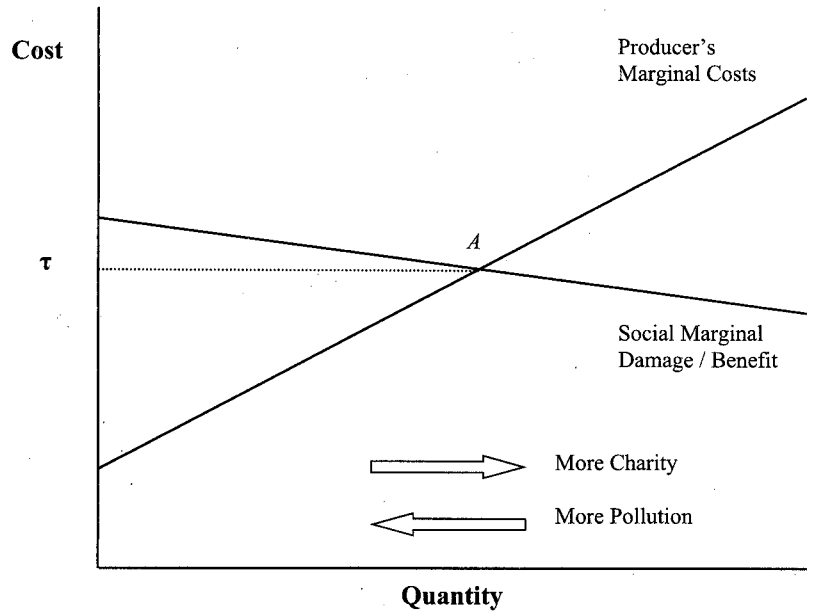
35. *Id.* at 139; Helfand et al., *supra* note 23, at 253.

36. GRUBER, *supra* note 32, at 122–23.

37. *Id.* at 124.

38. Note, importantly, that for simplicity we are assuming here that we should count the costs and benefits for the producer and everyone else equally. That's a controversial proposition, but I'll leave it aside here for ease of exposition.

Figure 1: Optimal Externality Production



In Figure 1, the upward-sloping line represents the marginal cost curve for the externality producer: as we trace the line rightwards, each additional unit of pollution reduction (say, one ton less of carbon) or charitable output (say, another bed in a homeless shelter) is costlier to achieve.[39] The downward-sloping line is the marginal social benefit curve: each unit is slightly less beneficial than the last.[40] At point *A*, the two lines intersect. This is the optimal point.[41] Anywhere to the right of *A*, the costs of charity or pollution reduction outweigh the benefits. To the left, we've left cost-effective improvements on the table.

We could imagine a few ways of achieving production at this level *A*. If government knew the shapes of the two curves, it could calculate the

39. This reflects the likelihood that firms will undertake the cheapest efforts first and then have to work harder and harder to achieve further milestones. For instance, at some point, adding more beds means building a new building.

40. Again, diminishing marginal utility is a standard assumption here. We probably house the neediest persons first, and at some point we're offering shelter space to Bill Gates.

41. I'm simplifying here for the sake of exposition. A more rigorous approach to setting the optimal quantity would also account for other factors that might affect the efficiency of the regulation. For example, if the regulation imposes costs, and the expectation of those costs changes behaviors other than the production of the externality—for example, distorts consumer choices among products—the ideal regulation might balance disruption of these expectations against pollution control. *See* Helmuth Cremer et al., *Externalities and Optimal Taxation*, 70 J. PUB. ECON. 343, 346 (1998).

quantity of output at *A* and simply mandate that producers achieve it, with jail for those who refuse.

Another approach is to set a price for producers. In the case of pollution, government could impose a fee or tax on each unit of carbon in an amount equal to the producer's marginal cost at point *A*; this price is labeled *tau* in Figure 1. For producers whose costs of eliminating the next unit of carbon are below *tau*, they will eliminate it, saving themselves *tau* minus their cost. For producers whose costs are above *tau*, they will simply emit the carbon and pay the tax. Thus, just as with the mandate, rational producers should produce exactly the amount of carbon at point *A*. Or, similarly, government could pay producers to eliminate carbon or produce charity. Once more, if the government offers a price *tau*, only producers who can fill a shelter bed for less than *tau* will take the offer.

Economists often call the first of these approaches "quantity regulation"[42] and the second two "price instruments."[43] Lawyers may be more familiar with the similar divide between what Calabresi and Melamed termed "property rules" and "liability rules."[44]

Most commentators strongly favor price instruments over quantity regulation, except in settings where special administrative considerations make prices impractical.[45] As Kaplow and Shavell show, prices can be used to duplicate most of the features of mandates.[46] Prices provide vital information to the government that regulation supposedly does not, as we'll see in more detail shortly.[47] Further, prices are said to provide for revenues that the government can use for other projects, while regulations do not.[48]

Glaeser's critique of nudges and similarly novel behavioral forms of regulation is typical in this regard. Glaeser argues that both taxes and nudges create economic distortions, but only taxes bring in money to help

---

42. GRUBER, *supra* note 32, at 140.

43. THOMAS STERNER, POLICY INSTRUMENTS FOR ENVIRONMENTAL AND NATURAL RESOURCE MANAGEMENT 148, 214–15 (2003).

44. Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092 (1972); *see also* Louis Kaplow & Steven Shavell, *Property Rules Versus Liability Rules: An Economic Analysis*, 109 HARV. L. REV. 713, 715 (1996).

45. GRUBER, *supra* note 32, at 140; Maureen L. Cropper & Wallace E. Oates, *Environmental Economics: A Survey*, 30 J. ECON. LITERATURE 675, 686 (1992); Don Fullerton et al., *Environmental Taxes*, *in* DIMENSIONS OF TAX DESIGN: THE MIRRLEES REVIEW 423, 429 (James Mirrlees et al. eds., 2010); Cameron Hepburn, *Regulation by Prices, Quantities, or Both: A Review of Instrument Choice*, 22 OXFORD REV. ECON. POL'Y 226, 228–29 (2006). As an example of a special consideration, price instruments may be riskier than quantity regulation when the marginal social damage curve is steep but its exact shape is uncertain, GRUBER, *supra* note 32, at 143–46, and the policy maker cannot sharply vary the tax rate to account for this risk.

46. Louis Kaplow & Steven Shavell, *On the Superiority of Corrective Taxes to Quantity Regulation*, 4 AM. L. & ECON. REV. 1, 7–10 (2002).

47. *Id.* at 4.

48. *E.g.*, Helfand et al., *supra* note 23.

offset those losses.[49] Perhaps graphic images of the harms of cigarette smoking printed on the sides of packs would be repulsive enough that the smokers switch to cigarillos or pipe tobacco, which are nearly as harmful but which they enjoy less. Or perhaps some smokers cannot quit, but also suffer added pain as a result of the imagery.[50] Further, unlike a cigarette tax, the graphic images don't bring in any revenues that could be used to improve the lives of smokers or anybody else. To take another example, workers who are defaulted into a savings program, who are unwilling to pay the costs of the opt-out mechanism, and who genuinely would prefer not to save are worse off than in the absence of the nudge.[51] A number of other economists have recently made a similar point about the preferability of a carbon tax over other regulatory alternatives: regulations change consumption patterns, creating deadweight loss, but bring in no offsetting dollars.[52]

Glaeser's point echoes a much older debate over the most efficient form of punishment for crimes. Becker, and later Polinsky and Shavell, have argued that in many situations fines are superior to imprisonment.[53] Both reduce the utility of the offender. The fines, though, can be used to transfer that loss to someone else, resulting in greater overall social welfare.[54]

Over the remainder of the Article I want to first flesh out, and then question, many of these well-established assumptions about the superiority of prices to other regulatory alternatives. As we'll see, some long-standing claims may not hold up to close scrutiny, especially once we factor in some of the unique aspects of the newest regulatory alternatives.

---

49. Glaeser, *supra* note 20, at 135, 150; *see also* Lucas, *supra* note 30, at 726–30 (observing that optimal cigarette tax rates differ from person to person because of heterogenity among smokers); Mitchell, *supra* note 19, at 1268, 1274 (explaining that the cost of increasing benefits to irrational persons will often be borne by rational ones); Rizzo & Whitman, *supra* note 18, at 960–61 (discussing problems of over- and under-inclusion that arise under one-size-fits-all tax models).

50. Andrew Caplin, *Fear as a Policy Instrument*, *in* TIME AND DECISION: ECONOMIC AND PSYCHOLOGICAL PERSPECTIVES ON INTERTEMPORAL CHOICE 441, 442, 452 (George Loewenstein et al. eds., 2003); *see* Lee Anne Fennell, *Willpower Taxes*, 99 GEO. L.J. 1371, 1415 (2011) (making this point about imperfectly targeted nudges generally).

51. *See* Jeffrey J. Rachlinski, *Cognitive Errors, Individual Differences, and Paternalism*, 73 U. CHI. L. REV. 207, 224–25 (2006); *cf.* Schlag, *supra* note 16, at 917 (noting that nudges may have more dramatic effects on behaviors than command-and-control regulation).

52. Helfand et al., *supra* note 23, at 287; Ian W.H. Parry et al., *When Can Carbon Abatement Policies Increase Welfare? The Fundamental Role of Distorted Factor Markets*, 37 J. ENVTL. ECON. & MGMT. 52, 52 (1999); *see also* Ian Ayres, *Regulating Opt-Out: An Economic Theory of Altering Rules*, 121 YALE L.J. 2032, 2091 (2012). However, Ayres also notes the potential targeting advantages of using what he calls "sticky defaults." Ayres, *supra*, at 2091–92.

53. Becker, *supra* note 22, at 193–99; Polinsky & Shavell, *supra* note 22, at 407–20.

54. Becker, *supra* note 22, at 180; Polinsky & Shavell, *supra* note 22, at 408. Becker credits an early version of this point to Bentham. Becker, *supra* note 22, at 193 n.40.

II. Policy Instruments in Three (or More) Dimensions

In order to assess the new evolving set of policy tools, it may be useful to situate them in the context of more familiar mechanisms. The new tools, I'll argue, share many features in common with those we already know. To see those features and how they interact, in this Part I'll attempt to break down the potential policy toolkit into its component parts. Since each part has distinctive features—adds to the treasury or draws from it, redistributes funds or doesn't, and so on—when we rebuild our tools from their components we'll be better able to see the consequences of the tool we choose. Other authors have made some of these distinctions before, so not all the components will be wholly new. But the overall picture, and its lessons, are novel.

A. Price vs. Quantity and Refinements

As we just saw, the distinction between price instruments and quantity regulation is a fundamental divide. In the simplest terms, price instruments are usually distinguished by the fact that they involve money transfers. They can include measures familiar from first-year law courses, such as tort liability, as well as sin taxes, sometimes called "Pigouvian" taxes after the economist most strongly associated with them.[55] So, for example, speed and blood-alcohol limits are common quantity regulations aimed at the dangers of the road, while tort lawsuits and tolls are price instruments aimed at the same problem.

Though the price/quantity dichotomy is widely accepted, and appears in virtually any textbook on the economics of regulation,[56] it's overly simplistic in a couple of different ways. Most importantly for my purposes, it collapses what ought to be four categories into just two. Once more, an archetypical price does two things that regulations usually don't: prices reveal information about the subjective valuation of the party who chooses to pay, and they result in the transfer of resources from one party to another.[57] As we saw, both these differences have important policy implications.

With some reflection, though, we can see that not all transfers reveal information, nor do all prices result in net transfers. Eminent domain without compensation provides an example of the first: the government takes title to property from a private landowner and gives it to someone else. While the government can perhaps investigate the subjective cost of the taking for the original landowner, the taking itself does nothing to reveal

55. GRUBER, *supra* note 32, at 135.

56. *E.g.*, *id.*

57. *See supra* text accompanying notes 46–54.

the landowner's preferences. Replevin and specific performance are other familiar examples.

Rebates supply an instance of prices without transfers. Suppose that we impose a soda tax and then rebate the proceeds equally to each household. Families that consume the household average amount of soda will be no richer or poorer than they were before. But each individual retains a *marginal* incentive to cut back on pop because if they consume more than average they will lose money on net. Each additional cup still costs an extra nickel or dime. In effect, we can observe the strength of people's preferences without transferring much between them. Louis Kaplow extends this idea, proposing that all regulation should be evaluated as if it were enacted together with a perfectly offsetting tax or rebate; under his scheme, all prices are on net transferless.[58]

Therefore, I'll subdivide the price/quantity categories. We have policies that transfer value and those that are what I'll call "transferless." And then we have policies that reveal information about subjective cost or valuation, which we can call "price," and those that don't, which I'll just call "priceless."

B. Public vs. Private

If policy instruments transfer value, it may be important to know who are the winners and losers. To simplify a bit, I'll call the two main possibilities "private" and "public." Tort proceeds are paid directly to victims, while tax revenues usually end up in the public fisc. There can also be middle ground between the two categories. If tort awards are subject to an income tax, then a portion of the judgment ends up in public hands. Sin taxes may be set aside for the benefit of a particular group, as in the case of the federal gasoline tax, which is mostly used for road improvements.[59] In the case of carrots, we can think of "public" and "private" as the question of who pays, rather than who benefits.[60] For instance, patents resemble tax credits for research and development, but the

58. LOUIS KAPLOW, THE THEORY OF TAXATION AND PUBLIC ECONOMICS 13–34 (2008). I do not disagree, and even agree that such perfect offsets might often be theoretically ideal. My goal is only to consider the second-best outcomes in the absence of optimal offsets. That is, I analyze the implementation of the price instrument in isolation from any such offsets, which so far have not been observed in practice.

59. STAFF OF JOINT COMM. ON TAXATION, 111th CONG., JCS-I-II, PRESENT LAW AND BACKGROUND INFORMATION ON FEDERAL EXCISE TAXES 2–10 (Joint Comm. Print 2011), *available at* https://www.jct.gov/Publications.html?func=startdown&id=3721.

60. That is not to say that governments cannot be subject to carrots and sticks. But that subject deserves more detailed treatment than I can offer here.

costs of the patent are borne mostly by consumers, while tax credits empty public coffers.[61]

Allocating winners and losers can dramatically affect both the economics and politics of a policy as prior commentary has recognized.[62] Economically, in most cases, it is more efficient to spread the costs of a policy as widely as possible.[63] Whether labeled a "tax" or not, the costs of paying for someone else's benefits distort our behavior, and the welfare loss from that distortion grows exponentially.[64] To see why, think of a $1 tax on iPhones. Not many people will switch to Android for $1, and those that do really did not value Apple very highly, anyway. At a $50 tax, many people are switching, and now we are changing the behavior of people who were deeply bonded with Siri. And yet, the government raises not a dollar from anyone who switches, which is why economists term the distortion "deadweight loss." If we can spread the cost of a policy across more payors, we can lower the costs each of them face, reducing deadweight losses.

Transfers also have important effects on incentives. For instance, we may prefer that fines be paid to the public, rather to the victims of crimes, in order to give victims the correct set of incentives to mitigate the harm they suffer.[65] As we will see, many commentators assume that carrots give potential beneficiaries incentives to start doing bad deeds, so that they will be paid to stop. But as I have pointed out elsewhere, that is less true if it turns out that the beneficiaries also mostly pay for their own carrots.[66]

Transfers are not the only way a policy can be public or private. The power to initiate an enforcement action, control over how it proceeds, the costs of administering and enforcing it—all these can be divided between public and private. These factors are significant at times, as well; "corrective justice" tort scholars point to private control, rather than the recipients of the money judgment, as a key feature of the tort system.[67]

---

61. *See* Michael Abramowicz, *Perfecting Patent Prizes*, 56 VAND L. REV. 115, 200–07 (2003); Nancy Gallini & Suzanne Scotchmer, *Intellectual Property: When Is It the Best Incentive System?*, *in* 2 INNOVATION POLICY AND THE ECONOMY 51, 54 (Adam B. Jaffe et al. eds., 2002).

62. For a review, see Galle, *supra* note 27, at 809–12, 840–45.

63. *See* A.B. Atkinson & J.E. Stiglitz, *The Design of Tax Structure: Direct Versus Indirect Taxation*, 6 J. PUB. ECON. 55, 55–64 (1976) (explaining benefits of uniform, broad-based taxation).

64. *Id.* at 56 (arguing that basing a tax on characteristics under the control of an individual will distort the economy).

65. Kaplow & Shavell, *supra* note 44, at 738; Donald Wittman, *Liability for Harm or Restitution for Benefit?*, 13 J. LEGAL STUD. 57, 67–68 (1984).

66. Galle, *supra* note 27, at 826–27.

67. *See, e.g.*, John C.P. Goldberg, *The Constitutional Status of Tort Law: Due Process and the Right to a Law for the Redress of Wrongs*, 115 YALE L.J. 524, 532–58 (2005) (arguing that the basis of tort law is private property rights and the notion that individuals have personal rights to be redressed).

Private control over a nudge regime might complicate regulatory goals.[68] For the most part, though, I will not examine those questions closely here.

C. *Stick vs. Carrot*

Once policy makers decide to rely on a price instrument, they have a choice between rewarding or penalizing, between carrots and sticks. Both options have similar effects on the marginal incentives of externality producers.[69] Whether producers are rewarded, or nonproducers fined, giving an additional dollar saves donors money relative to not giving.[70] Each instrument can be priced so that the marginal cost of an additional unit of production is equal to the marginal damage suffered by society, so that in effect the producers internalize the full social cost of their decision.[71] However, the two mechanisms vary in a number of other important ways. As I have described before, which option is the better choice for a particular policy depends largely on these other factors.[72]

Sticks are, except in unusual circumstances, the more efficient tool for reining in the social overproduction of some negative-externality-laden good.[73] Sticks earn the government money, while carrots drain the treasury, wasting hard-won tax revenues.[74] Revenue is critical because raising taxes is costly: in addition to paying the tax, many people will also change their behavior to minimize taxes, causing deadweight loss.[75] In addition, carrots give producers more resources to create the unwanted good. Similarly, in many cases, as individuals get wealthier, they demand more of the undesirable product, a phenomenon known as the "income effect."[76] Carrots are also wasteful if producers plan to cut back on their activities anyway. And overproducers who know they will be paid to curtail their activities in the future have an incentive to begin overproducing, while the opposite is true of sticks.

---

68. Michael S. Barr et al., *The Case for Behaviorally Informed Regulation*, *in* NEW PERSPECTIVES ON REGULATION 27, 35–39 (David Moss & John Cisternino eds., 2009).

69. Helfand et al., *supra* note 23, at 277–78.

70. *Id.* at 278.

71. *Id.*

72. Galle, *supra* note 27, at 809–13. For a complementary framework that sometimes reaches different conclusions from mine, see Gerrit De Geest & Giuseppe Dari-Mattiacci, *The Rise of Carrots and the Decline of Sticks*, 80 U. CHI. L. REV. 341, 360–73 (2013).

73. For development of the points in this paragraph, see Galle, *supra* note 27, at 813–31.

74. The revenue benefit of sticks depends, however, on some assumptions about how the revenues are deployed. A. Lans Bovenberg & Lawrence H. Goulder, *Environmental Taxation and Regulation*, in 3 HANDBOOK OF PUBLIC ECONOMICS 1471, 1497–507 (Alan J. Auerbach & Martin Feldstein eds., 2002). For development of this point, see *infra* section III(B)(1).

75. GRUBER, *supra* note 32, at 591–601.

76. *Cf. id.* at 36 (defining income effect as higher prices causing a consumer to buy less when all else is held constant). For example, poorer commuters may take the bus, while richer ones may prefer to drive.

In contrast, carrots are more defensible for encouraging the production of a good with positive externalities, where we would expect social underproduction.[77] In that case, the fact that carrot recipients have more resources is desirable, since we want them to produce or demand more of the good.[78] On the other hand, it is still the case that the expectation of future carrots has unwanted incentive effects, encouraging producers to delay producing the good until the government agrees to pay them. And carrots remain costlier, especially when factoring in the possibility that some might altruistically produce the good without subsidy. So though carrots are less clearly dominated by sticks in the positive externality setting, there remains a question whether they are worth the cost.

Let me emphasize the limits of what this "choice of instruments" kind of analysis can accomplish. The goal is to measure the relative efficacy of each choice, given an arbitrary baseline: our world looks like this, what should we do now? So the claim is not that sticks are always efficient, only that they are usually *more* efficient than carrots, all else equal.[79]

D. Ex ante/Ex post

Another potentially important policy dichotomy is the timing of the policy lever, though I will not emphasize it much here because its parameters are already well explored in the literature.[80] Some incentives pay off before the externality producer acts, and some take effect afterwards. For example, zoning laws restrict development before it results in unwanted burdens on neighbors, while nuisance suits impose liability after the damage has begun.

In many cases time is irrelevant. As any 1L knows, *ex post* liability regimes like tort and criminal law assume that rational actors will take account of their expected future costs when planning their behavior.[81] So *ex post* is effectively the same as *ex ante*, at least for rationally-forward-

77. *Id.* at 43–53 (noting that an unregulated market tends to underproduce goods with positive externalities).

78. Galle, *supra* note 27, at 832.

79. *See* Helfand et al., *supra* note 23, at 270 (noting that a goal of economic analysis is usually to identify the welfare effects of a policy in comparison to its alternatives); *cf.* Daniel Shaviro, *The Minimum Wage, the Earned Income Tax Credit, and Optimal Subsidy Policy*, 64 U. CHI. L. REV. 405, 415 (1997) (explaining that departures from the status quo can be analyzed without attributing any special normative status to existing rules).

80. For more extensive treatment, see Brian Galle, *Myopia, Fiscal Federalism, and Unemployment Insurance: Time to Reform UI Financing* 5–18 (Bos. Coll. Law Sch., Legal Studies Research Paper No. 265, 2012), *available at* http://ssrn.com/abstract=2031728.

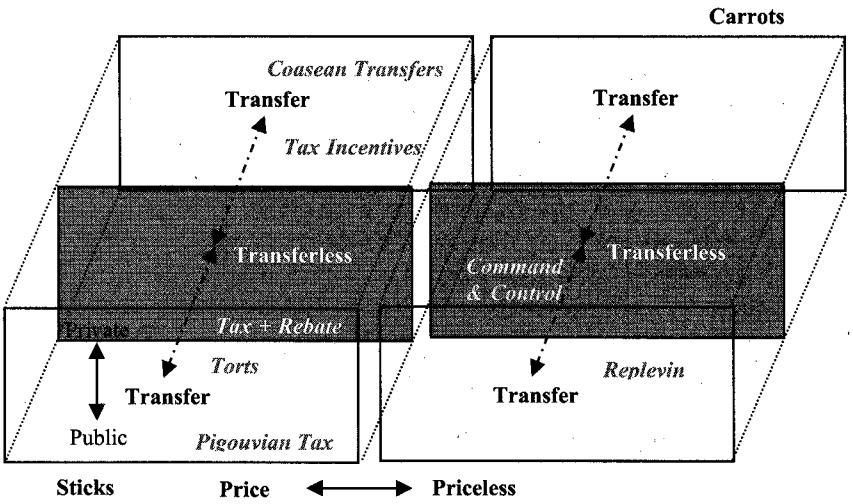
81. Christine Jolls, *On Law Enforcement with Boundedly Rational Actors*, *in* THE LAW & ECONOMICS OF IRRATIONAL BEHAVIOR 268, 272 (Francesco Parisi & Vernon L. Smith eds., 2005).

looking actors.[82] Information, myopia, and liquidity concerns can all disrupt this equivalence.[83]

E. A Summary So Far

By this point we already have many of the tools we need to compare, say, soda taxes against the New York beverage limits. We now see that many of the criticisms of the limits are really complaints that New York chose a priceless, transferless, and mostly private policy over a priced, public transfer. More generally, we can describe many policy instruments as a combination of the factors we've seen, as in Figure 2. Figure 2 actually does not depict the *ex ante*/*ex post* dimension, but readers can think of this page as the *ex ante* boxes, and then simply don their imaginary 3-D glasses and picture an identical set of boxes extending into the third dimension to represent *ex post*. The examples given, though, mix *ex ante* and *ex post*.

Figure 2: Components of Conventional Policy Instruments



82. Of course, time is money, so *ex post* liability must be greater at the time it is imposed than a comparable *ex ante* incentive, such as a Pigouvian tax; the present discounted value of each alternative should be identical *ex ante*.

83. Galle, *supra* note 80, at 15–18.

F. The New Regulation: Nudges, Defaults, and "Surprising" Effects

While we have seen much of the story, there remain important senses in which something like New York's soda policy differs from traditional regulation. As readers likely know, Sunstein and Thaler call these kinds of policies "nudges," and offer a long list of examples; for instance, they suggest painting roads to encourage more cautious driving and if that fails making organ donation the default choice on drivers' licenses.[84]

1. What's a Nudge?—For Sunstein, Thaler, and other proponents, two factors distinguish their ideas from more familiar approaches: nudges are policies whose effects are "surprising" and "asymmetric."[85] For example, classic rational-choice economic theory predicts that the default savings rate chosen by our employers should not affect how much we choose to save for retirement. Filling out a sheet of paper to change our plan takes ten minutes, and might be worth tens of thousands of dollars in the long term. Yet much evidence suggests that defaults matter a great deal: that is the surprising part.[86] The "asymmetry" is that the impact of the form is not uniform; some people are much more affected by having to fill out a form than others.[87]

A nudge, then, replaces traditional motivators, such as cash or jail time, with surprising and asymmetric incentives. If we know that individuals are slow to switch away from a default choice initially made for them, government can use defaults in place of commands.[88] Similarly, minor obstacles such as having to fill out a form or wait in a line can, at times, replace prescriptive regulation.[89] To the extent that the framing and presentation of information influences how we choose, government can

---

84. *See* THALER & SUNSTEIN, *supra* note 8, at 83–102, 172–84, 231–39, 257–68. See Russell Korobkin, *Libertarian Welfarism*, 97 CALIF. L. REV. 1651, 1662–64 (2009), for a pithy summary of the available tools.

85. *See* THALER & SUNSTEIN, *supra* note 8, at 85–86, 252–54; Camerer et al., *supra* note 17, at 1222 (arguing that paternalistic policies are justified when there is asymmetric information).

86. Shlomo Benartzi & Richard H. Thaler, *Heuristics and Biases in Retirement Savings Behavior*, J. ECON. PERSP., Summer 2007, at 81, 83–84.

87. *See id.* at 100–01; Chetty et al., *supra* note 14 (manuscript at 37–38).

88. Nudge proponents have mostly focused on internalities, but some scholars have extended their work to externalities or other regulatory goals as well. *See, e.g.*, Ayres, *supra* note 52, at 2086 (arguing that nudges can be used in the context of negative externalities); Anuj C. Desai, *Libertarian Paternalism, Externalities, and the "Spirit of Liberty": How Thaler and Sunstein are Nudging Us Toward an "Overlapping Consensus,"* 36 LAW & SOC. INQUIRY 263, 270 (2011) (discussing choice architecture in the context of negative externalities); Korobkin, *supra* note 84, at 1653 (stating that nudges can be used in the context of public goods by encouraging greater production of public goods); Matthew A. Smith & Michael S. McPherson, *Nudging for Equality: Values in Libertarian Paternalism*, 61 ADMIN. L. REV. 323, 335–39 (2009) (urging the use of the nudge concept to promote equality).

89. Brian Galle & Manuel Utset, *Is Cap-and-Trade Fair to the Poor? Shortsighted Households and the Timing of Consumption Taxes*, 79 GEO. WASH. L. REV. 33, 84–87 (2010).

influence the public towards more desirable outcomes without the need for law enforcement.[90]

These two features have important normative bite for nudge proponents. They argue that the objective burdens of overcoming a nudge in many cases are small.[91] Of course, in the moment that individuals face the nudge—when they are waiting on hold as *The Girl from Ipanema* plays tinnily through their phone's speaker—its costs appear too large to bear. So the claim that nudges are different depends partly on an assumption about the proper measure of individuals' utility: evidently we should count costs and benefits according to the perspective the individual would take in a temporally remote, "reflective" setting.[92] Seen from this point of view, the cost of waiting on the phone for a few minutes should look tiny.

Secondly, and less controversially, nudges differ from standard regulation in their ability to more closely approximate people's real preferences. Traditionally, critics of regulation have claimed that uniform government rules aimed at correcting people's own mistakes will necessarily impose a one-size-fits-all regime, forcing some people to change for the worse.[93] Social security, for instance, can be criticized as a form of forced savings that may reduce the subjective welfare of those who prefer to consume all their income immediately.[94]

Nudge defenders argue that asymmetry mitigates this problem because those who feel strongly about their own choices can easily overcome the government's default.[95] Although defenders acknowledge that for some people nudges can be hard to overcome, they suggest that asymmetric regulation is most defensible in those cases where the personality traits that make nudges tough to fight are the same traits that produce the behaviors the government is combating.[96] Impatient people won't opt out of default savings plans, but the impatient are also the most likely to be saving too

90. Camerer et al., *supra* note 17, at 1230–37.

91. THALER & SUNSTEIN, *supra* note 8, at 252–54; Camerer et al., *supra* note 17, at 1219, 1222; *see also* Ayres, *supra* note 52, at 2087 (describing costs of sticky defaults as "intermediate" between commands and free contract).

92. THALER & SUNSTEIN, *supra* note 8, at 12; Sunstein & Thaler, *Libertarian Paternalism*, *supra* note 17, at 1191. A more developed version of this argument is Eyal Zamir, *The Efficiency of Paternalism*, 84 VA. L. REV. 229, 237–84 (1998). *But cf.* Camerer et al., *supra* note 17, at 1253–54 (suggesting that nudges are preferable to traditional paternalistic regulation because of "uncertainty" about whether consumer choices are really mistakes).

93. *See, e.g.*, Richard A. Epstein, Exchange, *The Neoclassical Economics of Consumer Contracts*, 92 MINN. L. REV. 803, 806–07 (2008); Rizzo & Whitman, *supra* note 18, at 909–10.

94. Theodore R. Groom & John B. Shoven, *Deregulating the Private Pension System*, *in* THE EVOLVING PENSION SYSTEM 123, 126 (William G. Gale et al. eds., 2005).

95. Ayres, *supra* note 52, at 2091–92; Camerer et al., *supra* note 17, at 1222.

96. Camerer et al., *supra* note 17, at 1225–26; *see also* Allcott et al., *supra* note 12, at 2, 23 (discussing the correlation between susceptibility to defaults and propensity to suffer harm).

little. Therefore, even if nudges *are* costly for some people, these are generally the people who on net benefit from that cost.[97]

While prior nudge proposals generally do not rely on dollars, we can also imagine some surprising price instruments. When the metro D.C. area adopted a five-cent tax on plastic grocery bags, grocery-bag usage plummeted.[98] Consumers switched to alternatives that often were more expensive than plastic, even after accounting for the five-cent savings.[99] So it seems that it was not price alone that made the "bag tax" so effective. Commentators suggest that the tax might have provided new information to consumers about the harms of plastic bags.[100] Or it might have triggered a "norms cascade" in which it became shameful to be one of *those* people—the people who did not care about whether their trash would strangle a hapless sea bird.[101] Other monetary incentives may be especially effective because of the way they are timed and framed.[102]

In short, it seems as though we could expand our earlier set of boxes to include the possibility of surprising forms of any kind of policy tool. I illustrate this expanded universe with Figure 3, below. For some combinations, real-world examples are scarce, suggesting some new frontiers of policy experiments.

97. Camerer et al., *supra* note 17, at 1222.

98. Tatiana A. Homonoff, Can Small Incentives Have Large Effects? The Impact of Taxes Versus Bonuses on Disposable Bag Use 3, 6 (Mar. 27, 2013) (unpublished manuscript), *available at* https://appam.confex.com/appam/2013/webprogram/Paper6746.html.

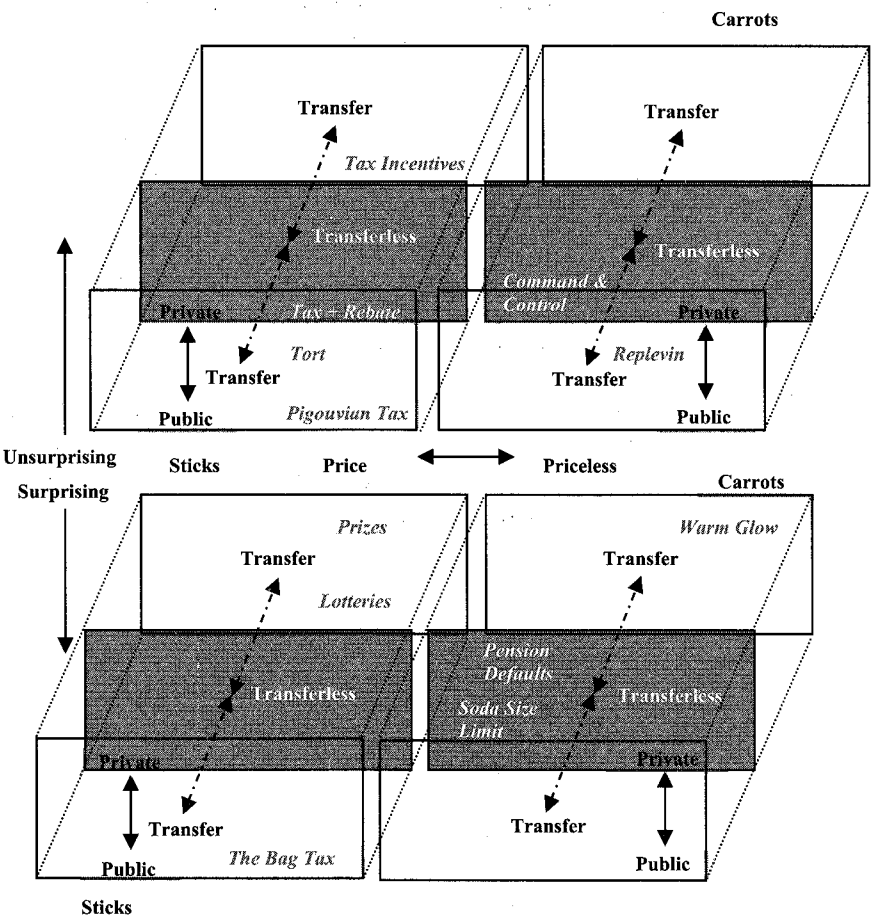
99. *Id.*

100. *Id.*

101. *Cf.* Cass R. Sunstein, *Social Norms and Social Roles*, 96 COLUM. L. REV. 903, 947–50 (1996) (explaining the concept of norms cascade and the potential government role in it).

102. AYRES, *supra* note 8, at 43, 53; *see* Kevin C. Volpp et al., *Financial Incentive-Based Approaches for Weight Loss*, 300 J. AM. MED. ASS'N 2631, 2636 (2008) (describing "supercharged" financial incentives).

Figure 3: Components of Policy Instruments



2. Why Do Nudges Work?—In order to do any serious policy analysis of surprising instruments, we'll need to understand what makes them effective. As I'll show in a bit, the psychological mechanisms that underlie different forms of nudge can translate into sharply varying social welfare implications for their widespread usage.

Nudges depend on humans' psychological foibles. Data show that we are overwhelmingly creatures of the present, and only through exercises of our limited pool of willpower can we force ourselves to take sufficient account of the future.[103] Relatedly, we tend to focus our attention on facts

103. For overviews of the literature, see Fennell, *supra* note 50, at 1375–94, and Shane Frederick et al., *Time Discounting and Time Preference: A Critical Review*, *in* ADVANCES IN BEHAVIORAL ECONOMICS 162, 166–79 (Colin F. Camerer et al. eds., 2004).

that are readily available to us or on items in plain sight, reacting automatically and emotionally to those immediate stimuli—mental processes Kahneman calls "system one."[104] Only with some effort do we turn our attention to the distant and the hidden, and engage our reasoning powers—Kahneman's "system two"—to reach better decisions.[105] We "anchor" on information we've already received and interpret new data selectively to fit with what we already know or want to be true.[106] In all of these areas evidence demonstrates that individuals vary considerably in their susceptibility to the behavior.[107]

These tendencies form the backbone of most surprising interventions. For instance, scholars who have studied "sticky defaults" argue that defaults are surprisingly persistent because many of us assign too much weight to the present burden of having to ponder finances and too little weight to the distant future benefits of savings.[108] Thus, policy makers who want to encourage savings by those who are "impatient" in this way can encourage employers to set the default to a high level of savings, rather than the prevailing zero or very low levels.

Other interventions employ our tendency to rely on system one processes. Choice architecture, for instance, aims to present us with options that we will find instinctively appealing.[109] New York City's new Active Design Guidelines are a literal example, encouraging builders to make stairs easy to find and elevators difficult, so as to encourage workers to climb to their offices.[110] Sunstein and Thaler mention Chicago's use of lines on the street to make drivers feel they are driving too fast and to reflexively slow down.[111]

New York's beverage limits draw on both impatience and attentiveness. Many studies show that consumers tend to eat or drink whatever is in front of them.[112] In one famous study, researchers found that consumers will eat as much soup as it takes to empty their bowl, even if the researchers are secretly pumping more soup into the bottom.[113] Items

---

104. Daniel Kahneman, *Maps of Bounded Rationality: Psychology for Behavioral Economics*, 93 AM. ECON. REV. 1449, 1451–57 (2003).

105. *Id.* at 1467–69.

106. JONATHAN BARON, THINKING AND DECIDING 203–24 (4th ed. 2008).

107. Ted O'Donoghue & Matthew Rabin, *Self-Awareness and Self-Control*, *in* TIME AND DECISION: ECONOMIC AND PSYCHOLOGICAL PERSPECTIVES ON INTERTEMPORAL CHOICE 217, 218–20 (George Loewenstein et al. eds., 2003).

108. Benartzi & Thaler, *supra* note 86, at 99–100 (pointing to evidence that education failed to change savings rates but that willpower-focused interventions did).

109. THALER & SUNSTEIN, *supra* note 8, at 3.

110. NYC CENTER FOR ACTIVE DESIGN, ACTIVE DESIGN GUIDELINES 70–81 (2010), *available at* http://www.nyc.gov/html/ddc/html/design/active_design.shtml.

111. THALER & SUNSTEIN, *supra* note 8, at 37–39.

112. See sources cited *supra* note 10.

113. WANSINK, *supra* note 10, at 47–52.

placed in front of us are more tempting, and more immediately available to our mind, than the distant benefits of restraint.[114] The burden of going back for another serving also triggers our tendency to put off future benefits, and since impatience is correlated with obesity,[115] this burden is largest for those who are most likely to overconsume. New York's nudge is to shrink the portion size, which should thereby also diminish consumption.

Importantly for my later analysis, evidence so far also suggests that some of us are more self-aware of our psychological failings than others. Consider the example of the mutual bank. Mutuals offer credit cards with relatively higher interest rates but promise "no hidden fees."[116] That combination of features seems most plausibly aimed at customers who know their own tendency to fall for the tricks played by other banks.[117] Mutuals command a small sliver of the credit market, however.[118] Similarly, many households report that they let the government keep too much in tax withholding each year so that they will not face the temptation to spend that money too soon—and then, ironically, some of these same households later pay very high fees to get access to their money a few weeks early.[119] Though other interpretations are possible, a reasonable inference is that our understanding of our own frailty, even if present, is often imperfect.

III. Choosing the Best Instrument

We're now fully prepared to compare a wide array of different policies. Again, most scholars so far agree that what they call price instruments—what in Figure 3 would fall into the stick–price–transfer boxes—dominate other options.[120] In this Part, I will argue that several of the presumed advantages of transfers and prices may be less important than has previously been understood.[121] While in some cases my claim depends

114. *See id.*

115. *E.g.*, Shinsuke Ikeda et al., *Hyperbolic Discounting, the Sign Effect, and the Body Mass Index*, 29 J. HEALTH ECON. 268, 268 (2010); Charles J. Courtemanche et al., *Impatience, Incentives, and Obesity* 17–26 (Nat'l Bureau of Econ. Research, Working Paper No. 17483, 2011), *available at* http://www.nber.org/papers/w17483.

116. Ryan Bubb & Alex Kaufman, *Consumer Biases and Mutual Ownership*, 105 J. PUB. ECON. 39, 45–50 (2013).

117. *Id.* at 48.

118. *Id.* at 45.

119. Michael S. Barr & Jane K. Dokko, *Third-Party Tax Administration: The Case of Low- and Moderate-Income Households*, 5 J. EMPIRICAL L. STUD. 963, 971–78 (2008).

120. Glaeser, *supra* note 20, at 150; *see also* sources cited *supra* note 52.

121. For simplicity's sake, when describing the effects of price instruments I mostly assume here that individuals respond rationally to the instrument. Obviously that is not necessarily so, especially in the internality context. For discussion of price instruments with irrational agents, see Brian Galle, *Carrots, Sticks, and Salience*, 66 TAX L. REV. (forthcoming 2014) *available at* http://lawdigitalcommons.bc.edu/lsfp/493/, and Garth Heutel, *Optimal Policy Instruments for Externality-Producing Durable Goods Under Time Inconsistency* (Nat'l Bureau of Econ.

on the potentially surprising nature of some transferless and priceless instruments, in other cases even conventional regulation is a better option than economists suggest.

A. Prices or Priceless?

Once more, commentators argue that price instruments best their priceless competitors by providing better information.[122] In particular, from a societal standpoint we would like to pay the least possible to get to the right level of externalities, but it's hard to know how to do that. For example, if Grungefirm can cut the first unit of emissions for $100, and Sparklefirm can clean up for only $80, it makes more sense to ask Sparkle first.[123] The problem is that government usually doesn't know the marginal cost of cleanup for each firm. Each firm has strong incentives to hide its capacity; if Sparkle can fool society into demanding reductions from Grunge instead, it saves $80. The opposite is true if we promise rewards for clean production: there, each firm wants to pretend to be the cheapest, rather than the costliest.[124]

Prices give externality producers incentives to reveal their private cost structure.[125] If we set taxes at $81 per unit of pollution, then Sparkle will spend $80 to clean up, saving $1. Grunge keeps polluting. Assuming that each is operating rationally, we can infer that Sparkle's marginal costs are less than $81 and Grunge's are more.[126] Same thing if we offer a bonus: if Sparkle accepts an $81 reward for reducing emissions, it must be the case that it costs Sparklefirm less than that to clean up its act.

This account has two potential flaws, one sketched previously by Jacob Nussim and another I'll lay out for the first time here. Nussim suggests that price instruments do not actually economize on information because they fail to reveal information about the least-cost avoider.[127] I'll

---

Research, Working Paper No. 17083, 2011), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1854185.

122. *See, e.g.*, Fullerton et al., *supra* note 45, at 430.

123. If the marginal cost of cleanup is increasing, as seems likely to be the case, then at some point it will become cheaper to switch to Grunge. At equilibrium, the marginal costs of remediation should therefore be equal at all firms. Any other result produces unnecessary social costs. Robert Stavins, *Environmental Economics*, *in* THE NEW PALGRAVE DICTIONARY OF ECONOMICS 882, 882–83 (Steven N. Durlauf & Lawrence E. Blume eds., 2d ed. 2008).

124. Competitive bidding is one classic solution to this problem, but that route has many complications, as well. JEAN-JACQUES LAFFONT & JEAN TIROLE, A THEORY OF INCENTIVES IN PROCUREMENT AND REGULATION 307–40 (1993).

125. *See* Kaplow & Shavell, *supra* note 46, at 4.

126. Of course, firms that are aware of the government's strategy can strategically conceal information. See Galle, *supra* note 27, at 822–23, 826–27, for discussion of that scenario.

127. *See* Jacob Nussim, Information Costs of Externality-Control Instruments 7–10 (Dec. 9, 2013) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2395152.

add that, even if price instruments do reveal all the information the regulator needs, in many cases priceless instruments could do just as well.

Nussim's point depends on the argument, familiar from tort theory, that sometimes it is victims who should change their behavior, not injurers. If the farmer can move her grain away from the fiery train tracks at a cost of only $50, while it would cost millions to relocate the tracks, society is better off if the farmer has to move (setting aside distributive considerations).[128] Thus the correct price for a price instrument is not, in our earlier terms, the marginal social damage but rather the lesser of the marginal social damage or the victim's cost of avoiding that damage.[129] If the price for the instrument is set at MSD, we won't learn much about victims' costs; those who can avoid the harm for less than the damage they suffer will do so, but we don't know *how much* less it costs them.

Kaplow and Shavell suggest, albeit fairly indirectly, a possible answer, though their focus is on the costs of injurers, not victims.[130] Suppose that instead of jumping directly to MSD, the government slowly phases in its new penalty over time. We then can observe the behavior of victims as the price changes. Do victims avoid injury when the damage to them is $40?[131] When it's $60? When it's $110? Nussim emphasizes that his argument is for the "static" case,[132] and perhaps this is why: in a dynamic setting, the policy maker can experiment with different values, and use the resulting observations to infer both producer and victim cost schedules.

If that is true, however, it implies that the government need not rely on price instruments. Or, more precisely, once the government has used price instruments *experimentally*, it can then switch to priceless regulation. What if the priceless instrument is superior in all respects to the price instrument, except for the fact that the priceless instrument cannot accurately account for private costs?[133] Why not gather that information, then use the more effective instrument going forward? If private costs change over time, the government could periodically introduce small-scale experiments to recalibrate.

---

128. *See generally* GUIDO CALABRESI, THE COST OF ACCIDENTS 135–97 (1970).

129. Or, putting this point another way, the true MSD is the lesser of harm or avoidance costs.

130. *See* Kaplow & Shavell, *supra* note 46, at 6 (suggesting use of varying tax schedules over time to reveal private cost information of externality producers). I am grateful to Louis for pointing out that this same argument can apply to the question of victims' costs.

131. Note that we are assuming that the government has ready access to the MSD "schedule," or at least to the expected MSD schedule. That is, the government can draw the MSD and producer's cost curves. It therefore can trace a line from the intersection of the cost of the price instrument and the producer's cost curve, and find the corresponding point on the MSD curve. That would allow it to infer the damage suffered by the average victim.

132. Nussim, *supra* note 127, at 7–8.

133. *Cf.* Calabresi & Melamed, *supra* note 44, at 1119 (arguing that, if not for information problems, property rules would usually be preferable).

No doubt it would be complex and costly to switch all of society from one regulatory instrument to another, but such large-scale disruptions would often be unnecessary. Unless every externality producer is perfectly unique, information gathered about some producers will very likely be relevant for others. Carbon mitigation costs may vary dramatically between different "generations" of power plant, but those using similar technologies will probably have similar costs. So small-scale experiments—for instance, in state "laboratories"—can provide data for later nationwide expansions.[134]

Another experimental method for replicating the informational benefit of a price instrument is to find what we could call the "shadow price" of its priceless alternative. Though many commentators seem to assume that quantity regulations are something like an absolute command,[135] more realistically any form of punishment can be priced and used in an optimal deterrence framework.[136] This shadow price can, like a traditional price, be set at the optimal level by matching it to the marginal social cost of the internality or externality.[137]

Social science should be able to estimate a person or firm's dollar-equivalent responsiveness to a priceless instrument.[138] Suppose that 50% of the population saves at the government's target savings rate when they are automatically enrolled in a retirement plan. To measure the subjective "cost" of opting out of the default, we can set up a parallel experiment in which we measure what dollar amount would produce an equal 50% participation rate. That equivalent dollar amount is the shadow price of the

---

134. But note that this likely requires carefully directed experiments rather than unguided state policy making. *See* Michael Abramowicz et al., *Randomizing Law*, 159 U. PA. L. REV. 929, 946–47 (2011) (noting that "[f]ederalism . . . does not easily facilitate a scientific approach to experimentation"); Brian Galle & Joseph Leahy, *Laboratories of Democracy? Policy Innovation in Decentralized Governments*, 58 EMORY L.J. 1333, 1368–70 (2009) (summarizing evidence that state "experiments" fail to meet the needs of real experimentation).

135. *See* Robert Cooter, *Prices and Sanctions*, 84 COLUM. L. REV. 1523, 1535 (1984) ("If officials obtain the necessary information and create an optimal legal standard, then private persons will be required to obey the legal standard upon pain of suffering the sanction attached to it."); Hepburn, *supra* note 45, at 241 ("Current speed limits are clearly a very crude approximation to 'optimal' speed limits[,] . . . where travel in excess of the speed limit is possible provided a fine is paid.").

136. *See* Becker, *supra* note 22, at 182–83 (using punishment, among other variables, in an optimal deterrence framework to determine the socially optimal levels at which "crime does not pay"); Rasmusen, *supra* note 22, at 524–27, 538 (incorporating the expected cost of "stigma" into his optimal deterrence framework).

137. Indeed, Kaplow and Shavell rely on the notion that regulatory commands can be interchangeable with prices elsewhere in their argument for price instruments. Kaplow and Shavell, *supra* note 46, at 9. They critique the traditional view that quantity regulation, and only quantity regulation, can create a "hard cap" or mandatory limit on externalities by arguing that sharp price increments can have similar deterrent effects. *Id.* at 7–10.

138. For a real-world example, see Marianne Bertrand et al., *What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment*, 125 Q.J. ECON. 263 (2010) (comparing effects of framing and price changes on likelihood of consumer borrowing and estimating a value for each framing technique).

nudge. Alternately, we could run the two experiments together by setting the regulation in place and allowing those subject to it to buy their way out, as in the famous case of the Civil War draft. Draftees were allowed to pay someone else to serve in their place, providing evidence of the subjective cost of military service for both parties.[139]

To be sure, these shadow price measures are imprecise and may vary widely across individuals.[140] For that reason, some criminal law scholars seem skeptical that alternatives to fines, such as jail or shaming, can be fit seamlessly into the optimal deterrence framework. Variations in individuals' vulnerability to harms in prison, in their adaptability to adverse circumstances, and in their subjective experiences of punishment can make it difficult to determine an average "cost" of jail time.[141]

I don't want to diminish these criticisms, but in many respects they can also be said of instruments denominated in dollars.[142] For example, individuals can also adapt to their financial situation. Just as those in prison can experience "hedonic adaptation" in which they find the experience of punishment is not as severe as they expected,[143] so too can households grow accustomed to their wealth. Researchers who study happiness argue fiercely over whether greater wealth correlates with greater happiness.[144] Hedonic adaptation to household wealth levels seems at least a plausible explanation for why it is so difficult to demonstrate this correlation: humans can find joy in whatever we have and perhaps grow blasé with familiar wealth.[145]

---

139. MICHAEL J. SANDEL, JUSTICE: WHAT'S THE RIGHT THING TO DO? 76–77, 80 (2009). As the draft example also infamously illustrates, an information difficulty for price instruments is that dollars do not have the same value for everyone. *Id.* at 82–83.

140. *See* Robert W. Hahn & John A. Hird, *The Costs and Benefits of Regulation: Review and Synthesis*, 8 YALE J. ON REG. 233, 240–43 (1990) (discussing limits on tools for translating "hedonic" preferences into dollar units); Hill, *supra* note 18, at 453 (giving examples of goods or taxes that provide different values to different individuals).

141. *See* John Bronsteen et al., *Happiness and Punishment*, 76 U. CHI. L. REV. 1037, 1046–55 (2009); Adam J. Kolber, *The Subjective Experience of Punishment*, 109 COLUM. L. REV. 182, 187–96 (2009). For a different take on the relevance of these data, see Dan Markel & Chad Flanders, *Bentham on Stilts: The Bare Relevance of Subjectivity to Retributive Justice*, 98 CALIF. L. REV. 907, 959–64, 967–73 (2010).

142. *See* Vandenbergh et al., *supra* note 29, at 735–36 (arguing that the impact of psychological or social factors may be more predictable than the effect of prices).

143. Bronsteen et al., *supra* note 141, at 1048–49.

144. *See* Richard A. Easterlin et al., *The Happiness–Income Paradox Revisited*, 107 PROC. NAT'L ACAD. SCI. 22,463, 22,463–67 (2010) (positing that over the long term, an increase in happiness does not follow an increase in wealth and critiquing alternative findings); Mike Morrison et al., *Subjective Well-Being and National Satisfaction: Findings from a Worldwide Survey*, 22 PSYCHOL. SCI. 166, 169–70 (2011) (finding that national satisfaction has a greater effect on overall life satisfaction for lower income individuals than it does on higher income individuals, partially compensating for the satisfaction discrepancies created by wealth disparities).

145. Thomas D. Griffith, *Progressive Taxation and Happiness*, 45 B.C. L. REV. 1363, 1389–90 (2004).

In addition, dollars could be less predictable than other forms of regulation to the extent that they are more or less salient. Of course, an incentive is usually only effective when people are aware of it. A growing body of real-world evidence suggests that consumers and other actors are not always fully attentive to dollar prices.[146] Math, and our understandable desire to avoid the pain of having to think about it, may help explain why people neglect prices.[147] Presumably that would be less of an obstacle for regulations that confront individuals with experiences or sensations rather than numbers. On the flip side, some researchers also find that dollar-denominated incentives are at times so visceral that they crowd out other, less tangible motivations,[148] which could make dollar-denominated sticks *more* salient than policy makers intend.

On the other hand, price instruments may retain their advantage in situations where experiments of the kind I have suggested are impractical. Suppose, for example, that landowners have a highly varying and idiosyncratic attachment to their property. If protecting that attachment is an important policy goal, price instruments might be more likely to achieve it.[149]

B. Transfer or No Transfer? Public or Private?

Another putative key advantage of taxes over regulation, or of fines over prison, is that in each case the transfer of money produces a better outcome than pure punishment or imposition.[150] As I will argue in this subpart, these claims may well be less true of nudges and other surprising forms of regulation. They also are less accurate when transfers flow not to

---

146. *See* Raj Chetty et al., *Salience and Taxation: Theory and Evidence*, 99 AM. ECON. REV. 1145, 1153–56 (2009); Aradhna Krishna et al., *A Meta-Analysis of the Impact of Price Presentation on Perceived Savings*, 78 J. RETAILING 101, 101–18 (2002).

147. See Brian Galle, *Hidden Taxes*, 87 WASH. U. L. REV. 59, 81–85 (2009) for a discussion of the uncertain basis for the low salience of some taxes.

148. John Condry & James Chambers, *Intrinsic Motivation and the Process of Learning*, *in* THE HIDDEN COSTS OF REWARDS: NEW PERSPECTIVES ON THE PSYCHOLOGY OF HUMAN MOTIVATION 61, 61–66 (Mark R. Lepper & David Greene eds., 1978); *see also* Stephanie J. Byram, *Cognitive and Motivational Factors Influencing Time Prediction*, 3 J. EXPERIMENTAL PSYCHOL.: APPLIED 216, 233 (1997) (reporting that financial incentives for speed of performance exacerbate biased performance).

149. Kaplow and Shavell suggest that when idiosyncratic valuation is high, property rules are superior to liability rules. Kaplow & Shavell, *supra* note 44, at 759–63. But that argument deals with the special situation of takings, *id.*, which actually more closely resemble quantity regulation than price instruments: the initial owner has no choice about whether to sell the property, and so although there is a "price" that the government pays, that price reveals nothing about the owner's preferences. In my framework, compensated takings are priceless and (to the extent that compensation is imperfect) partial transfers.

150. *See* Helfand et al., *supra* note 23, at 287 (arguing that price-based instruments like effluent taxes are more efficient than quantity-based standards); *see also* Korobkin, *supra* note 84, at 1668–69 (asserting that a system allowing for trading ensures a better maximization of utility than an imposition of a one-size-fits-all standard).

the public coffers but instead to other private parties. And prior commentators seem not to have considered the possibility that transferless instruments may be superior to transfers in the case of carrots—that is, if the transfer would be *to* the externality producer, rather than away.

1. Of Transfers and Nudges.—To understand fully the claim that transfers are superior to transferless instruments, we must take a short detour into the economics of taxation. Recall that virtually all taxes produce deadweight loss, or economic waste resulting from changes in actors' behavior in response to the tax. But computing the net loss of a Pigouvian tax or other transfer is a bit complicated, although in the case of externalities it has now been thoroughly examined by economists.[151]

The welfare effects of the stick transfer are a combination of several factors. Like any tax on a specific commodity, the transfer changes people's decisions about what goods to put in their market basket, producing deadweight loss.[152] It can also reduce their "real" returns to labor. That is, when laborers decide whether to get out of bed and go to work, they implicitly are deciding whether the utility payoff of their salary is worth the opportunity cost of more pillow time. Since taxes on goods reduce the utility payoff from salary, economists typically predict that a consumption tax will also affect this labor/leisure decision; this effect is sometimes called the "tax-interaction" effect because it is compounded in the presence of existing taxes on labor itself.[153]

In the case of Pigouvian taxes or other transfer instruments, the funds can be "recycled" by using them to cut other, distortionary taxes.[154] Depending on how well that recycling is targeted, the gains from offsetting other taxes may or may not exceed the deadweight losses the transfer produces.[155] And, of course, when consumers switch away from the taxed

---

151. For an accessible overview, see Parry & Oates, *supra* note 23, at 604–10, and for a more technical summary, see Bovenberg & Goulder, *supra* note 74, at 1486–507. Few previous commentators have examined the efficiency of a Pigouvian tax for internalities, or harms we do to ourselves. The major exception is Ted O'Donoghue & Matthew Rabin, *Optimal Sin Taxes*, 90 J. PUB. ECON. 1825, 1834–35 (2006), who omit consideration of the effects of the tax on labor supply. As I show here, that can be a major factor in the efficiency of the tax.

152. For instance, suppose Massachusetts taxes orange things but not pink ones. I prefer oranges to grapefruit. When I shop in Massachusetts, though, I might buy grapefruit to avoid the orange tax. My utility is lower, and Boston has no more money in its treasury. That is deadweight loss from a differential tax on consumption goods.

153. Parry & Oates, *supra* note 23, at 605–06. It is also likely that there are tax-interaction effects for taxes on capital, but these have not been thoroughly explored in the existing literature.

154. *Id.* at 606–07.

155. *See* Bovenberg & Goulder, *supra* note 74 (concluding that "small" environmental taxes offset the labor tax increase, whereas "large" environmental taxes do not offset the increased labor tax burden and lead both to a reduction in real wages and a drop in employment). Of course, revenues could also be used for new government programs rather than to reduce taxes. One way to think about why we focus on the cost of raising revenues instead of the benefit of these

good, they reduce harmful externalities for others. It might be helpful to think of the various effects as the terms of a simple equation:

$$U_t = E - L - C + R.$$

That is, the utility effect of the transfer includes E, externality gains; losses from the tax-interaction effect on labor, L; losses from changes in consumption choices, C; and gains from revenue recycling, R.[156]

Recall, too, that environmental economists argue that even regulations that do not explicitly put a price on behavior can also cause deadweight losses, especially when those regulations are enacted in the context of an existing income tax.[157] But unlike the transfer, the regulation does not bring in any new funds. In effect, the regulation gives us a utility result:

$$U_r = E - L - C.$$

Prior commentators therefore argue that a tax is unambiguously better by the amount of the quantity, R.[158] That is the logic that seems to be motivating critics who complain that New York's soda-cup default is a worse policy than a soda tax.

A critical assumption in this line of argument is that the labor-supply effects of the transfer and transferless policies are identical. In many instances that assumption is perfectly defensible. If the government has accurately determined the optimal quantity of the externality, presumably the cost of achieving that optimal level of production is identical under either instrument. Rational actors will account for that cost when they decide how hard to work.

An example here could be helpful. Let's take a soda drinker, Albert. The City of Novum Eboracum wants to reduce the burden of Albert's future health costs on its budget and determines that the optimal soda tax for Albert is $.50 per centiliter, which results in five liters of soda consumption on average. At $250 per week, that tax will likely significantly reduce

---

programs is that the benefits would be the same no matter how we paid for them. So the only variable is how socially costly it is to acquire the money to fund them.

156. This equation follows from, but simplifies, the calculations in Bovenberg & Goulder, *supra* note 74, at 1486–503. In my view the assumption that there are distortions in the commodities market, modeled here as the quantity C, should be controversial. The claim seems to be that consumers have clearly formed preferences prior to imposition of the tax and that the tax distorts these. But arguably the tax itself shapes or helps consumers to revise preferences, as was reportedly the case of the Washington, D.C.-area tax on shopping bags. *Cf.* Barr et al., *supra* note 68, at 28–29 (arguing that preferences are constructed during decision processes). If so, it isn't clear that this effect should count as a distortion. Rather than take a definitive position on the question, I will simply assume for now that these changes should count as welfare losses.

157. Bovenberg & Goulder, *supra* note 74, at 1502.

158. *See* sources cited *supra* note 52.

Albert's real returns to labor, and therefore figure prominently in his decision how hard to work. Alternately, the City can make obtaining soda annoying or unpleasant, such as by selling it only in very small containers. By definition, the shadow price of the annoyance, when it is set optimally, is $250—the exact amount of annoyance that Albert experiences when he pays the optimal soda tax. If Albert rationally anticipates the irritation of going back to the store for many tiny bottles of soda, that, too, should figure into his labor–leisure decision.

Except that we know from psychological research that Albert actually is unlikely to anticipate his future subjective costs of complying with the City's nudge. System one processes are automatic, not deliberative.[159] As such, their operation can be difficult to notice in real time, and even more difficult to predict.[160] Nudges that rely on our tendency to draw on automatic behavior, such as placing healthy food close to the cash register or using visceral, emotionally charged images rather than detailed information, affect us without triggering conscious thought.[161] Unsurprisingly, many studies find that consumers are extremely poor at predicting their own susceptibility to private firms' use of these kinds of techniques.[162]

Thus, Albert may not even anticipate that the nudge will change his behavior. Recall that portion size affects consumption exactly because we don't think about how much we're consuming, instead letting ourselves be guided by whatever we're presented with.[163] Reducing the size of a soda

---

159. Kahneman, *supra* note 104, at 1450–51.

160. *See* Roland Bénabou & Jean Tirole, *Self-Knowledge and Self-Regulation: An Economic Approach*, *in* 1 THE PSYCHOLOGY OF ECONOMIC DECISIONS 137, 137–38 (Isabelle Brocas & Juan D. Carrillo eds., 2003) (summarizing the authors' findings that individuals may not know "what actions they would take in a given situation until the very moment when they actually experience it"); George Loewenstein & David Schkade, *Wouldn't It Be Nice? Predicting Future Feelings*, *in* WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY 85, 92–98 (Daniel Kahneman et al. eds., 1999) (discussing how people consistently overestimate their own willpower and often inadequately predict how they will act in an excited state); *see also* Kahneman, *supra* note 104, at 1451 ("The operations of System 1 are fast, automatic, effortless, associative, and often emotionally charged; they are also governed by habit, and are therefore difficult to control or modify.").

161. *See* THALER & SUNSTEIN, *supra* note 8, at 20–22.

162. *See, e.g.*, Stefano DellaVigna & Ulrike Malmendier, *Contract Design and Self-Control: Theory and Evidence*, 119 Q.J. ECON. 353, 364 (2004) (noting that consumers often choose contracts with initial discounts, such as low credit interest, more often than they anticipate, even though these contracts eventually lapse into greatly augmented interest rates and profit for the issuing firm); Sha Yang et al., *Unrealistic Optimism in Consumer Credit Card Adoption*, 28 J. ECON. PSYCHOL. 170, 181 (2007) (concluding that consumers with unrealistic and self-serving optimism regarding future borrowing tend to prefer credit cards with features that are not in their best interest and are thus targeted customers for lending institutions).

163. *See* sources cited *supra* note 10. However, individuals may respond to nudges *after* they experience them. Some evidence suggests that nudged food consumers compensate for healthier choices by eating unhealthier foods afterwards. *See* Matteo M. Galizzi, *Label, Nudge or Tax? A Review of Health Policies for Risky Behaviours*, 1 J. PUB. HEALTH RES. 14, 16 (2012).

portion could diminish consumption without drinkers even really noticing, let alone changing their work habits in response.

Researchers are less clear on our ability to predict our own willpower. Though our exact predictions are poor, those who struggle with temptation usually know it and take steps accordingly.[164] Nonetheless, if surprising regulations often work because we are unwilling to exert the mental effort to overcome them, it seems unlikely we would exert the mental effort to compute and adjust our labor supply in response.[165] Others may refuse to face the fact of their own failings in order to preserve their self-esteem or to reduce feelings of internal conflict.[166]

On the other hand, the prediction that nudges will have little effect on labor supply depends to some extent on the assumption that labor supply is a system two process—that we think and plan about how hard to work. What if labor supply instead is itself a fairly instinctual process, where we force ourselves out of bed each morning based on some gut sense of how rewarding work will be? Perhaps Albert has some vague sense of dissatisfaction with the returns of his salary, without being able to identify that it is related to his newly healthy diet. No empirical work has yet examined the labor-supply effects of nudges, and so this seems to be a critical area for future research.

In any event, it is at least theoretically plausible that surprising policy instruments could have surprisingly low effects on labor supply, relative to more traditional transfers. In the most dramatic cases, labor effects could be negligible. Then it would be ambiguous whether transfers are superior; in terms of the earlier equations, transfers are superior only where $R > L$:

$$[U_t = E - L - C + R] > < [U_n = E - C].$$

That is, since these forms of transferless instruments eliminate labor distortions but bring in no revenues, they are the better choice when the transfer's labor distortions reduce welfare by more than its revenues increase welfare.[167] As it turns out, the relationship between R and L depends on whether transfers are public or private. Let's turn there next.

---

164. *See* Nava Ashraf et al., *Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines*, 121 Q.J. ECON. 635, 636–37 (2006); Ted O'Donoghue & Matthew Rabin, *Doing It Now or Later*, 89 AM. ECON. REV. 103, 105 (1999).

165. *Cf.* Patrick Bolton & Antoine Faure-Grimaud, *Thinking Ahead: The Decision Problem*, 76 REV. ECON. STUD. 1205, 1205–08 (2009) (modeling costs of deliberating about future decision settings).

166. *See* Roy F. Baumeister, *The Self*, *in* 1 THE HANDBOOK OF SOCIAL PSYCHOLOGY 680, 688–92 (Daniel T. Gilbert et al. eds., 4th ed. 1998).

167. A key assumption here is that the efficacy of the nudge-type intervention, represented here by the E term, can approximate the efficacy of other tools. For skepticism on that point, see Ryan Bubb & Richard E. Pildes, *How Behavioral Economics Trims Its Sails and Why*, 127 HARV. L. REV. (forthcoming 2014) (manuscript at 70–73), *available at* http://papers.ssrn.com/sol3/papers

2. Public or Private Transfers?—We saw earlier that not all sticks transfer money to the government, nor does the public fisc pay for all carrots.[168] As prior commentators note, this shift can have important welfare implications.[169] Others have not recognized, though, that these welfare consequences generally weaken the case for transfer instruments. Transfers to private parties are typically not as efficient as transfers to the government, and so this diminishes the advantage of the transfer relative to a transferless policy. Likewise, transfers paid for by private parties are less efficient than if the government pays.

Here again the results follow from some basic principles of tax economics. Recall that the deadweight loss of taxes rises exponentially with the size of the tax; doubling the tax is four times as bad, roughly speaking. All else equal, when there are more taxpayers contributing to the cost of a given program, the tax rate for each is of course lower. Putting these two facts together, it is well known that society is typically better off if costs can be spread as widely as possible.[170] Privately funded carrots are often less efficient, then, because the group who pays for the carrot will almost certainly be smaller than the group of all taxpayers, resulting in greater total deadweight loss in most cases.[171]

A bit less obviously, dedicating transfer funds to private beneficiaries can be inefficient for similar reasons.[172] Think of the transfer to the private group as a special-interest income tax cut for that group.[173] The opportunity cost of the transfer is that we could have used the same money to fund an income tax cut for everyone. Which would have generated greater welfare? The answer often has to be the tax cut for everyone. Why? Because the deadweight-loss savings of the first dollar of income tax savings is the biggest—that is, going from $100 to $99 is a bigger savings than going from $99 to $98, and so on. Multiplying that large gain times the whole of the taxpaying public is a much greater gain than getting only the gains

---

.cfm?abstract_id=2331000, and Lauren E. Willis, *When Nudges Fail: Slippery Defaults*, 80 U. CHI. L. REV. 1155, 1226–29 (2013).

168. I am still assuming for the sake of simplicity that all externality producers are private parties so that sticks are paid and carrots collected privately.

169. *E.g.*, Gallini & Scotchmer, *supra* note 61.

170. *See* sources cited *supra* note 63.

171. Gallini & Scotchmer, *supra* note 61. Some of the qualifying language in this sentence derives from the fact that I am omitting any consideration of the marginal utility of money. But note that, in general, diminishing marginal utility will also weigh against private transfers; even if everyone starts with the same amount of money, the marginal cost of the last dollar paid by the private funders will be greater than the cost of the last dollars paid by the general public, since the general public will have more money left over after paying.

172. *See* Thomas D. Griffith, *Should "Tax Norms" be Abandoned? Rethinking Tax Policy Analysis and the Taxation of Personal Injury Recoveries*, 1993 WIS. L. REV. 1115, 1123–27.

173. Lump-sum transfers would be even less efficient. By granting the transfer in the form of a tax cut, we can reduce the deadweight loss of taxation in addition to enriching the transferees, while a lump-sum transfer only accomplishes the latter.

from, say, $100 to $50 for a few people. I add a numerical example in the margin below for interested readers.[174]

In the analysis so far I haven't mentioned any distributive considerations, such as the marginal utility of money for the payors and payees, and for good reason. The trade-off in allocating transfers to or from private parties, we have just seen, is no different than the trade-off involved in designing the income tax itself: should some people get a tax break, at the expense of everyone else?[175] If that were a good trade-off, presumably income tax designers would already have made it, or could do so soon, unless for some reason it is administratively easier to effect redistribution through the policy instrument, rather than the tax system.[176]

We're now in a good position to revisit the relative size of R and L from the previous section. Remember that R is the social welfare gained from cutting income taxes by the amount of the revenues earned by a transfer instrument, while L is the amount of welfare lost to distortions in labor supply from that instrument. A transferless instrument that creates neither, or only a negligible L, is a better choice when $L > R$ for the transfer instrument alternative. When would that happen? Usually L and R are both derived from changes in labor supply, so that R is just the bonus we get by reducing our labor-distorting income tax. Since both L and R involve transfer of the same amount of money—L is the loss from a tax hike of $\$X$, while R is the gain from a tax cut of $\$X$—if both are distributed evenly across the whole population there is no net change.

More likely, the population subject to L is private—it's the small group of externality producers. Externality producers are going from paying, say, 100 to 110, while the public is getting a cut from 100 to 99. Because, again, distortions rise exponentially with the tax rate, the welfare losses, L, involved in the producers' jump are larger per dollar of revenue than the benefits, R, the public gets. As environmental economists have recognized, it is therefore not very likely that we can come out ahead by

---

174. Suppose Priya Vat faces a tax of $\$100x$. Assume for simplicity the deadweight loss of the tax is the square of the tax, so Priya's tax comes with a total deadweight loss of $\$10{,}000x$. Should we reduce Priya's tax to zero or give 100 other people, each also currently paying $\$100x$, a $\$1x$ cut? Priya's break saves us $\$10{,}000x$. Each $\$1x$ cut we give to someone else saves us ($\$10{,}000x - \$9{,}801x = \$199x$). Obviously, $100 \times \$199x = \$19{,}900x$, much more than we would save from Priya's cut. Note, though, that if it turns out that Priya is paying a much higher tax rate than other members of the public, such that the deadweight loss of taxing her is considerably greater, we could have a closer question.

175. *Cf.* KAPLOW, *supra* note 58, at 152–56 (explaining that all transfers can be considered in an "aggregate" framework as aspects of the income tax system).

176. *Cf.* Gerrit De Geest, *Removing Rents: Why the Legal System Is Superior to the Income Tax at Reducing Income Inequality* 8, 35 (George Washington Univ. Legal Studies Research Paper Series, Paper No. 13-10-02, 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm ?abstract_id=2337720 (arguing that tax systems may have higher administrative and information costs than regulation).

taxing the small group to pay for cuts for everyone else.[177] So usually $L > R$. This answers the question we left open at the end of section III(B)(1): transferless instruments with little labor distortion *do* frequently offer an advantage over transfers.

I do not mean to suggest that the reshuffling of revenues is the only important aspect of the public/private question. Again, the distribution of benefits and burdens can affect incentives, not to mention the political economy of the instrument. For instance, one might defend the tax exemption for tort judgments on the grounds that the extra social costs of making the revenues wholly private are necessary to buy added enforcement and monitoring effort from private parties.[178] I will largely reserve those questions for elsewhere, however.

3. Transfer Carrots.—Finally, commentators have not previously acknowledged that transferless instruments have a deadweight loss advantage over carrots. If carrots are the only transfer instrument available, transferless alternatives may be preferable even if they do not provide as precise information. Surprisingly, carrots often can be a better bargain than sticks or transferless instruments on a per-unit basis, but their poor targeting ultimately makes them pricier.

Carrots look like a bargain because they don't reduce the labor supply of externality producers.[179] In the externality case, we can give the utility of offering a carrot as:

$$U_c = E - C - R_c$$

177. Bovenberg & Goulder, *supra* note 74, at 1498–503. That is less true if revenue gains can be made in a tax more distortive than the income tax, such as the corporate tax. *Id.* at 1505–07. Similarly here, if the government can pay for its carrots using a tax *less* distortive than the income tax, the carrot transfer is more appealing.

178. *Cf.* Dan Markel, *Overcoming Tradeoffs in the Taxation of Punitive Damages*, 88 WASH. U. L. REV. 609, 636–38 (2011) (suggesting that a plaintiff's claim on punitive damages can be thought of as a "finder's fee").

179. The subsidy creates no labor effect among marginal agents because by assumption the amount of the subsidy is just enough to leave them indifferent. *See* De Geest & Dari-Mattiacci, *supra* note 72, at 363 (distinguishing carrots from sticks by noting that carrots "fully compensate" producers). If the subsidy is less than this amount, it is never collected. If it is more than this amount, then it would increase real returns to labor, but it also would be pure waste from the perspective of the government—in our equation, a more positive L term would be offset by a diminished E and greater C terms. Of course, it is still possible that the pure exchange of higher taxes on some in return for increased labor for others could be welfare-enhancing; that is arguably the case for the Earned Income Tax Credit (EITC). *Cf.* Gregory Acs & Eric Toder, *Should We Subsidize Work? Welfare Reform, the Earned Income Tax Credit and Optimal Transfers*, 14 INT'L TAX & PUB. FIN. 327, 332 (2006) (observing that EITC offsets the negative work incentives of the payroll tax). But that would take us away from the Pigouvian tax setting that is my focus here.

where E and C are the same as before and R_c is the deadweight loss associated with increasing income or other taxes to pay for the carrot.[180] By comparison, we still have:

$$U_s = E - L - C + R \text{ and } U_r = E - L - C.$$

These equations imply that for any given unit of externality reduction, a carrot is welfare superior to transferless instruments when $L > R_c$ and superior to sticks when $L > R + R_c$. That is, the carrot is a better deal when the cost of paying for the carrot is less than the welfare loss that would be caused by the other instruments' effects on labor supply.

In all likelihood, however, the per-unit cost of the carrot must be paid many more times than the per-unit cost of a stick or transferless instrument. As Dari-Mattiacci and De Geest argue, all carrot beneficiaries will typically claim their carrot, regardless of whether the carrot changes the claimant's behavior.[181] The cost of the other instruments is only incurred, though, for those who would not otherwise have complied. For example, if the government paid people not to steal, it would have to pay almost the entire population, while if it nudged them away from theft, only the lightest fingered of the population would feel much burden. I will add that this large potential difference in total "price" also affects the size of the R_c term: when there are more carrot claimants, the tax rate needed to pay for them also rises, with exponential effects on the resulting welfare loss.

C. *Carrot, Stick, or Compromise?*

By definition, a transfer instrument must be either a carrot or a stick: value is either transferred to the externality producer or away from it. In my prior work on carrots and sticks, I argued that in many cases choosing either instrument involves trade-offs.[182] For instance, using carrots to encourage positive externalities may offer desirable income effects but threaten serious incentive problems. Transferless instruments offer a third way. Since they don't move money around, they lack both the benefits and detriments of carrots and sticks. Sometimes, that middle ground might be the best of all. Let's look at three different kinds of trade-off: income effects, redistribution, and incentives.

1. Income Effects.—It seems obvious that where sticks reduce the wealth of payors and carrots increase it, transferless instruments sometimes do neither. Though it is an obvious point, it is also potentially a very

180. Note that I assume that the subsidy creates a loss, C, from distortion in the product market because, like the penalty, it changes consumers' preferences.

181. Dari-Mattiacci & De Geest, *supra* note 27, at 369–76.

182. Galle, *supra* note 27, at 809–13.

significant one, and one that no other commentators seem to have focused on. Income effects often present some of the strongest arguments for choosing between carrots and sticks.[183] As we will see, the availability of a third option with intermediate income effects will often open new and potentially more efficient policy possibilities.

Of course, some transferless regulations do change wealth. Zoning changes or other land-use regulations could make my existing property more valuable or less. As Calabresi and Melamed noted, in the presence of transaction costs the simple act of assigning a legal entitlement, such as the right to exclude, to one party might subjectively enrich that party.[184]

Many other regulations affect welfare, but not *wealth*. That is an important distinction, since the income effect operates primarily from changes in a person's budget; as our capacity to engage in trade expands, we may want different things.[185] When I lose a $20 bill, I can buy less, but stubbing my toe and suffering $20 worth of throbbing pain has no impact on the groceries I can take home. Shaming, imprisonment, and the mental hassle of opting out of a sticky default all could fall into this category.

Internalities present a more complex picture. Government policy that helps individuals overcome their own mistakes may help individuals to better allocate their spending. The consumer now can buy more of her highest-priority goods instead of wasting money on tempting alternatives. In effect, her budget has expanded. Or, alternately, we can think of the internality correction as having provided the consumer with a free service, such as credit counseling or a "commitment device," that is, a reliable way of helping people commit to not spending foolishly.[186] Evidence suggests that many households are willing to pay considerable amounts for commitment devices.[187]

In the case of normal goods, this income effect can somewhat offset the substitution effect on the consumer's consumption of the internality good. For example, once Lindsay is no longer spending as much money each month on her morning vodka, she can more easily pay rent.[188] With her housing stable, it is rational for her to consume more of the less important items in her budget, including the occasional glass of wine with dinner.

---

183. *Id.* at 832–38.

184. *See* Calabresi & Melamed, *supra* note 44, at 1098–99.

185. *See* GRUBER, *supra* note 32, at 35–37.

186. *Cf.* Markus Haavio & Kaisa Kotakorpi, *The Political Economy of Sin Taxes*, 55 EUR. ECON. REV. 575, 580 (2011) ("[S]ophisticated consumers might value sin taxes as a way of committing to a lower level of consumption in the future.").

187. Ashraf et al., *supra* note 164; David Laibson, *Life-Cycle Consumption and Hyperbolic Discount Functions*, 42 EUR. ECON. REV. 861, 868 (1998).

188. *Cf.* Chetty et al., *supra* note 146, at 1173–74 (discussing the income effect of improved allocation of consumer choices).

Although the nudge does, therefore, have some potential income effects, that effect is still an intermediate position between sticks and carrots. An internality-correcting carrot would have an even larger income effect: it would both expand the household's budget and also improve its allocation. And internality-correcting sticks would have both positive and negative income effects (better allocation, less money), such that it is unclear which dominates in any particular instance. But since the positive income effect of correcting the internality (improved allocation) seems identical no matter the instrument, the stick's propensity to increase demand would be unambiguously less than that of the nudge.

For pure internalities, though, the income effect of a government correction may not matter much. By assumption, society's only interest is in helping the household get to its unbiased preferred consumption of each good. The household's demand for the internality does not drop as far as it would in the absence of income effects. But the new level of consumption is still the efficient level for the household, given its new wealth and preferences.

The income effects of correcting an internality are most clearly problematic in the case of goods with both internalities and externalities.[189] Imagine that the Shvitz household has an old, inefficient air conditioning unit. They receive a government subsidy to buy a new one. Though they will spend less on energy consumption keeping cool, they also will be able to afford to run their air conditioner more often. If they had instead been threatened with a fine and self-financed the purchase of a new air conditioner, they would have had less money to run the new unit. Also note in the energy case that households with higher wealth can consume other goods that produce externalities. Even though the Shvitzes are subjectively better off with their new unit, they also now have more money to drive around or heat their house in winter.

2. Distributive Effects.—Next, carrots and sticks differ considerably in the way they redistribute wealth, and that difference is important for many commentators.[190] Carrots move money from taxpayers or private payors to externality producers, while sticks do the opposite.[191] Transferless instruments, in contrast, can be distributively neutral.[192] That seemingly banal distinction has some potentially important policy consequences.

---

189. *Cf.* Allcott et al., *supra* note 12, at 11–24 (modeling effects of subsidies for energy-efficient durable goods on marginal energy consumption).

190. *See* Calabresi & Melamed, *supra* note 44, at 1121 (using the example of a factory polluting a wealthy section of town to show how different solutions, among them carrots and sticks, produce markedly different trade-offs between economic efficiency and distributional goals).

191. *Id.*

192. *Id.*

For one, I have argued before that the distributive consequences of sticks may be a reason to prefer carrots when programs affect poorer households.[193] Transferring funds away from taxpayers who are already indigent runs contrary to basic distributive justice principles. Indeed, the logic of redistribution seems to have driven the design of both the Affordable Care Act (ACA) and the cap-and-trade climate change bill passed by the U.S. House of Representatives in 2009. Both legislative schemes relied primarily on a traditional stick to control externalities. The ACA, famously, imposes a penalty tax on households that fail to purchase insurance, while the cap-and-trade bill required businesses to purchase licenses to emit greenhouse gases.[194] Each, though, made exceptions for low-income families. The ACA exempts households that cannot find "affordable" insurance from its mandate, providing them with subsidies instead.[195] The climate-change legislation offered lump-sum refunds to each household, which in effect converted it into a transferless price.[196]

Carrots and transferless instruments have another practical advantage over sticks. It is well known that imposing liability on households that might be unable to afford to pay the full stick price would blunt the incentive effects of the price instrument.[197] If insurance is unavailable, the implication is that a different regulatory option—imprisonment, for example—may be necessary to ensure that poorer households face the correct marginal incentives.[198] Though other commentators have not mentioned it, this same argument can be a reason to substitute carrots for sticks.

If both carrots and transferless prices can account for low-income or judgment-proof producers, which one is the better choice? The ACA or the climate-change bill? There are good arguments for each, depending on context. As we have already seen, when income taxes are a perfectly viable redistributive tool, transferless instruments may be superior because they reduce the social cost of paying for carrots.

On the other hand, carrots may be a strong choice when policy makers cannot easily use the income tax for redistribution. Obviously, a carrot transfer will move more money to poor households than a transferless

193. Galle, *supra* note 27, at 817–20.

194. 26 U.S.C. § 5000A(b)(3) (Supp. V 2012); American Clean Energy and Security Act of 2009, H.R. 2454, 111th Cong. § 311 (2009).

195. 26 U.S.C. § 5000A(e)(1).

196. H.R. 2454 § 431.

197. *See* Kaplow & Shavell, *supra* note 44, at 739–40; *see also* Helfand et al., *supra* note 23, at 297 (noting that judgment-proof firms are also difficult to adequately deter); Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193, 1208 (1985) (suggesting that nonmonetary penalties for crimes can be justified where a defendant has resources to pay a fine but those resources are illiquid).

198. *See* Kaplow & Shavell, *supra* note 44, at 740. See Polinsky & Shavell, *supra* note 22, at 411–12, 420–22, for development of this idea in the criminal-enforcement context.

instrument would. As we saw in the last section, redistributive carrots are often indistinguishable from using the income tax for the same purpose: we are raising some people's taxes to cut others'. Kaplow argues it therefore would be foolish to adopt an externality-correcting policy in order to effect some redistribution.[199] Since the policy offers only what is already available through existing institutions, we should not pay to set up a new regulatory structure unless it is worthwhile on nonredistributive grounds.

But this assumes a fully operational income tax. Even putting aside political economy concerns, redistributive income taxation can be impractical for some governments, or at least less efficient than a privately funded carrot alternative. Municipal governments offer a classic example. If a city imposes an income tax, it drives taxpayers into surrounding suburbs, where they can still enjoy most of its amenities.[200] User fees on the amenities make that kind of free riding more difficult.[201] These fees can then be used to subsidize amenity consumption by poorer users. For example, revenues from tolls or congestion fees can be set aside to pay for public transportation or other transit assistance for the indigent. For central cities, transfer policies may be the only practical way of achieving their preferred level of redistribution.[202]

3. Games and Mitigation.—A third set of major differences between transferless and rival instruments is their respective effect on incentives for future behavior. Nudges and other surprising instruments also have some additional features that distinguish them from ordinary transferless policies.

First, transferless instruments split the difference between carrots and stick transfers when it comes to the incentives of victims. Recall that efficient laws generally give victims the incentive to mitigate their own exposure to harm when they are the least cost avoider.[203] Victims who can collect damages for their full harm, despite failure to take precautions, may lack incentive to avoid injury. Therefore most commentators suggest that fines or punitive damages should be paid to the state rather than victims.[204] In my terminology, these proposals convert the private transfer into a public one.

---

199. KAPLOW, *supra* note 58, at 32.

200. Richard Briffault, *Our Localism: Part II—Localism and Legal Theory*, 90 COLUM. L. REV. 346, 351–53 (1990).

201. WILLIAM G. COLMAN, A QUIET REVOLUTION IN LOCAL GOVERNMENT FINANCE 9–10 (1983).

202. Federalism theorists also suggest other alternatives, such as revenue sharing, but these have their own problems. Richard M. Bird & Michael Smart, *Intergovernmental Fiscal Transfers: International Lessons for Developing Countries*, 30 WORLD DEV. 899, 900–09 (2002).

203. RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 172 (7th ed. 2007).

204. *See* sources cited *supra* note 65.

I have argued before that even fully public transfers can fail to achieve the goals commentators set for them.[205] If victims are also taxpayers, they share partially in the benefits of the public transfer. Often that share is tiny enough to be irrelevant, but that is certainly less true of, say, a large firm in a small town. Reciprocally, carrots might be preferable to sticks, even if the carrots' costs are fully public, because potential victims or beneficiaries must bear a fraction of the treasury cost. The public therefore retains an incentive, if only partial, to reduce the cost of bribing externality producers. And of course this incentive would be stronger if the costs were borne by a small, private group.

By escaping revenue effects altogether, transferless prices fall in between these two cases. They avoid giving any reward, even a fractional one, to victims. If victim mitigation is an important concern, transferless instruments, such as prison, might be more efficient than transfer sticks, such as fines, contradicting what has been thought until now a fundamental tenet of the economics of crime.[206] Admittedly, though, transferless instruments also may fail to offer even the fractional incentive a carrot could provide.[207]

Secondly, surprising instruments add another wrinkle to the incentives game. Carrots and sticks differ crucially in their effects on forward-looking actors. According to Coase and many others, carrots' fatal flaw lies in their tendency to encourage new harms by producers who want to be paid to stop.[208] Similarly, in the case of positive externalities, carrots can crowd out good behavior or encourage strategic delays, as the producer dawdles until the government agrees to pay.[209] In contrast, a producer who anticipates that her activity will be punished has good reasons to take steps to mitigate her harm in advance.[210]

Transferless instruments can duplicate these *ex ante* effects, but only to the extent that externality producers recognize in advance that they will perceive the instrument as costly or rewarding. Will knowing that I might have to get up to drink a bigger serving of soda next year make me want to cut back on Coke today? Probably not to the extent a future soda tax

---

205. Galle, *supra* note 27, at 824–27.

206. Crime theorists have made this argument at least since Bentham. Becker, *supra* note 22, at 193 n.40.

207. Some transferless instruments, such as imprisonment, can also be costly to construct and administer. But the carrot is typically more efficient because if set optimally its total cost is equal to the total harm, while the cost of the prison system is essentially random, and therefore could greatly over- or under-incentivize mitigation.

208. Wiener, *supra* note 27, at 726 & n.186.

209. *See* Cass R. Sunstein, *Television and the Public Interest*, 88 CALIF. L. REV. 499, 546–47 (2000) (examining the incentive effects of subsidies for public-interest broadcasting).

210. Saul Levmore, *Changes, Anticipations, and Reparations*, 99 COLUM. L. REV. 1657, 1663 (1999).

would.[211] Objectively, going back for a second serving seems like a trivial cost; it's only in the moment that we face it that it seems unduly burdensome.[212]

Nudges can therefore offer a third alternative to sticks and carrots, which may be useful in some policy scenarios. For example, I've argued that, despite the deep political obstacles to a successful global stick to prevent climate change, policy makers should avoid carrots due to their negative incentive effects.[213] Climate nudges, such as those Sunstein proposes in a forthcoming book chapter,[214] could be a third possibility, allowing for some incremental progress without triggering the kinds of gamesmanship carrots would. On the other hand, the people for whom nudges are most effective are exactly those who are the least forward-looking—they aren't the kind of people who weigh the future heavily in their present decisions. So anticipation effects of all kinds are smaller for that population, reducing the difference between nudges and more traditional instruments.[215]

4. Framing.—Finally, as I mentioned briefly in *The Tragedy of the Carrots*, and as Eyal Zamir discusses at length in his recent work, sticks may be more effective than carrots because of the way that humans perceive redistribution.[216] Some evidence suggests that we tend to respond more strongly to events we perceive as losses than we do to events framed as gains.[217] I posited that, because these framing effects are often manipulable and may be temporary, they likely should not be a central component of

---

211. *Cf.* Jonathan Gruber, *Tobacco at the Crossroads: The Past and Future of Smoking Regulation in the United States*, J. ECON. PERSP., Spring 2001, at 193, 202–03 (summarizing studies finding that expectation of future price changes affects current consumption of willpower goods).

212. Camerer et al., *supra* note 17, at 1219, 1222.

213. Galle, *supra* note 27, at 845–46.

214. Cass R. Sunstein, *Behavioral Economics, Consumption, and Environmental Protection*, *in* HANDBOOK ON RESEARCH IN SUSTAINABLE CONSUMPTION (Lucia Reisch & John Thøgerson eds., forthcoming) (manuscript at 2–3), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2296015.

215. *Cf.* George Loewenstein & Ted O'Donoghue, *"We Can Do This the Easy Way or the Hard Way": Negative Emotions, Self-Regulation, and the Law*, 73 U. CHI. L. REV. 183, 189 (2006) (noting that *ex post* incentives are ineffective for individuals with unusually high time discounting).

216. Galle, *supra* note 27, at 816; Eyal Zamir, *Loss Aversion and the Law*, 65 VAND. L. REV. 829, 843–85 (2012).

217. Zamir, *supra* note 216, at 834–43. For skepticism about some but not all of this evidence, see Charles R. Plott & Kathryn Zeiler, *The Willingness to Pay–Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations*, 95 AM. ECON. REV. 530, 537–38 (2005) and Gregory Klass & Kathryn Zeiler, *Against Endowment Theory: Experimental Economics and Legal Scholarship* 27–42 (Georgetown, Pub. Law and Legal Theory Research Paper Series, Paper No. 13-013, 2013), *available at* http://ssrn.com/abstract=2224105.

price-instrument policy.[218] Zamir, though acknowledging the manipulability of framing, suggests that loss aversion is nonetheless pervasive enough to be the source of important moral intuitions, such as tort law's differential treatment of negligent injury and negligent failure to rescue.[219]

Whether Zamir is closer to right than I am or not, transferless instruments could potentially represent a middle path of loss aversion. As far as I am aware, there is no clear evidence on whether individuals perceive the cost of overcoming defaults or other nonmonetary inconveniences as "losses." But given that we know some actors do not even notice that defaults have changed their behavior, it would be surprising if *on average* individuals viewed defaults as being as costly as explicit prices of similar magnitude. Nudges therefore offer policy makers a third option in the loss aversion continuum. Loss aversion presents policy makers with a trade-off. With lower loss aversion, they may get less deterrence per dollar of penalty.[220] But they also get less bitter political opposition from incumbent producers. Therefore, nudges might not be as effective as sticks, but they might be more politically achievable.[221]

D. Targeting

A final important area where price instruments may diverge is their ability to be targeted or "tagged" most precisely. As others have shown, asymmetric policies can help to resolve serious targeting problems in the regulation of internalities.[222] Prior commentary has not explored whether this same argument still applies for the regulation of externalities; the basic logic of the targeting argument seems implausible in that context. I will

---

218. Galle, *supra* note 27, at 808–09; *see also* Yuval Feldman, *The Complexity of Disentangling Intrinsic and Extrinsic Compliance Motivations: Theoretical and Empirical Insights from the Behavioral Analysis of Law*, 35 WASH. U. J.L. & POL'Y 11, 38–39 (2011).

219. Zamir, *supra* note 216, at 884, 887–90. For recent evidence that the framing of policies as tax or subsidy matters, see Homonoff, *supra* note 98, at 3–4.

220. For discussion of whether it would ever be optimal for producers to perceive prices as being in excess of their true cost, see Galle, *supra* note 121 (manuscript at 31–33) (short answer: probably not).

221. Another way in which nudges might be more politically viable is if internality sufferers are aware of their own problems but underestimate them. Then demand for an internality-correcting stick will be low. *See* Strnad, *supra* note 30, at 1256–57. But these same households would also presumably underestimate the cost of a future nudge, which could allow for a much costlier nudge than would be possible if the commitment device were structured as a monetary penalty. *Cf.* Katrina Fischer Kuh, *When Government Intrudes: Regulating Individual Behaviors that Harm the Environment*, 61 DUKE L.J. 1111, 1175–76 (2012) (noting that individuals may not oppose regulations that impose costs on them only indirectly); Michael P. Vandenbergh, *Order Without Social Norms: How Personal Norm Activation Can Protect the Environment*, 99 NW. U. L. REV. 1101, 1103–04 (2005) (arguing that changing conservation norms through information and other informal regulatory devices is more politically viable than price-based mandates).

222. *See supra* notes 95–97 and accompanying text.

argue that asymmetric approaches can have targeting advantages for positive externalities, mixed internality/externality cases, and instances in which human error weakens the power of conventional instruments.

1. Targeting Externalities.—The power of targeting lies in its ability to reduce unnecessary welfare costs of regulation. When individuals vary in their propensity for errors and willpower lapses, a uniform price or regulation may inefficiently distort incentives or otherwise produce deadweight loss.[223] For example, taxes on cigarettes might help some smokers who want to quit to steady their resolve. But other smokers might be "rational addict[s]," in Gary Becker's famous turn of phrase: they are well-informed, respond fully to the long-term costs they face, and accept them.[224] For them, the tax simply imposes pain or misshapes their preferences, a classic case of deadweight loss.[225]

Nudge proponents claim that asymmetric regulation accounts for heterogeneity by allowing costs to vary together with the need for correction.[226] That is, those who treat the costs of a nudge as larger also may tend to be those who suffer from internalities. The irrational smoker perhaps smokes because he focuses excessively on his present satisfaction. That same trait will make the burden of, say, putting on his coat and stepping outside to smoke much more irksome than it would be for others who weren't similarly present biased. So an indoor-smoking-ban nudge corrects the internality for those who suffer it while imposing rather small costs on those who don't.

It seems much harder to tell this story about externalities.[227] Nudges' targeting works best where susceptibility to the nudge is correlated with the harm to be prevented.[228] Toxic waste, though, is equally harmful no matter who emits it.

Nudges may, however, help overcome the problems caused by inframarginal producers of positive externalities, which can be thought of as an aspect of targeting.[229] Inframarginality can make carrots an especially

---

223. Strnad, *supra* note 30, at 1252, 1254–55.

224. Gary S. Becker & Kevin M. Murphy, *A Theory of Rational Addiction*, 96 J. POL. ECON. 675, 676–78 (1988).

225. *See* De Geest & Dari-Mattiacci, *supra* note 72, at 362–65 (pointing out that sticks may burden actors with higher than average costs of compliance). If the tax is very small, though, the behavioral effect on rational consumers should be negligible, in which case internality gains should exceed any deadweight loss. O'Donoghue & Rabin, *supra* note 151, at 1835.

226. *See* Allcott et al., *supra* note 12, at 2, 23.

227. This difficulty has not dissuaded nudge proponents. *See* sources cited *supra* note 88.

228. Camerer et al., *supra* note 17, at 1219, 1222.

229. *See* Lily L. Batchelder et al., *Efficiency and Tax Incentives: The Case for Refundable Tax Credits*, 59 STAN. L. REV. 23, 45–46 (2006) (arguing that targeting is important for positive externalities because of the social cost of paying for subsidies). The marginal actor is the person who is just on the knife-edge of deciding what to do; with a bit of stick or carrot, they will change

wasteful policy choice. Carrots are generally awarded to everyone who goes along with the regulator's goals, so that there is no easy way to sort out those who would have done so anyway from those who needed some extra incentive.[230] Every itemizer who donates to charity gets a charitable contribution deduction, even if they would have given out of religious obligation or personal generosity.[231] That's most problematic for positive externalities because our intuition is that voluntary reductions in negative externalities are relatively rare. Any transferless instrument, including a nudge, can help on this front by reducing the social cost of transfers to inframarginal producers.

Both carrot and stick transfers can also actually reduce positive contributions from inframarginal producers by crowding out their internal motivation.[232] Researchers find that offering explicit monetary rewards can diminish voluntary contributions.[233] The psychological mechanism is uncertain.[234] Some psychologists suggest that monetary incentives are particularly apt to generate resistance because they reduce our sense of autonomy.[235] Possibly the dollar award attracts excessive focus, distracting volunteers from the more abstract reasons they held previously.[236] Being paid may also diminish the "warm glow" signal that donors usually experience: some individuals may behave altruistically because they want to be recognized by others as altruistic, and when there is an explicit monetary incentive, that signal is muddied.[237]

---

their behavior. "Inframarginal" agents are so committed to their path that the incentive effects of the price instrument are not important.

230. De Geest & Dari-Mattiacci, *supra* note 72, at 370; Galle, *supra* note 27, at 820–21, 833–34. Under an effective enforcement regime, sticks will fall only on actors who defy the government's objectives, and so this form of inframarginality problem doesn't arise.

231. *See* I.R.C. § 170 (2006). I omit mention of some technical limits that don't affect my point.

232. Wendy J. Gordon, *Of Harms and Benefits: Torts, Restitution, and Intellectual Property*, 21 J. LEGAL STUD. 449, 457 (1992); Andrew Green, *You Can't Pay Them Enough: Subsidies, Environmental Law, and Social Norms*, 30 HARV. ENVTL. L. REV. 407, 432–35 (2006).

233. For a helpful overview, see Feldman, *supra* note 218, at 23–29.

234. *Id.* at 24.

235. Edward L. Deci & Richard M. Ryan, *The Empirical Exploration of Intrinsic Motivational Processes*, *in* 13 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 39, 61 (Leonard Berkowitz ed., 1980).

236. *See* Bruno S. Frey & Alois Stutzer, *Environmental Morale and Motivation*, *in* THE CAMBRIDGE HANDBOOK OF PSYCHOLOGY AND ECONOMIC BEHAVIOUR 406, 412–13 (Alan Lewis ed., 2008) (suggesting that crowding out may be due to an individual's shift in focus from internal to external motivations).

237. *See* Ernst Fehr & Bettina Rockenbach, *Detrimental Effects of Sanctions on Human Altruism*, 422 NATURE 137, 140 (2003); William M. Landes & Richard A. Posner, *Salvors, Finders, Good Samaritans, and Other Rescuers: An Economic Study of Law and Altruism*, 7 J. LEGAL STUD. 83, 99 (1978); *cf.* E. Tory Higgins & Yaacov Trope, *Activity Engagement Theory: Implications of Multiply Identifiable Input for Intrinsic Motivation*, *in* 2 HANDBOOK OF MOTIVATION AND COGNITION: FOUNDATIONS OF SOCIAL BEHAVIOR 229, 240–43 (E. Tory

Nudges can offer a partial solution.[238] Several studies find that nudge-like interventions have improved altruistic behavior.[239] Because the nudge's incentive effect is not easily visible to others, it may not confuse the altruistic signal to the same extent dollars do. And the implicit "price" of the nudge is more subtle, and thereby perhaps less likely to reduce self-perceived autonomy or to assume more salience than the donor's other motives in his mind.

2. Mixed Internality/Externality Cases.—Another plausible story for why asymmetric regulations might be effective for externalities is if there are also internalities happening at the same time. For example, we might want to help the planet, but lack the willpower to dedicate ourselves on a daily basis to such a long-term goal.[240] Or we might fail ourselves in ways that also impose large harms on others, such as by overconsuming expensive energy sources or leaving ourselves vulnerable to risks that others will ultimately have to help us overcome.[241] In these scenarios, there is once again potentially a strong correlation between the subjective costs of the nudge or other surprising instrument and the need for regulation. For instance, instant energy feedback from "smart meters" will not do much to change the behavior of those who are already very energy conscious, but then it will not cost them much, either.

O'Donoghue and Rabin argue that taxes, too, can achieve the targeting benefit that they (in their joint work with Camerer and Issacharoff) attribute to nudges, but their claim relies on a questionable assumption. They posit that since only low willpower individuals will continue to consume tempting goods subject to a tax, rational consumers will not pay the tax.[242] But the optimal tax may vary considerably across individuals. For instance, soda drinkers who are more prone to diabetes may represent a greater

---

Higgins & Richard M. Sorrentino eds., 1990) (noting that extrinsic incentives may reduce perception that an actor was self-motivated).

238. *See* Amir & Lobel, *supra* note 19, at 2130–32 (suggesting that behaviorally informed regulation may be able to reduce crowding-out effects).

239. *E.g.*, Alberto Abadie & Sebastien Gay, *The Impact of Presumed Consent Legislation on Cadaveric Organ Donation: A Cross-Country Study*, 25 J. HEALTH ECON. 599, 606–13 (2006); James Andreoni & Justin M. Rao, *The Power of Asking: How Communication Affects Selfishness, Empathy, and Altruism*, 95 J. PUB. ECON. 513, 515–17 (2011).

240. Andrew Green, *Self Control, Individual Choice, and Climate Change*, 26 VA. ENVTL. L.J. 77, 78–79 (2008); *see also* Leonhard K. Lades, *Impulsive Consumption and Reflexive Thought: Nudging Ethical Consumer Behavior*, J. ECON. PSYCHOL. (forthcoming) (manuscript at 9–10) (on file with author); *cf.* Ayres, *supra* note 52, at 2088 (proposing use of sticky defaults to account for heterogeneity in production of externalities); Sunstein & Thaler, *Libertarian Paternalism*, *supra* note 17, at 1192–93 (discussing "libertarian benevolence," which appears to be the use of nudges to encourage positive externalities).

241. *See* Galle & Utset, *supra* note 89, at 72–77 (discussing the impact of time inconsistency, poor planning, and procrastination by certain households in energy consumption).

242. O'Donoghue & Rabin, *supra* note 151, at 1831, 1835.

potential social cost. High taxes on drinkers with low propensity to create externalities would simply represent deadweight loss, undermining their targeting argument. Nudges may also be better targeted in the sense that they may be better capable of changing the behavior of individuals who are usually inattentive to costs and benefits and so would not be much influenced by a tax.[243] But this is not to say that advances in tax design could not potentially match nudges' targeting potential in the future.[244]

3. Better than Errors.—Lastly, asymmetric regulation might be an effective supplement or alternative when human errors render less surprising instruments ineffective. Take the case of individual retirement arrangements, or IRAs, an important tax incentive for retirement savings.[245] One justification for IRAs is that, if households do not save now, the public will have to care for them when they are old and infirm.[246]

Despite their enormous annual cost—more than $125 billion annually[247]—IRAs and related retirement provisions don't seem to work, and it is easy to see why. Imagine that the government will pay you to overcome your tendency to procrastinate planning for retirement, but in order to collect your reward you have to read some program rules written in bureaucrat, find household records that establish your eligibility, go through some complex calculations, and fill out and submit government forms. Quite probably, serious procrastinators are the very last people who would benefit from that program.[248] But that is exactly the structure of IRAs.[249] Unsurprisingly, then, Chetty et al. find massive mistargeting of similar retirement incentives in Denmark, with about 85% of the beneficiaries, by their estimation, receiving subsidies that do not meaningfully change behavior.[250]

Asymmetric regulation is useful for these situations because it helps to patch gaps created by variations in the public's responsiveness to traditional regulation. Though some people are too inattentive or impatient for

---

243. *Cf. id.* at 1835 (acknowledging that internality-correcting taxes are inefficient unless "people with self-control problems are sensitive to tax changes").

244. O'Donoghue & Rabin, *supra* note 164, at 109, offer some examples.

245. STAFF OF JOINT COMM. ON TAXATION, 110TH CONG., JCX-53-08, PRESENT LAW AND ANALYSIS RELATING TO INDIVIDUAL RETIREMENT ARRANGEMENTS 16 (2008), *available at* https://www.jct.gov/publications.html?func=startdown&id=1286.

246. POSNER, *supra* note 203, at 498–99.

247. CONGRESSIONAL BUDGET OFFICE, THE DISTRIBUTION OF MAJOR TAX EXPENDITURES IN THE INDIVIDUAL INCOME TAX SYSTEM 6 tbl.1 (2013).

248. This is not a problem entirely exclusive to carrots. For example, sticks that are imposed long after the unwanted behavior are poorly targeted because of excessive time discounting. Loewenstein & O'Donoghue, *supra* note 215.

249. *See* STAFF OF JOINT COMM. ON TAXATION, *supra* note 245, at 16–46 (discussing the present law surrounding IRAs).

250. Chetty et al., *supra* note 14 (manuscript at 43–44).

monetary incentives to be fully effective, others may be more sensitive. Even in the Chetty et al. study, 15% of Danes did respond quite readily to retirement incentives.[251] Government could increase the size of its incentives to try to grab people's attention, but then we also are showering extra funds—or extra penalties—on those who were already paying attention, leading to complicated trade-offs.[252] So again we have a targeting problem, with potentially a strong correlation between some kinds of mental errors and the need for additional government regulation. Asymmetric regulation targeted to affect those who are especially insensitive to prices would therefore seem a good policy alternative.

I should acknowledge, though, that traditional prices may have an advantage over nudges when the optimal tax schedule is complex. Recall that—setting aside some possible complications—the optimal Pigouvian price should be set equal to marginal social damage. For some externalities, that damage could vary considerably depending on, say, the consumer's prior health history, his family situation, where he lives, and so on.[253] Alcohol consumption is a likely example, especially since small amounts of alcohol may actually improve some health outcomes.[254] With enough information, an *ex ante* tax can approximate these effects, and with a reliable enough system of proof, an *ex post* liability system can as well.[255]

It isn't clear whether nudges can. If susceptibility to the nudge happens to be closely correlated with propensity to produce externalities, the impact of the nudge could vary with the marginal damage, but this may not always be possible. But most commentators believe that the informational demands of such a flexible tax are also usually unrealistic,[256] so this may not be a significant weakness of nudges.

IV. Examples

We now have the tools to evaluate New York City's beverage-size limits, and a number of other innovative policy proposals, too. The results are a bit surprising. The superiority of taxes or other stick transfers, which prior commentators have almost universally assumed, in some instances is not so clear. Maybe less surprisingly, carrots often look even worse than they did when nudges were not in the picture, as nudges in many cases can substitute for carrots without presenting the same risks.

---

251. *Id.* at 36.

252. For extended analysis of these trade-offs, see Galle, *supra* note 121 (manuscript at 13–35).

253. *See* Kaplow & Shavell, *supra* note 46, at 4–5 (indicating marginal social harm can be nonlinear or unfixed).

254. Strnad, *supra* note 30, at 1244.

255. *See* Jon D. Hanson & Kyle D. Logue, *The Cost of Cigarettes: The Economic Case for Ex Post Incentive-Based Regulation*, 107 YALE L.J. 1163, 1268–74 (1998).

256. O'Donoghue & Rabin, *supra* note 151, at 1830; Strnad, *supra* note 30, at 1271–72.

A. Soda

Let's begin our examples with the recent controversy over New York's sugary-beverage policy. The City Health Department justified its proposal primarily with an externality story: soda contributes significantly to obesity, which in turn imposes serious cost burdens on publicly funded health services.[257] The policy may also help to protect consumers against themselves, but for my purposes here not much changes if we also include this "internality" story.[258] Either way, the factors I have identified somewhat favor nudge-type approaches, such as the city's cap, over a soda tax or similar stick-like instrument, such as cutting subsidies to beverage ingredients or increasing tort liability for beverage producers.[259]

Size limits are better targeted at soda drinkers' potential internalities than a tax would be. The default size is most binding on individuals with high discount rates and excessive focus on the present.[260] Caffeine quaffers who excessively discount the future will more likely view the bother of obtaining a second cup as disproportionately large relative to the later benefits of quenching their thirst.[261] Similarly, those who are the most focused on their immediate surroundings would be the most likely to be influenced by the size of the portion in front of them.[262] These two groups are also those who predictably will not accurately account for the future cost of their consumption when they make present drinking decisions.

---

257. Thomas A. Farley, *NYC Health Commissioner: Limiting Soda Is the Right Way to Protect the Health of New Yorkers*, FORBES (June 11, 2013, 7:30 AM), http://www.forbes.com/sites/robwaters/2013/06/11/nyc-health-commissioner-limiting-soda-size-is-the-right-way-to-protect-the-health-of-new-yorkers/; *see also* Korobkin, *supra* note 84, at 1681–82. For evidence that soda consumption contributes to obesity, and therefore to obesity-related health problems, see generally Cara B. Ebbeling et al., *A Randomized Trial of Sugar-Sweetened Beverages and Adolescent Body Weight*, 367 NEW ENG. J. MED. 1407 (2012); and Janene C. de Ruyter et al., *A Trial of Sugar-Free or Sugar-Sweetened Beverages and Body Weight in Children*, 367 NEW ENG. J. MED. 1397 (2012).

258. Jeff Strnad, in his exhaustive 2005 analysis, argued that "fat taxes" in general were best defended as a form of implicit insurance premium charged to consumers who would later put demands on the health-care system rather than as internality-correcting. Strnad, *supra* note 30, at 1234, 1267–68. But Strnad did not argue there were no internalities, only that taxes could not be targeted accurately enough at internality sufferers. *Id.* at 1322. Nudges may improve targeting enough to overcome Strnad's objections.

259. For discussion of the role of government subsidies in excess beverage consumption, see Adam Benforado et al., *Broken Scales: Obesity and Justice in America*, 53 EMORY L.J. 1645, 1791–95 (2004).

260. *See* James J. Choi et al., *Optimal Defaults*, 93 AM. ECON. REV. (PAPERS & PROCS.) 180, 180–81 (2003) (modeling properties of defaults generally).

261. For evidence that obesity may be the product of impatience, see sources cited *supra* note 115.

262. *See* Chandon, *supra* note 10, at 16 (connecting overeating to temptation and misperceptions of the true size of food portions); Andrew B. Geier et al., *Unit Bias: A New Heuristic That Helps Explain the Effect of Portion Size on Food Intake*, 17 PSYCHOL. SCI. 521, 524 (2006) (suggesting that "immediate presence" of temptation helps to explain the influence of portion size on consumption).

In contrast, a soda tax would likely fall on all consumers, including those who are not at any risk of obesity and those who have rationally concluded that the risks are worth the costs. At least for those at low risk, the tax simply imposes pain or distorts behavior without any offsetting gain. On the other hand, this argument presumes that any soda tax would necessarily have to be linear—that is, that we would impose the same price per serving for all consumers. If a more flexible schedule were possible, such that those who are at greater risk of health consequences paid higher prices, then the tax might be better targeted than the nudge. It is very unlikely the ideal portion size is identical for all consumers.[263] But realistically it also seems very improbable that either the tax or the portion size could be set to vary with real marginal costs.

The size limit may also be better targeted in the sense that it reduces the extent to which internality sufferers substitute into other unhealthy behaviors. For example, taxes on soda could encourage consumers to switch to other unhealthy choices.[264] Will soda drinkers similarly switch to sugary juices in order to be able to buy them in larger sizes? Though of course time will tell, the soda nudge might not produce much of this kind of switching. To induce switching, the would-be consumer must recognize in advance that she will want additional consumption and also recognize that she will then be unwilling to pay the price to overcome the default. As we have seen, both of these are uncertain: the consumption decision may be the product of the portion size the consumer experiences, and her ability to predict her perception of the price may be limited.

On the other hand, the soda tax certainly brings in more dollars than the size limit, but the welfare effects of that swap are less clear. As I argued earlier, it is possible that consumers would perceive an explicit tax to reduce their returns to labor, while not noticing or even appreciating the effects of a similar nudge. The beverage size limit seems a good example of where it is plausible that consumers would not connect the nudge to their labor/leisure decision, since again there will be consumers who do not even notice that the smaller portion size changed their preferences. If so, then the greater revenues of the soda tax also come at some additional social cost, and it is ambiguous whether the opportunity they offer to cut other taxes (or invest in worthwhile new government programs) makes society better off on net.

The nudge option does seem to have better distributive outcomes. Studies find that the population at greatest risk from excessive sugary

263. Strnad, *supra* note 30, at 1321.

264. Eric A. Finkelstein et al., *Impact of Targeted Beverage Taxes on Higher- and Lower-Income Households*, 170 ARCHIVES INTERNAL MED. 2028, 2033 (2010) (raising the substitution question); Jason M. Fletcher et al., *The Effects of Soft Drink Taxes on Child and Adolescent Consumption and Weight Outcomes*, 94 J. PUB. ECON. 967, 972–73 (2010) (reporting evidence that taxes cause substitution to other unhealthy beverages).

beverage consumption tends to be rather poorer on average than others.[265] So the soda tax has a good chance to be even more regressive than a standard sales tax. Some commentators have suggested mitigating that unfortunate outcome either through paying back the tax's proceeds to low-income households through cash rebates or by offering subsidies for healthy food options.[266] But note that some of these alternatives are either identical to or (if subsidies exceed taxes collected) inferior on revenue terms to the nudge. If all revenues are rebated, the nudge and tax are identical, except that the nudge is better targeted.[267]

Finally, and cutting in the other direction, in the absence of a rebate taxes could have an advantage when it comes to income effects. Both the tax and the default could help the consumer to better allocate her available budget, creating a positive income effect—as the household feels richer, they demand more goods, including foods that could contribute to obesity. The tax, however, reduces the consumer's household wealth, likely diminishing her demand for soda.[268] On the other hand, some evidence suggests that junk food is an inferior good,[269] in which case the nudge is better: by leaving the household with more money, the nudge diminishes its demand for the unhealthiest foods.

On net, the case for surprising alternatives to a soda tax is surprisingly good. That is not to say that the 16-ounce cap is the best such policy. Right now, we have no particular reason to think that 16 ounces is the optimal serving size. But additional policy experiments can help to identify which transferless policies are best.

B. Retirement

It is interesting also to consider an instance where nudges could replace carrot transfers. Retirement savings offer a major example. In their study of Danish workers, Chetty et al. appear to endorse proposals replacing

---

265. Valerie Gebara & Leena Gupta, *Consumption of Sugar Sweetened Beverages (SSBs) in New York City*, EPI DATA BRIEF (N.Y.C. Dep't of Health and Mental Hygiene), May 2011, at 1–2, *available at* http://www.nyc.gov/html/doh/downloads/pdf/epi/databrief4.pdf.

266. Battle & Brownell, *supra* note 30; Pratt, *supra* note 30, at 124–25; *see also* Sugarman & Sandman, *supra* note 30, at 1489 (proposing rebates to help states cover the costs of obesity reduction).

267. The tax might produce extra revenue if only a fraction is rebated—for instance, if only lower-income households collect a refund. But note that a means-tested rebate in effect is an income tax. *See* KAPLOW, *supra* note 58, at 153–54. Such a rebate would thus create additional economic distortions.

268. *See* Gideon Yaniv et al., *Junk-Food, Home Cooking, Physical Activity and Obesity: The Effect of the Fat Tax and the Thin Subsidy*, 93 J. PUB. ECON. 823, 826–27 (2009) (arguing that subsidies for healthy foods could increase demand for unhealthy foods through income effect).

269. Matthew Harding & Michael Lovenheim, The Effect of Product and Nutrient-Specific Taxes on Shopping Behavior and Nutrition: Evidence from Scanner Data 16 (Mar. 4, 2013) (unpublished manuscript), *available at* http://www.stanford.edu/~mch/resources/Harding_Nutrition.pdf.

tax incentives for retirement contributions, such as the 401(k) plan (and its lesser-known cousins, such as § 403(b)) with employer-administered default contributions to workers' retirement accounts.[270] Though they do not frame it in precisely the terms I have set out here, in essence their claim is that defaults are better targeted and less costly for the government.[271] Again, their study finds that "inattentive" investors save more when default contributions are ratcheted up, but that those investors ignore (while still benefiting from) tax incentives.[272] And inattentive investors make up 85% of the Danish working population.[273] Thus, they claim that default contributions both require little government investment and also reduce the likelihood of giving money to inframarginal agents.[274]

My analysis supports this story, but suggests some possible qualifications. For one, the welfare benefits of default savings may be smaller than Chetty et al. assume. Eliminating § 401(k) could save on the order of $125 billion annually, allowing the government to lower overall tax rates and reduce the deadweight loss of federal taxation.[275] But the default may also generate deadweight loss, not only because it may not match the preferences of some "inattentive" investors but also because it might affect their labor supply. If workers who ignore retirement are in fact motivated only by today's take-home pay, they may perceive an extra 6%

---

270. *See* Chetty et al., *supra* note 14 (manuscript at 43–44) (suggesting that automatic contributions are preferable to tax incentives). For a summary of past legislative efforts, see STAFF OF JOINT COMM. ON TAXATION, *supra* note 245, at 48–49. For other academic supporters of opt-in defaults as a solution to the retirement savings problem, see generally J. Mark Iwry & David C. John, *Pursuing Universal Retirement Security Through Automatic IRAs*, *in* AGING GRACEFULLY: IDEAS TO IMPROVE RETIREMENT SECURITY IN AMERICA 45 (Peter Orszag et al. eds., 2006); William G. Gale et al., *The Saver's Credit: Savings for Middle- and Lower-Income Americans*, *in* AGING GRACEFULLY, *supra*, at 77; Karen C. Burke & Grayson M.P. McCouch, *Social Security Reform: Lessons from Private Pensions*, 92 CORNELL L. REV. 297, 308–10 (2007); Camerer et al., *supra* note 17, at 1227–29; Edward A. Zelinsky, *The Defined Contribution Paradigm*, 114 YALE L.J. 451, 523–24 (2004).

271. *See* Chetty et al., *supra* note 14 (concluding that one reason automatic contributions to savings are more effective than price subsidies is that "policies that influence the behavior of passive savers have lower fiscal costs, generate relatively little crowd-out, and have the largest impacts on individuals who are paying the least attention to saving for retirement").

272. *Id.*

273. *Id.*

274. *Cf. id.* at 38–39 (discussing how a government price subsidy has a small impact on savings because the subsidy is an inframarginal transfer with a low interest elasticity of savings for active savers). Chetty et al. also claim that tax incentives don't increase net savings even among attentive households because in their sample, tax incentives just encourage investors to move money from existing savings into tax-favored accounts. *Id.* at 43. This is an important result, but it doesn't necessarily imply that the incentives are fruitless. Eligible, retirement-savings vehicles may be much stickier than other savings—among other reasons, because there is a statutory penalty for withdrawal. If the government's goal is long-term savings, moving money into these stickier accounts may, therefore, still be a somewhat good investment.

275. Or, of course, the government could spend the $125 billion in some other way, if the alternative generates greater welfare gains than the tax cut.

set-aside out of current earnings as the equivalent of a 6% tax.[276] That might be a bigger effective hike than any cut that could accompany the $125 billion windfall.[277]

Chetty et al. also appear to assume that switching away from the 401(k) carrot will better align income effects, but that isn't necessarily the case. They echo a common criticism of carrots for retirement, which is that increases in household wealth tend to stimulate consumption, while the goal of the policy is to encourage savings and, therefore, to reduce current consumption.[278] As we have seen though, it is possible that a very well-targeted default could also be perceived as expanding the household's budget. Workers could see their returns to labor as higher, since they will not be wasting as much money on short-term temptations. So, in short, a more complete assessment of their proposal requires us to know more about how inattentive investors respond to default savings.[279]

Perhaps a central theme to both the soda and retirement examples is that the labor-supply effects of a surprising instrument depend on the nature of the error individuals are making. When what is happening is a failure of will, rather than knowledge, labor supply seems most likely to increase. In this scenario, some households know that they are getting a valuable commitment device from the government and recognize the improved budget allocation that device allows them. In contrast, when the error is a mistake of attention or understanding, families could well reduce their labor in response to the nudge because they do not see that the government has actually made them better off. Future empirical work devoted to better identifying how people are going wrong could therefore have significant policy implications.

---

276. *See* Louis Kaplow, *Myopia and the Effects of Social Security and Capital Taxation on Labor Supply* 1–2 (Nat'l Bureau of Econ. Research, Working Paper No. 12452, 2006), *available at* http://www.nber.org/papers/w12452 ("[E]ven an actuarially fair social security retirement scheme . . . might be imagined to have similar effects to a current tax on labor if individuals are significantly myopic . . . .").

277. Chetty et al. also suggest that employers could be convinced to offer default retirement accounts with relatively small government incentives, perhaps just enough to cover the costs of administration. *See* Chetty et al., *supra* note 14. That, too, is unclear. If employers currently capture some or all of the benefit of the government subsidy, persuading them to agree politically to the swap would likely require a promise to replace much of the current savings. Of course, Chetty et al. might argue that since workers seem not to care about retirement savings, we might think that employers cannot save much in the way of lower salaries, and therefore cannot capture much of the subsidy. But if that is true, then workers would likely perceive default savings as a tax, which means that the social welfare benefits of their proposal are smaller than they suggest.

278. GRUBER, *supra* note 32, at 650.

279. Another complication in the analysis is that shifting income from one period of life to another also alters the marginal utility of a dollar for the worker. This factor can interact with myopic preferences in complex ways. For a more complete discussion, see Kaplow, *supra* note 276, at 4–13.

C. *Pollution*

Regulators and inventive commentators have proposed and sometimes even road-tested a variety of nudge-like instruments for reducing carbon and other forms of pollution.[280] Some of these interventions have been aimed at consumers, such as the various kinds of cues and defaults to reduce household energy consumption championed by Richard Thaler and Cass Sunstein and by Ian Ayres.[281] Although it's unclear that even in combination all the proposed climate nudges could achieve the needed levels of carbon reduction,[282] energy-conserving nudges look to be at least a valuable component of any strategy. Consider first the household-level nudges, such as "smart" energy meters that offer instant feedback to households on their energy usage.[283] It seems pretty straightforward that these kinds of efforts are preferable to carrot transfers, such as paying families to adopt conservation strategies, subsidizing weatherproofing, and so on. Given the potentially vast number of inframarginal claimants for such subsidies, the nudges almost certainly cost less, and either way will have a lesser unwanted income effect on household energy consumption.

Less intuitively, the constellation of nudges could outshine a carbon tax. At first glance, the nudges seem to sacrifice any possible revenue recycling benefits from the carbon proceeds. But nearly all carbon-tax proposals include efforts to mitigate the severe regressivity of taxes on carbon, which function as a broad-based sales tax due to the energy involved in manufacturing and transporting nearly any consumer good.[284] In many proposals, most or all of the revenues from the carbon tax or its equivalents (such as the 2009 cap-and-trade bill I mentioned earlier) would be devoted to cash rebates for low-income households, making the carbon tax close to transferless.[285] With some experiments, the nudges could be matched to a shadow price, allowing them to approximate the informational

---

280. Vandenbergh et al., *supra* note 29, at 763–79.

281. THALER & SUNSTEIN, *supra* note 8, at 257–61; AYRES, *supra* note 8, at 138–42.

282. For consideration of that question, see Thomas Dietz et al., *Household Actions Can Provide a Behavioral Wedge to Rapidly Reduce US Carbon Emissions*, 106 PROC. NAT'L ACAD. SCI. 18,452, 18,452 (2009) (estimating that behavioral changes can reduce U.S. emissions by 7.4% over the next decade).

283. Leslie Kaufman, *Utilities Turn Their Customers Green, with Envy*, N.Y. TIMES, Jan. 30, 2009, http://www.nytimes.com/2009/01/31/science/earth/31compete.html?_r=0.

284. *E.g.*, Gilbert E. Metcalf & David Weisbach, *The Design of a Carbon Tax*, 33 HARV. ENVTL. L. REV. 499, 514 (2009); David A. Super, *From the Greenhouse to the Poorhouse: Carbon-Emissions Control and the Rules of Legislative Joinder*, 158 U. PA. L. REV. 1093, 1190–96 (2010).

285. For discussion of several alternatives, see Terry Dinan & Diane Lim Rogers, *Distributional Effects of Carbon Allowance Trading: How Government Decisions Determine Winners and Losers*, 55 NAT'L TAX J. 199, 205–06 (2002).

benefits of the carbon tax.[286] So on these two traditional criteria, nudges and taxes are similar.

Alternately, if the carbon tax does not include a rebate or rebates only a portion of poorer households' average costs, nudges could present a trade-off between revenues and fairer distribution. A nudge does still arguably impose costs on some families—those who must exert effort to avoid or ignore it. For instance, with greater feedback I might feel bad that I am using too much energy. But these deadweight losses are more likely to be equitably distributed across households rather than being borne most heavily by the poorest. It certainly could be that the subjective mental costs of avoiding energy-conserving nudges are greater for individuals with less wealth, but it is not immediately obvious why that would be so. Early empirical evidence on the distribution of the costs of mental effort are unclear, with one or two papers actually finding that richer people seem to view effort as more costly.[287]

So far, nudges and carbon taxes are roughly equivalent, but nudges are probably better targeted in a couple of different respects. Even with rebates, households retain a marginal incentive to conserve energy under a carbon tax since, if my rebate is determined by everyone else's average costs, I can come out ahead by being thriftier than they are.[288] Again, if government cannot readily connect this marginal incentive with my effective wealth, then a marginal dollar in incentives will overmotivate the poor while undermotivating the very rich. Many conservation nudges in contrast can be designed to affect primarily those who need greater interventions. A thermostat set to automatically lower temperatures on winter evenings is more likely to change the behavior of households who are inattentive to energy use than those who are already paying attention. These are also families who may well derive some additional internality benefit from the nudge.[289]

A final factor to consider is crowd-out. Even if Glaeser is right that in an economic sense nudges are every bit as "coercive" as taxes, not everyone may see things the same way.[290] As we saw earlier, express dollar-denominated incentives tend to replace other intrinsic motivation, but

---

286. Another important goal of carbon taxes is to not only reduce overall energy consumption but also to shift the sources of energy to less carbon-intensive uses. Nudges can also be designed to encourage switching. For instance, in addition to reporting total usage, smart meters could report the mix of sources being drawn from the grid and allow the household to dynamically adjust which source it prefers.

287. *E.g.*, Jacob Goldin & Tatiana Homonoff, *Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity*, 5 AM. ECON. J.: ECON. POL'Y 302, 331 (2013).

288. *See* KAPLOW, *supra* note 58, at 2–3 (explaining that rebates do not change first-best analysis of Pigouvian taxes).

289. Carbon taxes could likewise help the family to better prioritize its spending, but at some overall cost to them.

290. *See* sources cited *supra* note 18.

nudges might not. So carbon taxes could reduce altruistic energy conservation, while nudges might leave it unchanged or even improve it. Nudges might even help altruistic but low willpower individuals achieve the greater conservation levels they desire.[291]

D. Positive Externalities

Positive externalities offer a particularly fertile area for developing new nudges. For the most part, carrots are the dominant U.S. instrument for encouraging many important positive externalities, ranging from copyright protections for artists to tax deductions for charitable contributions and research and development.[292] As I suggested earlier, in many cases nudges can replace carrots in instances where sticks are problematic.

Charitable contributions are a possible example. Many of the tools others have designed for pension savings could also be employed for charitable giving. For example, employees could by default have a small portion of their earnings in excess of a certain threshold distributed among a short list of charities they had previously selected—say, 3% of income above $40,000.[293] Employees also could commit to donating a portion of future earnings, as Thaler and Sunstein suggest.[294] More radically, and taking a cue from Germany, the United States could collect donations for charities through the tax system without subsidizing them.[295] Realized gains on investment properties could be "taxed" an extra few percentage points unless the taxpayer opts out, with the revenues flowing to their designated charities.

---

291. Sunstein, *supra* note 214 (manuscript at 11–12).

292. *See* Galle, *supra* note 27, at 840 ("[C]arrots are commonplace—and more are, shall we say, sprouting up all the time.").

293. The mean itemizing household currently donates about 2% of personal income to charity. STAFF OF JOINT COMM. ON TAXATION, 113TH CONG., JCX-4-13, PRESENT LAW AND BACKGROUND RELATING TO THE FEDERAL TAX TREATMENT OF CHARITABLE CONTRIBUTIONS 45 (2013). If the employee never gets around to designating any beneficiary organizations, the firm could select them or the money could be distributed to charities like the United Way that do the choosing for their donors. A critic of the proposals might argue that the proposals somewhat arbitrarily cap the amount of "subsidy" the government offers. In contrast, the deduction allows donors to determine the amount of matching dollars the government will provide without limit as long as annual contributions do not exceed 50% of adjusted gross income. That is accurate, but note that it isn't inevitable that the deduction will always have this advantage. Several serious legislative proposals over the past few years would cap the annual amount of subsidized contributions for each donor. *See* ROGER COLINVAUX ET AL., EVALUATING THE CHARITABLE DEDUCTION AND PROPOSED REFORMS 12 tbl.5 (2012), *available at* http://www.urban.org/UploadedPDF/412586-Evaluating-the-Charitable-Deduction-and-Proposed-Reforms.pdf.

294. THALER & SUNSTEIN, *supra* note 8, at 231–32.

295. For an overview of the German system, see Stephanie Hoffer, *Caesar as God's Banker: Using Germany's Church Tax as an Example of Non-Geographically Bounded Taxing Jurisdiction*, 9 WASH. U. GLOBAL STUD. L. REV. 595, 601–06 (2010). Of course, aiding collection is itself a bit of a subsidy, but a much smaller one than the charitable contribution deduction currently offers.

Charitable nudges along these lines are likely superior on deadweight loss terms to the current income tax deduction for charitable contributions, while offering a somewhat less useful income effect than the deduction. Obviously the nudge would not reduce government revenues to the extent the deduction does.[296] As with retirement savings though, it is possible that donors could view the default as reducing their real returns to labor, resulting in deadweight loss that mimics the cost of the lost revenue. Donors who perceive the donation as a loss could also see themselves as poorer, reducing their demand for charity. And of course the donor pays a higher tax than she would with the deduction in place, which could further diminish her demand.

The nudges have other advantages as well. They are almost certainly better targeted than the deduction, much of the value of which is presently claimed by donors who likely would give a substantial amount regardless of the subsidy.[297] Unlike the present design of the deduction, a nudge does not reduce the progressivity of the tax system.[298] Another criticism of the deduction is that, because it offers larger rewards for higher income givers, it tends to produce a charitable sector slanted towards the interests of the rich.[299] I have also argued that, unless charities can more firmly be separated from the political sphere, the deduction distorts our politics.[300] Like a credit, the nudges I mentioned would somewhat mitigate these tendencies, though of course wealthier donors will still have more to give.

Similar nudges could also be used to supplement or replace the estate tax and its accompanying deduction for charitable bequests.[301] Though the purposes behind the income tax deduction for charitable contributions have been closely interrogated by commentators, the estate-tax deduction has mostly escaped scrutiny.[302] Most of those who have examined it are generally cheerful about its effects: in addition to subsidizing charities, it

---

296. *See* STAFF OF JOINT COMM. ON TAXATION, *supra* note 293, at 44 tbl.2 (estimating $37.6 billion in 2012 federal tax savings for charitable contributors).

297. *See* MOLLY F. SHERLOCK & JANE G. GRAVELLE, CONG. RESEARCH SERV., R 40919, AN OVERVIEW OF THE NONPROFIT AND CHARITABLE SECTOR 49 (2009) (stating that the charitable contribution tax subsidy, based on one estimate, induces $0.50 of giving for each $1.00 of revenue loss).

298. *See* STAFF OF JOINT COMM. ON TAXATION, *supra* note 293, at 36 ("[T]he charitable contribution deduction reduces a taxpayer's after-tax cost of giving by relatively more, the higher his marginal tax rate."). Note, however, that the regressivity of the deduction could be offset by increasing tax rates for the income brackets of individuals who tend to donate more. COLINVAUX ET AL., *supra* note 293, at 10.

299. Mark P. Gergen, *The Case for a Charitable Contributions Deduction*, 74 VA. L. REV. 1393, 1405–06 (1988).

300. Brian Galle, *Charities in Politics: A Reappraisal*, 54 WM. & MARY L. REV. 1561, 1600–03 (2013).

301. *See* 26 U.S.C. § 2055(a) (2006).

302. Miranda Perry Fleischer, *Charitable Contributions in an Ideal Estate Tax*, 60 TAX L. REV. 263, 264–67 (2007).

serves to break up dynastic wealth, much like the estate tax generally.[303] That is true, but the two instruments get there in very different ways—most obviously, one deposits money into the Treasury, while the other does not.[304] Whether the remnants of dynasty should be allocated by the public or the dynasts seems like it should be a question of some interest.

All that I want to say about the institutional design of dynasty-breaking here is that nudges represent a third possibility. With nudges, the choice of how to allocate wealth are framed and influenced by the public's agents while the ultimate choices remain in the hands of donors. If the nudge replaces the estate-tax deduction, we must decide whether the incremental loss of private control and the deadweight losses to those who do not surrender it are worth the revenue gains. Alternately, the nudge (if effective enough) could replace the estate-tax system altogether. Then the question would be whether the incremental *gains* in private control are worth the lost dollars.

Conclusion

I have attempted here to offer the first extended consideration of the relative efficiency of nudges and other surprising regulation alongside the more traditional price-instrument and quantity-regulation alternatives. As with any initial academic forays into untrodden ground, no doubt I have made some missteps or overlooked some important areas for exploration. For now though, it looks as though present widespread skepticism of nudges and other surprising new policies may be misplaced. As a result, New York's soda law and many other forms of asymmetric regulation may merit closer consideration than others have so far been willing to offer. My arguments also warrant at least some reconsideration of older regulatory forms, such as prison or command-and-control regulation, whose inferiority to price instruments has become a central tenet of law and economics.

---

303. *E.g.*, James R. Repetti, *Democracy, Taxes, and Wealth*, 76 N.Y.U. L. REV. 825, 856 (2001); *see also* John G. Simon, *Charity and Dynasty Under the Federal Tax System*, 5 PROB. LAW. 1, 33 (1978) ("[I]t is doubtful that the charitable deduction results in a less egalitarian distributional pattern than what we would have in a world without deductions."). *But cf.* Fleischer, *supra* note 302, at 276–83 (agreeing with this point but cautioning that it does not explain all the legal features of the existing deduction).

304. *Cf.* Mark L. Ascher, *Curtailing Inherited Wealth*, 89 MICH. L. REV. 69, 135–36 (1990) (suggesting that an estate-tax deduction might be justified because it diversifies providers of public goods); Ray D. Madoff, *What Leona Helmsley Can Teach Us About the Charitable Deduction*, 85 CHI.-KENT L. REV. 957, 965–66 (2010) (pointing out that a charitable estate-tax deduction allows the very wealthy to effectively control use of government funds).

Book Reviews

Still Hazy After All These Years: The Data and Theory Behind "Mismatch"

MISMATCH: HOW AFFIRMATIVE ACTION HURTS STUDENTS IT'S INTENDED TO HELP, AND WHY UNIVERSITIES WON'T ADMIT IT. By Richard Sander & Stuart Taylor, Jr. New York, New York: Basic Books, 2012. 348 pages. $28.99.

Reviewed by William C. Kidder* & Angela Onwuachi-Willig**

A decade ago Professor Richard Sander authored a controversial *Stanford Law Review* article marshaling empirical evidence to advance the argument that affirmative action at U.S. law schools harmed African Americans' performance and resulted in a net decrease in the number of black lawyers.[1] Lawyer and journalist Stuart Taylor favorably wrote about Sander's findings and thesis at the time,[2] and years later the book *Mismatch* is the result of their collaboration,[3] one which also includes U.S. Supreme Court amicus briefs criticizing affirmative action[4] in the recent cases of

---

* Assistant Executive Vice Chancellor, University of California, Riverside. The views stated herein about UC are my research views and do not necessarily reflect the views of the UC/UCR administration. We wish to thank the following scholars for their comments and suggestions on earlier drafts of this Review: Mario Barnes, William Bowen, Erwin Chemerinksy, Matthew Chingos, Cheryl Harris, Kevin Johnson, Richard Lempert, Shana Levin, and Catherine Smith.

** Visiting Professor of Law, Yale Law School; Charles M. and Marion J. Kierscht Professor of Law, University of Iowa. I thank Dean Gail Agrawal and Charles M. and Marion J. Kierscht for their research support. I also give special thanks to my husband, Jacob Willig-Onwuachi, and our children, Elijah, Bethany, and Solomon, for their constant love and support.

1. Richard H. Sander, *A Systemic Analysis of Affirmative Action in American Law Schools*, 57 STAN. L. REV. 367, 373–75 (2004).

2. Stuart Taylor Jr., *Opening Argument—Do Racial Preferences Reduce the Number of Black Lawyers*, NAT'L J. (Dec. 4, 2004), http://www.nationaljournal.com/magazine/opening-argument-do-racial-preferences-reduce-the-number-of-black-lawyers—20041204.

3. RICHARD SANDER & STUART TAYLOR, JR., MISMATCH: HOW AFFIRMATIVE ACTION HURTS STUDENTS IT'S INTENDED TO HELP, AND WHY UNIVERSITIES WON'T ADMIT IT (2012).

4. Brief Amici Curiae for Richard Sander and Stuart Taylor, Jr. in Support of Neither Party at 2, Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411 (2013) (No. 11-345) [hereinafter Brief for Sander & Taylor, *Fisher* Case]; *see also* Brief Amicus Curiae for Richard Sander in Support of Petitioner at 13–15, Schuette v. Coal. to Defend Affirmative Action, No. 12-682 (U.S. July 1, 2013) [hereinafter Brief for Sander & Taylor, *Schuette* Case] (Stuart Taylor, Jr. signing as counsel of record).

Fisher v. University of Texas at Austin[5] and *Schuette v. Coalition to Defend Affirmative Action*.[6]

We were assigned a word limit for our Review, and we understand that Professor Sander was provided an opportunity to reply,[7] so we have narrowed our Review to a few areas in Parts II and III of *Mismatch*, which are where many important data claims are found. We hope that Sander's response squarely addresses these areas and not other affirmative action topics that are important in their own right (e.g., mismatch in law school[8] and STEM—i.e., science, technology, engineering, and mathematics—fields, consideration of socioeconomic background in college admissions), but not substantively discussed herein. In this Review, we have focused our attention on Sander and Taylor's claims that purported mismatches between students and institutions give rise to lower graduation rates and wages, that Proposition 209 (Prop. 209) has resulted in "warming effects" that have increased the attractiveness of the University of California system to underrepresented minorities, and that affirmative action causes its beneficiaries to feel stigmatized.

Our comprehensive review will show that the authors of *Mismatch* cherry-pick the data to support a series of unwarranted claims, for the social science data overall (and particularly the best peer-reviewed works) do not support Sander and Taylor's assertions that affirmative action causes lower overall college graduation rates or earnings for African Americans and Latinos. Additionally, the review shows that totality of social science evidence does not support Sander and Taylor's dubious claim that Prop. 209 ushered in a "warming effect" that reduced stigma and led to African Americans and Latinos becoming more likely to accept admission offers from the University of California.

We believe our Review and the themes we have chosen to address are timely and of policy relevance, as confirmed in the Supreme Court's October 2013 oral argument in *Schuette*, where Sander and Taylor's

---

5. 133 S. Ct. 2411 (2013).

6. 133 S. Ct. 1633 (2013), *granting cert. to* Coal. to Defend Affirmative Action v. Regents of Univ. of Mich., 701 F.3d 466 (6th Cir. 2012).

7. For context, we had not seen a draft of Sander's forthcoming reply at the time our substantive edits were completed and submitted to the *Texas Law Review*.

8. One of us analyzes recent law school mismatch research in William C. Kidder & Richard O. Lempert, *The Mismatch Myth in U.S. Higher Education: A Synthesis of the Empirical Evidence at the Law School and Undergraduate Levels*, *in* AFFIRMATIVE ACTION AND RACIAL EQUITY: CONSIDERING THE EVIDENCE IN FISHER TO FORGE THE PATH AHEAD (Uma M. Jayakumar & Liliana M. Garces eds., forthcoming 2014). Moreover, both of us separately coauthored earlier pieces responding to Sander's 2004 article on law school mismatch. *See infra* notes 36, 64; Kevin R. Johnson & Angela Onwuachi-Willig, *Cry Me a River: The Limits of "A Systemic Analysis of Affirmative Action in American Law Schools"*, 7 AFR.-AM. L. & POL'Y REP. 1, 4 (2005). In a couple of spots in this Review, we refer to Sander's law school mismatch claims to the extent there is an illuminating parallel on a technical point, but we do not delve into a substantive discussion of the law school mismatch literature.

mismatch hypothesis (and book) and the University of California's post-affirmative action college graduation rates were topics of discussion between the Justices and counsel.[9] We end by highlighting a revealing "mismatch" between Sander and Taylor's extensive focus on underrepresented minorities and affirmative action versus their inattention to the implications of mismatch for white students such as plaintiff Abigail Fisher. Under Sander and Taylor's worldview—highly flawed and contradictory as it is—Ms. Fisher's academic credentials indicate strong concerns about "academic mismatch" similar to many of the admitted students of color at the University of Texas at Austin for whom Sander and Taylor claim that mismatch is a serious problem.

I. Graduation Rates and Earnings: Lack of Depth, Lack of Breadth

In Chapter 6 ("The Breadth of Mismatch") and Chapter 9 ("Mismatch and the Swelling Ranks of Graduates"), and at several points throughout their book, Sander and Taylor argue that the purported mismatches caused by affirmative action bring about lower graduation rates and wages for African American and Latino beneficiaries of the policy. As we reveal in this Part, however, such claims are spurious, as the supporting evidence used by Sander and Taylor is either outdated or cherry-picked and dependent upon incomplete information and analyses. Even assuming that Sander and Taylor's evidence is reliable (and it is not), the overwhelming weight of social-science evidence bears against their contentions about the impact of mismatch on underrepresented students' graduation rates and wages.[10]

A. Introductory Points About Graduation Rates

Sander and Taylor make unsupported contentions that the findings in Bowen and Bok's *Shape of the River* are wrong,[11] and that "[s]tudies that

9. Transcript of Oral Argument at 11, 13, 16, 50–51, Schuette v. Coal. to Defend Affirmative Action, No. 12-682 (U.S. argued Oct. 15, 2013).

10. *See infra* Table 2.

11. SANDER & TAYLOR, *supra* note 3, at 106–07. Sander and Taylor claim that Bowen and Bok's findings about elite colleges' graduation rates reflect, "very plausibly, [those] students who were on average substantially less mismatched than were black students at less elite schools." *Id.* at 107. But for the subset of schools for which they had admission data for the 1989 cohort, Bowen and Bok reported an average black–white SAT score gap at the College & Beyond (C&B) schools of 209 points, compared to a nationwide gap of 200 points for the U.S. college-going population that year. *See* WILLIAM G. BOWEN & DEREK BOK, THE SHAPE OF THE RIVER 375 tbl. D.2.1 (1998). Inferring how these figures would likely translate on an apples-to-apples basis comparing within-institution gaps across the spectrum of colleges (i.e., standard deviation units) suggests that the black students at the C&B schools encounter a larger "credential gap"—at least as far as test scores—than is the case more generally in U.S. higher education. Sander and Taylor provide no data for their claims about Bowen and Bok and lesser mismatch at elite universities. Additionally, although their book promised a more technical analysis and critique of Bowen and Bok on the MISMATCH book website, SANDER & TAYLOR, *supra* note 3, at 107, Sander and

examine broader swaths of American higher education often find strong evidence that racial preferences produce lower college graduation rates."[12] These two claims by Sander and Taylor are simply not supported by contemporary social-science evidence, including the best peer-reviewed studies.[13] In fact, the two studies examining "broader swaths of American higher education"[14] that Sander and Taylor use to support their argument about lower graduation rates give the impression of being stuck in a time warp from ten or fifteen years ago.[15] Specifically, the first study that Sander and Taylor use—Loury and Garman—is outdated because it looked at students graduating high school in 1972.[16] Similarly, the second study that Sander and Taylor rely on—Light and Strayer—used a 1979 survey (students born in the late 1950s and early 1960s).[17] Undeniably, there have

---

Taylor failed to deliver, even sixteen months after their book went to press. Even the Thernstroms, who champion the mismatch hypothesis and whom Sander and Taylor reference in connection with Bowen and Bok, *id.* at 106, acknowledge that in theory the mismatch hypothesis would predict that "the dropout rate for blacks should be higher at Yale [and other elite C&B universities] than at a less selective school" because of the larger credential gap at elite C&B universities. Stephan Thernstrom & Abigail Thernstrom, *Reflections on* The Shape of the River, 46 UCLA L. REV. 1563, 1603 (1999) (book review). We believe that the relevant claims of both Sander and Taylor (i.e., narrower credential gap is the cause of less mismatch) as well as the Thernstroms (i.e., there is wider credential gap at elite schools, but grade inflation and resource differences between public and private universities obscure mismatch) are questionable, but the positions they stake out are somewhat incompatible. Sander and Taylor's critique of Bowen and Bok in their *Fisher* amicus brief is equally unavailing. *See* Brief for Sander & Taylor, *Fisher* Case, *supra* note 4, at 10 & n.26.

12. SANDER & TAYLOR, *supra* note 3, at 107; *see also id.* at 278.

13. *See infra* Table 2.

14. SANDER & TAYLOR, *supra* note 3, at 107.

15. For example, a year before *Mismatch*, Bastedo and Jaquette wrote:

> In the 1980s and 1990s, critics of affirmative action argued that racial minorities were damaged by affirmative action through lower graduation rates and that minority students would perform better—earn higher GPAs and be more likely to graduate—if they attended colleges that "fit" their academic profile (e.g., Cole & Barber, 2003; Light & Strayer, 2000; Thernstrom & Thernstrom, 1999; Trow, 1999). These claims were largely refuted by empirical data (Alon & Tienda, 2005; Bowen & Bok, 1998; Melguizo, 2008). The debate played out again over affirmative action at law schools, after a legal scholar conducted an analysis showing far lower bar pass rates for minority students graduating from elite law schools (Sander, 2004, 2005). These claims were also largely refuted through more sophisticated empirical analysis (Ho, 2005).

Michael N. Bastedo & Ozan Jaquette, *Running in Place: Low-Income Students and the Dynamics of Higher Education Stratification*, 33 EDUC. EVALUATION & POL'Y ANALYSIS 318, 319 (2011). The Sander and Taylor book relies upon many of these very same stale and/or refuted sources. For a critique of Cole and Barber's (and Sander and Taylor's) underlying claims about SAT scores and affirmative action, see William C. Kidder, *Misshaping the River: Proposition 209 and Lessons for the* Fisher *Case*, 39 J.C. & U.L. 53, 91–99 (2013).

16. Linda Datcher Loury & David Garman, *College Selectivity and Earnings*, 13 J. LAB. ECON. 289, 294 (1995).

17. Audrey Light & Wayne Strayer, *Determinants of College Completion: School Quality or Student Ability?*, 35 J. HUM. RESOURCES 299, 306 (2000).

been significant shifts in education and, more so, college admissions since 1972 and 1979. Also, Sander and Taylor's reliance on both the Loury and Garman and the Light and Strayer studies is faulty for other reasons. In particular, their reliance on the Loury and Garman data is flawed because those data were strongly swayed by historically black colleges and universities (HBCUs), where African American students had higher graduation rates than black students with similar credentials who attended predominantly white institutions.

It is imprudent for Sander and Taylor (via Loury and Garman) to rely upon the HBCUs as the workhorse behind their claim for "strong evidence" that mismatch lowers college graduation rates. After all, sound empirical scholarship properly identifies and rules out plausible alternative hypotheses,[18] and with respect to graduation rates of African Americans at HBCUs, there are rival hypotheses conspicuous in the literature that caution against making causal inferences regarding mismatch. For instance, researchers have found that the HBCUs often have a more supportive campus climate and have indicated that numerical diversity (both student and faculty) is one important contributing factor in boosting African Americans' grades and graduation rates at HBCUs[19] (we return to these themes later in our Review).

Indeed, other studies, such as Thomas Kane's, have utilized a more appropriate method for examining the impact of what Sander calls "mismatch" on graduation rates by separately accounting for HBCUs.[20] Specifically, Kane, using the nationally representative High School and Beyond data sample, concluded that "even if a students' characteristics are held constant, attendance at a more selective institution is associated with higher earnings and higher college completion rates for minority

18. *See* Leland Wilkinson & The Task Force on Statistical Inference, *Statistical Methods in Psychology Journals: Guidelines and Explanations*, 54 AM. PSYCHOLOGIST 594, 600 (1999) ("Inferring causality from nonrandomized designs is a risky enterprise. Researchers . . . have an extra obligation . . . to alert the reader to plausible rival hypotheses that might explain their results." (emphasis omitted)); MARK A. OLSON, STATISTICS FOR EXPERIMENTAL ECONOMISTS 20 (2012) (same).

19. *See* Walter R. Allen, *The Color of Success: African-American College Student Outcomes at Predominantly White and Historically Black Public Colleges and Universities*, 62 HARV. EDUC. REV. 26, 39 (1992) (finding, in an influential article, that on predominantly white campuses African Americans emphasized feelings of alienation and episodes of discrimination, whereas HBCUs had more favorable outcomes and the HBCUs tended to emphasize a greater sense of engagement, connection, and feeling encouraged in their educational pursuits); *see also* Walter R. Allen et al., *Historically Black Colleges and Universities: Honoring the Past, Engaging the Present, Touching the Future*, 76 J. NEGRO EDUC. 263 (2007).

20. *See* Thomas J. Kane, *Racial and Ethnic Preferences in College Admissions*, *in* THE BLACK-WHITE TEST SCORE GAP 431, 445–47 (Christopher Jencks & Meredith Phillips eds., 1998); *see also* MICHAEL K. BROWN ET AL., WHITEWASHING RACE: THE MYTH OF A COLOR-BLIND SOCIETY 116 (2003) (noting that Kane's study of 1982 high school seniors "flatly contradicts" the earlier Loury and Garman study of 1972 high school seniors).

students."[21] Furthermore, there is mixed, more recent evidence regarding whether African Americans at HBCUs have higher graduation rates than African American students at non-HBCU schools, all other things being equal,[22] and this research indicates that HBCUs may yield no benefit on earnings and may even result in a wage penalty controlling for other factors.[23]

Regarding the other national study cited by Sander and Taylor, the two *Mismatch* authors neglect to point out that, only two years after Light and Strayer's study based on 1979 survey data, Light and Strayer published a different study with the same data set that is more directly relevant. That study concludes that affirmative action "in college admissions boost minorities' chances of attending college and that retention programs directed at minority students subsequently enhance their chances of earning a degree."[24]

Thus far, the evidence proffered by Sander and Taylor is consistent with the title of our Review: *Still Hazy After All These Years*. After all, a decade ago in the *Stanford Law Review*, Sander relied on Loury and Garman and on Light and Strayer as his main supporting literature (that is national in scope and that is outside the STEM area) regarding undergraduate mismatch,[25] and today Sander and Taylor are unable to

---

21. Kane, *supra* note 20, at 432, 452.

22. *See, e.g.*, Ronald G. Ehrenberg et al., *Do Historically Black Colleges and Universities Enhance the College Attendance of African American Youths?*, *in* A NATION DIVIDED: DIVERSITY, INEQUALITY AND COMMUNITY IN AMERICAN SOCIETY 171, 171–88 (Patricia Moen et al. eds., 1999) (HBCUs increased graduation rates); Stella M. Flores & Toby J. Park, *Race, Ethnicity, and College Success: Examining the Continued Significance of the Minority-Serving Institution*, 42 EDUC. RESEARCHER 115, 125 (2013) (in study of Texas, net of other factors, finding HBCU graduation rates were essentially the same); Mikyong Minsun Kim & Clifton F. Conrad, *The Impact of Historically Black Colleges and Universities on the Academic Success of African-American Students*, 47 RES. HIGHER EDUC. 399, 417–19 (2006) (similar B.A. rates at traditionally white and HBCUs, but this was notable given the lower funding received by HBCUs).

23. Kane found HBCU status had "no statistically significant relationship with earnings." Kane, *supra* note 20, at 445. And Fryer and Greenhouse's more recent study reveals that students enrolling in HBCUs by the 1990s incurred wage penalties relative to similarly prepared students at traditionally white institutions. Roland G. Fryer, Jr. & Michael Greenstone, *The Changing Consequences of Attending Historically Black Colleges and Universities*, 2 AM. ECON. J.: APPLIED ECON. 116, 118 (2010). This finding is inconsistent with the Sander and Taylor mismatch account, and more so because the wage penalty at HBCUs accrued even though test score differences compared to traditionally white institutions slightly decreased between the 1970s and 1990s. *Id.* at 118, 141, 144.

24. Audrey Light & Wayne Strayer, *From* Bakke *to* Hopwood*: Does Race Affect College Attendance and Completion?*, 84 REV. ECON. & STAT. 34, 43 (2002).

25. *See* Sander, *supra* note 1, at 451. At that time Sander was aware of and attempted to distinguish Kane's criticism of Loury and Garman. *Id.* at 451 n.225. Putting aside the HBCU issue, one should note that Holzer and Neumark critique Loury and Garman on methodological grounds in a manner that more directly responds to Sander's earlier observation:

> Datcher Loury and Garman do not analyze differences in outcomes for blacks and whites over the entire range of college quality; they merely compare schools with

muster new or more robust analyses to bolster their claims that affirmative action harms African Americans' and Latinos' college graduation rates nationally.

As a first step before our more detailed review of the social-science literature on graduation rates and mismatch, we lay a foundation with comprehensive descriptive statistics responsive to Sander and Taylor's point about examining "broader swaths of American higher education"[26] in the context of college graduation rates. Here we also provide a framework for evaluating Sander and Taylor's elaboration of a "cascade effect" model in Chapter 2, which they claim "in key respects mirror[s] real-world data closely,"[27] and which they argue results in "perhaps the greatest harm done by the racial preferences used at the very elite schools[:] . . . their cascading effect on somewhat less elite schools."[28] Our data in Table 1 are not intended as causal proof refuting the mismatch hypothesis. Rather, our modest goal with Table 1 is to help readers have enough context to gain an intuitive appreciation about the extent to which the mismatch hypothesis—that underrepresented minority students will obtain *higher* graduation rates if they cascade to less selective universities—is empirically "swimming upstream" vis-à-vis the contemporary factual landscape at U.S. research

---

> average SAT scores above and below 1000. And, in their simulations where the net effects of college selectivity on overall graduation and earnings outcomes are determined, they only compare schools having median scores of 900 and 1000. But Kane, as well as Long (2004), have shown that the primary effects of affirmative action are in admission to the top quintile of schools, which are above these categories in quality. If this is true, the analysis in Darman-Loury and Garman seems to miss the most relevant part of the college quality spectrum with regards to affirmative action.

Harry J. Holzer & David Neumark, *Affirmative Action: What Do We Know?* 25 J. POL'Y ANALYSIS & MGMT. 463, 479 (2006).

Fast forward a decade, and the more fundamental point is that we are still dissecting a couple of old studies only because Sander and Taylor have failed to meet their burden of proof in support of their claim that there is "strong evidence that [affirmative action programs] produce lower college graduation rates." SANDER & TAYLOR, *supra* note 3, at 107.

26. SANDER & TAYLOR, *supra* note 3, at 107.

27. *Id.* at 21–25. To avoid confusion, note that Sander and Taylor deploy the term "cascade" to indicate the harmful effects of affirmative action, but traditionally affirmative action critics have deployed the cascade metaphor to describe the benefits of affirmative action bans. For a critique of the latter, see Michael N. Bastedo, Cascading Minority Students in Higher Education: Assessing the Impact of Statewide Admissions Standards (May 19, 2009) (unpublished manuscript), *available at* http://www-personal.umich.edu/~bastedo/workingpapers.html. The original formulation of the cascade metaphor (by Heriot, the Thernstroms, Trow and others), *see id.* at 1, is even more objectionable. With its serene imagery of gently flowing water or champagne bubbling downward among stacked crystal glasses, the original cascade metaphor obfuscates a core theme in our Review: ending affirmative action means closing doors of opportunity and success in American society.

28. SANDER & TAYLOR, *supra* note 3, at 107.

universities.[29] Table 1 covers the four most recent freshmen cohorts' six-year graduation rates (combining 2003–2004 through 2006–2007 cohorts) at all one hundred universities with the "Research University-Very High" (RU-VH) classification by the Carnegie Foundation and sufficient data using the federal/NCAA graduation rates.[30] The table displays the African American and Latino freshmen graduation rates, organized into quintiles (with 20 schools each); the most "selective"[31] quintile is on the left, and the least selective quintile is on the right. With four years of data at a hundred universities, Table 1 represents almost 90,000 African American and over 100,000 Latino freshmen.

---

29. To be sure, even the top one hundred research universities represent a modest share of the U.S. higher education picture overall. which includes community colleges, nonselective four-year public universities, modestly selective private colleges, and so on. At the same time, Sander and Taylor rely on Kane, *supra* note 20, for the proposition that only the top fifth or quarter of colleges use race-conscious affirmative action. SANDER & TAYLOR, *supra* note 3, at 309 n.21.

30. For information on the RU-VH Carnegie Institutions, see *Classification Description*, CARNEGIE FOUND. FOR ADVANCEMENT TEACHING, http://classifications.carnegiefoundation.org/lookup_listings/. The four cohort graduation rates are from the federal-graduation-rate NCAA "FGR Reports," which are available at *Education & Research*, NCAA, http://fs.ncaa.org/Docs/newmedia/public/rates/index.html. A few additional RU-VH universities are not displayed either because data were unavailable or the combined sample for African Americans was below 100: Brandeis, Caltech, Montana State, Hawaii, Rockefeller, Utah, and Yeshiva. For Latinos, there were actually ninety-nine institutions rather than one hundred, and those included or excluded were almost the same but not identical. (For space reasons, the table lists only the schools used to calculate the African American figures.) For Latinos the RU-VH universities not displayed due to unavailable data or samples that were too small are Alabama-Huntsville, Alabama-Birmingham, Caltech, Case Western Reserve, Mississippi State, North Dakota, Rockefeller, and Yeshiva. These small differences in "excluded schools" also account for the small differences in the comparison white graduation rates (e.g., within the second quintile the white rate is 85.8% for the African American row and it is 85.2% for the Latino row). Table 1 and the accompanying text report unweighted averages for each quintile.

31. Somewhat similar to Fischer and Massey, discussed *infra*, we use SAT median scores as a proxy for selectivity. The SAT median data are from The Education Trust's *College Results* data set, *Choose a College*, COLLEGE RESULTS ONLINE, http://www.collegeresults.org/. Using the SAT as a rough proxy for selectivity is not the same thing as claiming it is a proxy for "merit" or that it is the strongest predictor of individual student performance in college. That said, the simple correlation between median SAT/ACT scores and *U.S. News* rankings for the top 50 universities is 0.89 even if the correlations are much smaller for effective teaching and other more complex educational metrics, for example. Ernest T. Pascarella et al., *Institutional Selectivity and Good Practices in Undergraduate Education: How Strong is the Link?*, 77 J. HIGHER EDUC. 251, 252, 379–80 (2006).

Table 1: African American and Latino Six-Year Graduation Rates at One Hundred Top U.S. Research Universities (Carnegie "RU-VH"), in Quintiles by Selectivity, 2003–2004 to 2006–2007 Freshmen Cohorts

Top Quintile (# 1–20)	2nd Quintile (# 21–40)	3rd Quintile (# 41–60)	4th Quintile (# 61–80)	Bottom Quintile (# 81–100)
African American Graduation Rates (with Black–White Gap in Graduation Rates)				
88.9% (5.4 point gap)	**76.0%** (9.8 point gap)	**67.3%** (11.8 point gap)	**56.1%** (11.1 point gap)	**43.2%** (13.7 point gap)
Latino Graduation Rates (with Latino–White Gap in Graduation Rates)				
90.9% (3.4 point gap)	**80.4%** (4.8 point gap)	**71.2%** (7.9 point gap)	**60.4%** (7.9 point gap)	**49.0%** (6.6 point gap)
Institutions Included in Each Quintile				
Harvard, Yale, Princeton, MIT, Chicago, Dartmouth, Stanford, Wash. U, Columbia, Brown, Notre Dame, Penn, Duke, North-western, Rice, Vanderbilt, Tufts, Georgetown, Cornell, Carnegie Mellon	Emory, Johns Hopkins, USC, Rensselaer, UC Berkeley, NYU, Case Western, Virginia, Georgia Tech, Rochester, North Carolina-CH, Tulane, Michigan, Maryland, G-W, Miami, Illinois U-C, UCLA, UC San Diego, Florida	Boston U, Wisconsin-M, Ohio State, Pittsburg, Minnesota, UT Austin, UConn, VA Tech, Texas A&M, U of Washington, Stony Brook, UCSB, Tennessee, Penn State, Rutgers-NB, South Carolina, UC Irvine, Delaware	Florida State, NC State, Oklahoma, Central Florida, Michigan State, Iowa, Missouri, Purdue, UMass-Amherst, LSU, Alabama-H, UCSC, U at Buffalo, Iowa State, Nebraska, Kentucky, South Florida, U at Albany	Colorado State, Kansas, Cincinnati, Louisville, Oregon, Alabama-B, Arkansas, Illinois-Chi., N. Dakota State, Virginia Comm., Houston, GA State, Wash. State, Arizona, Oregon State, Arizona State, Miss State, UC Riverside, New Mexico, Wayne State

Three notable patterns emerge from Table 1 and the associated school-level data. First, African American and Latino graduation rates are highest by a considerable margin at the most selective universities. In the top twenty universities, 89% of African Americans and 91% of Latinos graduate, with the rates being even higher at the top of this tier (e.g., 97% and 96% at Harvard; and 94% at Yale). The fact that African American and Latino graduation rates increase with selectivity is important given Sander and Taylor's acknowledgement that in a "world totally purged of racial preferences, the proportion of blacks at the most elite universities . . . could fall dramatically," possibly to "as low as 1 percent" of the student body or

at least drop by half after accounting for other factors like athletics and class-based affirmative action.[32] For instance, at Duke University, African American and Latino graduation rates are nearly equal or equal to white graduation rates[33] (92%, 95%, 95%); so if African Americans plunged from ten percent of the Duke student body to two or three percent, for example, it is difficult to conceive of circumstances where ending affirmative action could result in a net gain in the likelihood of graduation for those underrepresented minority students who might no longer attend schools like Duke without any consideration of race. (Decreases in minority graduation rates also have negative implications for the University and U.S. society, discussed later in this Review.)

A second and related pattern emergent from Table 1 is more directly responsive to Sander and Taylor's assertion that the "greatest harm" of affirmative action is the cascading effect at somewhat elite colleges, which they claim "greatly aggravat[es] the overall scale of the mismatch problem."[34] In fact, Table 1 suggests the exact opposite: that graduation rates would be lower if African Americans attended less elite colleges at each level in the cascade. Specifically, Table 1 shows that the average black and Latino graduation rates in the top quintile exceed the white average graduation rate in the second quintile (86.8%), just as the African American and Latino average graduation rates in the third quintile meet or exceed the average white graduation rate in the fourth quintile (67.4%), and the black and Latino graduation rates in the fourth quintile equal or exceed the average white graduation rate for the bottom quintile (56.9%).[35] By implication, if in the absence of affirmative action many African American and Latino students cascaded to the next quintile (e.g., from schools like Boston University to schools like the University of Massachusetts at Amherst), the data in Table 1 suggest that the likelihood is quite small that these students of color could systematically be *more likely* to end up graduating even if one makes generous assumptions about a post-affirmative action landscape improving performance.[36]

---

32. SANDER & TAYLOR, *supra* note 3, at 278–79.

33. Sander and Taylor discuss Duke in another mismatch context. *Id.* at 25, 176–79. Our reference to Duke's exceptional graduation rates cabins the policy relevance of those parts of the book.

34. *Id.* at 107; *see also id.* at 23–24.

35. The figures in the text are white graduation rates in relation to African Americans. In relation to Latinos, the corresponding white graduation rates are 85.2%, 68.3%, and 55.5%, respectively. As noted earlier, these modest differences are because there were some small differences regarding which schools were "tossed" due to low sample sizes.

36. If past experience offers any lessons, in the area of law school admissions Sander's post-affirmative action models relied on a combination of heroic assumptions, *see* Sander, *supra* note 1, at 473 & tbl.8.2, and his portrait of the post-affirmative action landscape benefited from internally contradictory positions and methods, *see* Richard H. Sander, *A Reply to Critics*, 57

Third, another stubborn fact in these data is that many of the premiere public universities in Table 1 without race-conscious affirmative action still have troublingly large black–white gaps in graduation rates, including Texas A&M[37] (19 points), UC Berkeley (17 points), UC Davis (14 points), UCLA (12 points), the University of Florida (11 points), Washington State (10 points), UC Santa Barbara (10 points), and the University of Washington (9 points).[38] Thus, the real world data caution strongly against the notion that graduation rates will ascend to significantly higher levels without affirmative action, and even Sander and Taylor soberly acknowledge that "some of the ostensibly race-neutral proxies for racial preferences have brought in students who encounter even greater mismatch problems" than those under affirmative action.[39] One of the explanations a number of economists have emphasized, consistent with Table 1, is that race-conscious affirmative action can simply tend to be more efficient in yielding academically successful underrepresented students[40] (and Sander and Taylor's conclusions on this point are also intermingled with their non-peer reviewed allegations about evasion and cheating in admissions, a topic

---

STAN. L. REV 1963, 2000–02 (2005). *But see* David L. Chambers et al., *The Real Impact of Eliminating Affirmative Action in American Law Schools: An Empirical Critique of Richard Sander's Study*, 57 STAN. L. REV. 1855 (2005); Richard O. Lempert et al., *Affirmative Action in American Law Schools: A Critical Response to Richard Sander's "A Reply to Critics"* (Univ. of Mich. John M. Olin Ctr. for Law & Econ., Working Paper No. 06-001, 2006), *available at* http://www.law.umich.edu/centersandprograms/lawandeconomics/abstracts/2006/Documents/06-001lempert.pdf.

37. As Professor Garces notes, after *Grutter*, UT Austin "announced that it would reinstate the use of race in undergraduate admissions decisions, whereas Texas A&M University opted not to reinstate the consideration of race in admissions." Liliana M. Garces, *Necessary But Not Sufficient: The Impact of* Grutter v. Bollinger *on Student of Color Enrollment in Graduate and Professional Schools in Texas*, 83 J. HIGHER EDUC. 497, 505 (2012).

38. Just as for the "with affirmative action" universities, one should note that a portion of the racial gap in graduation rates is sometimes related to intercollegiate athletics, more so at schools with "big time" athletic programs in the NCAA Division I Football Bowl Subdivision, such as with the examples above. At a campus like UC Berkeley, likely one of the upper-bound cases in Table 1 because it garnered recent negative media attention over its "rock-bottom graduation rates" for student–athletes, the 17 point gap between white and black graduation rates (91% versus 74%) narrows to 13 points if all student–athletes receiving grant-in-aid scholarships are removed from the calculation (91% versus 78%). *See* Nanette Asimov & Ann Killion, *Why Do Many Cal Athletes Not Graduate?* SFGATE, (Nov. 22, 2013, 11:00 PM), http://www.sfgate.com/collegesports/article/Why-do-many-Cal-athletes-not-graduate-5004343.php.

39. SANDER & TAYLOR, *supra* note 3, at 280.

40. For various theoretical elaborations of these issues by economists, see Jimmy Chan & Erik Eyster, *Does Banning Affirmative Action Lower College Student Quality?*, 93 AM. ECON. REV. 858 (2003); Roland G. Fryer, Glenn C. Loury & Tolga Yuret, *An Economic Analysis of Color-Blind Affirmative Action*, 24 J.L. ECON. & ORG. 319 (2007); Debraj Ray & Rajiv Sethi, *A Remark on Color-Blind Affirmative Action*, 12 J. PUB. ECON. THEORY 399 (2010); Brent R. Hickman, Pre-College Human Capital Investment and Affirmative Action: A Structural Policy Analysis of US College Admissions (July 2013) (unpublished manuscript), *available at* http://home.uchicago.edu/~hickmanbr/uploads/AA_Empirical_paper.pdf.

for another day[41]). And the aforementioned problems Table 1 poses for the mismatch hypothesis carry even more force if, in a counterfactual world without affirmative action, some or many African Americans and Latinos were to drop two quintiles rather than one.

In contrast to the "facts on the ground" reflected in the descriptive statistics in Table 1, Sander and Taylor sketch out a simplified cascade effect admission model in which they claim that affirmative action is what produces large academic-index-score gaps throughout middle tier colleges and even at nonselective colleges.[42] Though Sander and Taylor's simplified cascade effect admission model is foundational for the remainder of their book and they claim it suggests that second and lower tier colleges suffer substantial mismatch as a byproduct of affirmative action at the most selective institutions, we, in fact, know little else about their cascade effect model except that it is not actually linked to outcome data on graduation rates (real or simulated).[43] And while Sander and Taylor claim that "a fuller description of this model[] and the underlying data can be found"[44] on their book's website, nothing has been available even now, as we near the publication date for our Review (which is sixteen months after *Mismatch* went to press).[45]

Even for those who might be generally predisposed to find the mismatch theory plausible, including some Supreme Court Justices, there are, as we have highlighted, a couple themes that should serve as early warning signs about the unreliability of the empirical claims undergirding Sander and Taylor's book: (1) limited, stale, and slanted citations to the research literature on college graduation rates; and (2) claims about a damaging cascade effect that are untethered to robust real world outcome

41. *See* SANDER & TAYLOR, *supra* note 3, at 279–80, 286. Sander's recent claims about UCLA admissions were harshly criticized in two separate and independent reviews by Professors Stern and Lempert that were commissioned by the UCLA Bunche Center for African American Studies. *See* Richard Lempert, Observations on Professor Sander's Analysis of the UCLA Holistic Admissions System (2013) (unpublished manuscript), *available at* http://www.newsroom.ucla.edu/portal/UCLA/document/Lempert_Review-Sander.pdf; David Stern, Are There Racial Dis-parities in UCLA Freshman Admissions? (Nov. 23, 2012) (unpublished manuscript), *available at* http://www.newsroom.ucla.edu/portal/UCLA/document/Stern_Review-Sander.pdf. Likewise a rigorous analysis of UC Berkeley freshmen admissions by Professor Hout found that race only played a trivial role in post-Prop. 209 admissions. MICHAEL HOUT, BERKELEY'S COMPREHENSIVE REVIEW METHOD FOR MAKING FRESHMAN ADMISSIONS DECISIONS: AN ASSESSMENT 2, 49 (2005), *available at* http://academic-senate.berkeley.edu/sites/default/files/committees/aepe/hout_report_0.pdf.

42. SANDER & TAYLOR, *supra* note 3, at 19, 23–24.

43. *Cf. id.* at 21 & 309 n.21, 22–24 (using academic-index rankings based on GPA and SAT distributions to explain the cascade effect).

44. *Id.* at 24.

45. *See Mismatch Supplements*, MISMATCH, http://www.mismatchthebook.com/?p=4 (showing no such description as of February 2014).

data and that our large-scale graduation-rate data (Table 1) suggest are built upon a foundation of sand.

B. Studies of Graduation Rates Nationally (and in Texas)

Now we turn to the social science on affirmative action graduation rates and labor market outcomes in more detail, which shows unequivocally that the cumulative weight of the educational research is in conflict with Sander and Taylor's key claims. A principle response that Sander and Taylor have to opposing social science on mismatch is to reiterate their arguments about "selection effect[s]" from Chapter 5 in claiming that selection on unobservables "will skew the analysis to favor students attending more elite schools . . . . Taking this bias into account, these studies as a group provide substantial—if not definitive—evidence that mismatch reduces minority graduation rates."[46] In other words, Sander and Taylor contend that one reason why the purported negative effects of mismatch on African Americans and Latinos may not be as prevalent for students at elite schools as they are for such students at lower tier schools is because students at elite schools may have unmeasurable positive qualities that enable them to succeed despite mismatch. Apart from the very fact that this argument by Sander and Taylor effectively concedes that there are important qualities that can enable student success in college despite what Sander and Taylor call a "mismatch" in credentials, we note that Sander and Taylor's mismatch argument is flawed in other ways. The phenomena of selection bias is true enough as far as it goes, but Sander and Taylor's degree of overreach—in claiming "substantial" or "definitive" evidence of mismatch reducing minority graduation rates[47] is unfortunate and appears (as we will show) to be based upon compound supposition rather than an empirically corroborated claim. In addition to the studies mentioned earlier (Bowen and Bok, Loury and Garman, and Light and Strayer), the only other studies included in Sander and Taylor's discussion at this point in the book are Dale and Krueger (discussed further below), and Alon and Tienda (plus data on the University of California, discussed further below).

While Sander and Taylor acknowledge that Alon and Tienda found little evidence of mismatch,[48] they fail to mention that Alon and Tienda used multiple empirical methods to overcome selection bias (i.e., propensity score analysis and Heckman methods[49]) yet still found "the mismatch

---

46. SANDER & TAYLOR, *supra* note 3, at 107–08.

47. *Id.* For background about selection bias and the idea that Sander's position on this issue has evolved and been inconsistent, see Richard O. Lempert et al., *supra* note 36, at 4.

48. SANDER & TAYLOR, *supra* note 3, at 107.

49. Sigal Alon & Marta Tienda, *Assessing the "Mismatch" Hypothesis: Differences in College Graduation Rates by Institutional Selectivity*, 78 SOC. EDUC. 294, 296 (2005).

hypothesis is empirically groundless for black and Hispanic" students.[50] In fact, a number of other peer-reviewed studies—employing a range of empirical methods—reach conclusions that mirror those found in Alon and Tienda's study.

Looking beyond the studies referenced by Sander and Taylor, the literature on college graduation rates and retention is too voluminous to summarize here *and* do justice to all the methodological nuances, but our "tree-top" level summary of a body of peer-reviewed studies shows that the weight of social science supports the proposition that African American and Latino students attain higher graduation rates in connection with affirmative action at selective U.S. colleges and universities. For instance, in *Crossing the Finish Line*, Bowen, Chingos, and McPherson analyzed twenty-one public flagship universities, plus the public university systems in four states, and found there is "no support whatsoever for [the mismatch] hypothesis" and that students "are generally well advised to enroll at one of the most challenging universities that will accept them."[51]

Similar to Alon and Tienda, Melguizo used techniques to control for selection bias and looked at National Education Longitudinal Study (NELS) data spanning highly selective to nonselective institutions. She found: "[M]inorities benefit from attending the most elite institutions. . . . [T]he selectivity of an institution attended has a positive and significant impact on the college completion rates of minorities."[52]

Furthermore, a study by Small and Winship adds support to our contention that college graduation rates are higher for Latinos and African Americans at selective institutions. Though Sander and Taylor made a to-do in their book about the inaccessibility of the College and Beyond (C&B) data set[53] utilized by Bowen and Bok for their seminal work, *The Shape of the River*, in addition to the aforementioned Alon and Tienda study that used C&B data, Small and Winship also relied upon the C&B data in concluding the following: "[S]electivity increases the probability of

50. *Id.* at 309.

51. WILLIAM G. BOWEN ET AL., CROSSING THE FINISH LINE: COMPLETING COLLEGE AT AMERICA'S PUBLIC UNIVERSITIES 12–16, 227–28 (2009).

52. Tatiana Melguizo, *Quality Matters: Assessing the Impact of Attending More Selective Institutions on College Completion Rates of Minorities*, 49 RES. HIGHER EDUC. 214, 216–17, 223, 232 (2008); *see also* Tatiana Melguizo, *Are Students of Color More Likely to Graduate from College if They Attend More Selective Institutions?: Evidence from a Cohort of Recipients and Nonrecipients of the Gates Millennium Scholarship Program*, 32 EDUC. EVALUATION & POL'Y ANALYSIS 230, 242–44 (2010) (concluding that "highly motivated low-income students of color in good academic standing can thrive at the most and highly selective institutions and attain a bachelor's degree in a timely manner").

53. SANDER & TAYLOR, *supra* note 3, at 106, 236. A point relating to several studies cited herein is that although the elite C&B institutions were primarily private, because those public universities in the sample had larger student bodies, over 30% of the students in the 1976 and 1989 C&B cohorts were from public universities. *See* BOWEN & BOK, *supra* note 11, at xxxvii.

graduation. . . . Second, it is noteworthy that it helps blacks more than it does whites . . . . [T]he strong effects of selectivity demonstrate a clear benefit of Affirmative Action in elite institutions."[54] Small and Winship's study reached these findings after controlling for a number of institutional factors, including institutional wealth, grading difficulty/leniency, and expenditures on student resources.[55]

Convergent validity comes from a study by Fischer and Massey, who reapproached the C&B schools (and added the University of California at Berkeley) in creating a newer database with the National Longitudinal Survey of Freshmen.[56] Fischer and Massey concluded, "Our estimates provided no evidence whatsoever for the mismatch hypothesis. . . . If anything minority students who benefited from affirmative action earned higher grades and left school at lower rates than others . . . ."[57] Fischer and Massey's study also directly responded to a core tenet of Sander and Taylor's theory (and Loury and Garman's less effective test of that theory[58]) insofar as it looked at the greater distance ("mismatch") between minority students' SAT scores and the median SAT score in the same institution, with findings that were the opposite of what mismatch would predict. Fisher and Massey noted:

> Indeed, the degree of an individual's likely benefit from affirmative action is *negatively* related to the likelihood of leaving school, and the effect is highly significant. For each 10 points increase in the gap between the individual's SAT score and the institutional average, there was an 8.5% *decrease* in the likelihood of leaving college.[59]

And among nearly 40,000 freshmen attending a broad swath of public and private four-year institutions in Illinois, Gong similarly found a negative relationship between dropping out after the freshmen year and the "mismatch" distance between a student's ACT score and the college median ACT.[60]

Several of the studies in this genre, including Bowen and Bok and two studies by Espenshade and colleagues that rely on a subset of C&B

---

54. Mario L. Small & Christopher Winship, *Black Students' Graduation from Elite Colleges: Institutional Characteristics and Between-Institution Differences*, 36 SOC. SCI. RES. 1257, 1258, 1272 (2007).

55. *See id.* at 1267 tbl.3.

56. Mary J. Fischer & Douglas S. Massey, *The Effects of Affirmative Action in Higher Education*, 36 SOC. SCI. RES. 531, 534 (2007).

57. *Id.* at 544.

58. *See* Holzer & Neumark, *supra* note 25.

59. Fischer & Massey, *supra* note 56, at 541.

60. YUQIN GONG, ILL. EDUC. RESEARCH COUNCIL, THE DIVERGENCE OF THE RIVER: EXAMINING THE EFFECT OF ACADEMIC "MISMATCH" ON COLLEGE STUDENT'S EARLY ATTRITION (2006), *available at* http://www.siue.edu/ierc/presentations/pdf/Mismatch2006Symp.pdf (summarizing the findings from Gong's unpublished Ph.D. dissertation).

institutions, acknowledge that affirmative action has some tradeoff vis-à-vis students' college grade-point averages (GPAs), but nonetheless conclude that the net benefits as far as college graduation rates and later graduate and professional school attainment make affirmative action worthwhile from a social policy standpoint.[61]

Turning to studies about Texas, the previous affirmative action ban after *Hopwood v. Texas*[62] provided opportunities for analyzing "natural experiments" around what happened after the case's ruling took effect and ended affirmative action.[63] One such study by Cortes found that graduation rates for minorities actually decreased after *Hopwood*, rather than increased.[64] In this study, Cortes focused on those outside the top strata—the second and lower deciles in high school rank—and used the top decile students as a control group because their admission prospects were the same pre-*Hopwood* and under the Top Ten Percent Plan.[65] Cortes focused on outcomes at six Texas publics that included the two flagships (the University of Texas at Austin and Texas A&M at College Station), but also Texas Tech, Texas A&M at Kingsville, the University of Texas at San Antonio, and the University of Texas Pan American.[66] Thus, Cortes addressed a core criticism of Sander and Taylor by looking beyond a narrow set of elite institutions; yet, she found that the gap between minority and nonminority graduation rates among the students in her study grew from twenty-five percentage points in 1990–1996 (42% versus 67%) to

---

61. *See* Douglas S. Massey & Margarita Mooney, *The Effects of America's Three Affirmative Action Programs on Academic Performance*, 54 SOC. PROBS. 99, 114 (2007) (noting negative association with college grades but finding that "[c]ontrary to expectations derived from the critics, the stronger an institution's apparent commitment to affirmative action, the lower the likelihood minority students would leave school"); *see also* THOMAS J. ESPENSHADE & ALEXANDRIA WALTON RADFORD, NO LONGER SEPARATE, NOT YET EQUAL: RACE AND CLASS IN ELITE COLLEGE ADMISSION AND CAMPUS LIFE 233–36, 245 (2009) (finding that class rank distributions are "sharply differentiated by race," with URM students "disproportionately concentrated toward the bottom of their graduating class," but nevertheless stating that their results are "completely consistent with those found in the C&B data" and that they "would have to conclude that there is no support in [their] data for the mismatch hypothesis"); Joanne W. Golann et al., *Does the "Mismatch Hypothesis" Apply to Hispanic Students at Selective Colleges?*, *in* THE EDUCATION OF THE HISPANIC POPULATION: SELECTED ESSAYS at 209, 222–23 (Billie Gastic & Richard R. Verdugo eds., 2013). *Compare* BOWEN & BOK, *supra* note 11, at 72–28 (class rank), *with id.* at 160–72 (leadership), *and id.* app. D tbl.D.4.1 (percentage in the three tiers of C&B schools who went on to obtain M.D., J.D. Ph.D. and M.B.A. degrees).

62. 78 F.3d 932 (5th Cir. 1996).

63. *See id.* at 962 (concluding that the law school may not use race as a factor in admissions). Regarding the point about state affirmative action bans and natural experiments, see Susan K. Brown & Charles Hirschman, *The End of Affirmative Action in Washington State and Its Impact on the Transition from High School to College*, 79 SOC. EDUC. 106, 106 (2006).

64. Kalena E. Cortes, *Do Bans on Affirmative Action Hurt Minority Students? Evidence from the Texas Top 10% Plan*, 29 ECON. EDUC. REV. 1110, 1111 (2010).

65. *Id.* at 1111–13.

66. *Id.* at 1117 & n.17.

thirty points in 1998–1999 (39% versus 69%) after *Hopwood*, when affirmative action in Texas ended.[67] By contrast, Sander and Taylor make the hollow claims in *Mismatch* that "preferences on the scale used by [The University of Texas at Austin] are almost certain to backfire on the students they purport to help."[68]

C. *Graduation Rates at the University of California*

Turning to graduation rates in California, Sander and Taylor devote Chapter 9 to the University of California's experience after Prop. 209 ended affirmative action,[69] claiming:

> Perhaps the most important mismatch question we can consider from the UC move to putative race-neutrality is this: Did even a modest reduction in the net preferences received by blacks and Hispanics improve their graduation rates?
>
> The simple answer is an emphatic yes. Minority graduation rates rose rapidly in the years after Prop 209, and on-time (four-year) graduation rates rose even faster. . . . The increase in black six-year graduation was less dramatic (63 percent before and 71 percent after Prop 209) but still substantial.
>
> . . . Six-year graduation rates [for Hispanics] rose from 69 to 74 percent.[70]

These claims about "substantial" and even "stunningly improved rates"[71] of graduation warrant careful examination, particularly because the Michigan attorney general very recently cited Sander's related graduation-rate research in his merit brief in the *Schuette* Supreme Court case.[72] Indeed, during the October 2013 oral argument in *Schuette*, Michigan's solicitor general asserted that the University of California's under-represented minority graduation rates are "20 to 25 percent higher than [they were] before California's Prop. 209," suggesting this was caused by

---

67. *Id.* at 1120.

68. SANDER & TAYLOR, *supra* note 3, at 289. Sander and Taylor provide scant supporting evidence for this claim, *id.* at 288–89, and the same goes for their amicus brief in *Fisher*, where the claims are fleshed out in somewhat more detail, *see* Brief for Sander & Taylor, *Fisher* Case, *supra* note 4, at 5–10.

69. Prop. 209—passed by a majority of voters in November of 1996—amended the California Constitution to provide: "The State shall not discriminate against, or grant preferential treatment to, any individual or group on the basis of race, sex, color, ethnicity, or national origin in the operation of public employment, public education, or public contracting." CAL. CONST. art. I, § 31(a); Sherman J. Clark, Commentary, *A Populist Critique of Direct Democracy*, 112 HARV. L. REV. 434, 434 n.1 (1998).

70. SANDER & TAYLOR, *supra* note 3, at 146.

71. *Id.* at 143.

72. Brief for Petitioner at 31 & nn.5–6, 32, 35, Schuette v. Coal. to Defend Affirmative Action, No. 12-682 (U.S. July 1, 2013).

the affirmative action ban.[73] Chief Justice Roberts referenced Sander and Taylor's work on mismatch during the same oral argument.[74]

But more in-depth examination reveals that Sander and Taylor committed a serious flaw when they reported 63% and 69% as the pre-Prop. 209 African American and Latino freshmen 1992–1997 six-year graduation rates, respectively, and later used averages from 1998–2003 for those groups' post-Prop. 209 figures.[75] Although reporting averages for adjacent years is reasonable in other circumstances, here it was masking a trend in the data that actually cuts against Sander and Taylor's principal mismatch thesis. Using Sander and Taylor's same data, Figure 1 below shows that, for African Americans, the six-year graduation rate in the University of California (UC) system improved from 60% of entering freshmen in 1992 to 69% in 1997.[76] Thus, African Americans made a substantial, nine-point improvement in their graduation rate in the half-dozen years *before* Prop. 209, making the subsequent rise in the years after Prop. 209 (to 71% of entering freshmen by 1998, and 73% by 2003[77]) look much less impressive, if not disappointing. Likewise for Latinos, the graduation rate rose pre-Prop. 209 from 67% in 1992 to 72% in 1997.[78] In the years after Prop. 209 took effect, the Latino graduation rate fluttered between 72% and 75% (73.6% average), and without an upward trajectory.[79]

73. Transcript of Oral Argument, *supra* note 9, at 16. Michigan's solicitor general also relied on Sander to advance problematic claims about UC enrollment levels after Proposition 209, which is beyond the scope of this Review. For a critique of these claims citing several *Schuette* amici briefs, see William Kidder, *Michigan's Mangled Empirical Claims in the Schuette Affirmative Action Case*, AM. CONST. SOC'Y BLOG (Oct. 23, 2013), http://www.acslaw.org/acsblog/michigan%E2%80%99s-mangled-empirical-claims-in-the-schuette-affirmative-action-case.

74. Transcript of Oral Argument, *supra* note 9, at 50–51.

75. *See* SANDER & TAYLOR, *supra* note 3, at 146–47.

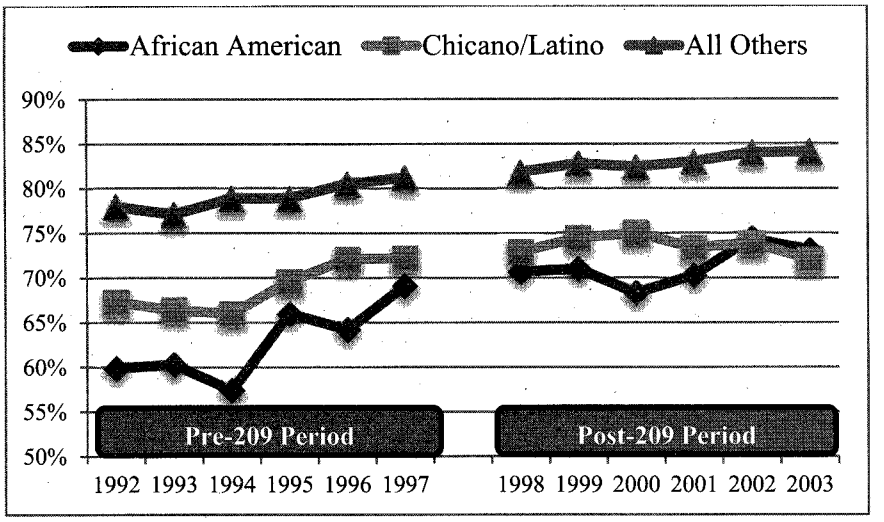
76. Additional details, with figures identical to those in *Mismatch*, are provided in a short paper by Sander from around 2010. Richard H. Sander, An Analysis of the Effects of Proposition 209 upon the University of California 6 (unpublished manuscript), *available at* http://www.seaphe.org/pdf/analysisoftheeffectsofproposition209.pdf. Michigan's solicitor general cites this same unpublished paper by Sander as the source for claims in his merit brief in *Schuette*. Brief for Petitioner, *supra* note 72, at 35.

77. Sander, *supra* note 76.

78. *Id.*

79. *Id.*

Figure 1: UC's Six-Year Graduation Rates, 1992–2003 Freshmen[80]



Sander and Taylor are even more celebratory about the post-Prop. 209 changes in four-year graduation rates.[81] UC's four-year graduation rates are not displayed in a figure because of an "apples-to-oranges" problem that is ignored by Sander and Taylor: The source data for UC in 1992–1994 do not include the fourth-year summer in the graduation rate, unlike the 1995–1997 data and the post-Prop. 209 (1998–2005) data.[82] This is not nearly as big of a deal for six-year rates (because a sixth-year summer adds a miniscule bump to graduation rates), but for the 1992–1994 four-year rates, (which constitute half of what makes up Sander and Taylor's pre-Prop. 209 average), the absence of fourth-year summer data deflates the graduation rates by about five percentage points.[83] Taking that into account as well as

80. The "all other" category is for all domestic, but not international, students.

81. *See* SANDER & TAYLOR, *supra* note 3, at 146 (saying that on-time (four-year) graduation rates rose even faster than six-year graduation rates).

82. Sander's original paper and the *Mismatch* book relied upon UC data from the UC Office of the President "Statfinder" website, *id.* at 323 n.143, which is no longer available due to budget constraints. But the library of tables produced in Statfinder would presumably have included this proviso. Likewise, the latest UC Accountability Report includes such a caution regarding data on pre-1995 graduation rates. *See* UNIV. OF CAL., ANNUAL ACCOUNTABILITY REPORT 42 n.1 (2013), *available at* http://accountability.universityofcalifornia.edu/documents/accountabilityreport13.pdf (qualifying its presentation of four-year graduation rates by stating that the rates after 1995 include fourth-year summers, but that data before 1995 do not).

83. This is a ballpark estimate. The 1992–1994 entering freshmen cohorts' second year persistence rate is the same as it was for the 1995 cohort (82.1% average versus 82.0%), but the 1992–1994 four-year graduation rate (35.3% average) is over five points lower than the rate in 1995 (40.7%). *See* Memorandum from the Univ. of Cal. Office of the President, University of California Undergraduate Student Persistence and Graduation Rates, Entering Cohorts: Fall 1992-Fall 2011 (Feb. 7, 2013) (on file with authors).

the fact that graduation rates were rising for all UC students in the pre- and post-Prop. 209 period[84] because of rising selectivity, more relevant than the averages Sander and Taylor report is the fact that the gaps in four-year graduation rates in 1997 were within two or three points of the first post-Prop. 209 (1998–2005) averages reported by Sander and Taylor.[85] The chart in Sander's working paper shows that the increase in four-year graduation rates for all other domestic students (i.e., non-underrepresented minorities) increased with a similar slope as for underrepresented minority (URM) students.[86] We are not the first, nor likely the last, to emphasize that paying proper attention to trend lines and other contextual factors is important when analyzing UC graduation rates and drawing inferences about Prop. 209.[87]

Sander and Taylor also rely on a recent working paper about Prop. 209 and graduation rates by Duke economists Arcidiacono et al.,[88] arguing that "[t]here is simply no other study that has so effectively handled the difficult problem of 'selection effects'" and that, "if anything, [the paper] underestimate[s] Prop. 209's true effects."[89] What Sander and Taylor do not emphasize, however, is that "mismatch" was third on the list in Arcidiacono et al.'s findings about what factors were most influential in explaining their results: (1) they attributed the largest share of the increase in minority graduation rates, 35%–50%, to increased selectivity (see our

84. Sander, *supra* note 76, at 4–6.

85. Again, post-Prop. 209, four-year graduation rates rose significantly between 1998 and 2005—mostly for selectivity reasons—but given the trend line associated with the period between 1992 and 1997, this rise certainly would have been the case as well in a counterfactual world where Prop. 209 never occurred. *See id.*

86. *Id.* at 4. Additionally suggestive of confounders, the combined (for all groups) four-year graduation rates at *non-UC* elite public universities likewise rose from 41% for the 1998 freshman class to 52% for the 2005 freshman class (the period corresponding to the initial years after Prop. 209). *See Freshman Graduation Rates*, U. CAL. ACCOUNTABILITY REP. 2013, http://accountability.universityofcalifornia.edu/index/4.1.

87. For example, Chang and Rose analyze 1994–2003 UC and UCB/UCLA graduation rates and conclude, "Proposition 209 added little to the momentum URM students already had going back at least to 1995. About two-thirds of the graduation-rate improvement occurred before students were subject to the Proposition 209 admissions requirements." Tongshan Chang & Heather Rose, *A Portrait of Underrepresented Minorities at the University of California, 1994–2008*, *in* EQUAL OPPORTUNITY IN HIGHER EDUCATION: THE PAST AND FUTURE OF CALIFORNIA'S PROPOSITION 209, at 83, 98 fig.5.5, 99 (Eric Grodsky & Michal Kurlaender eds., 2010); *see also* Brief for the President and Chancellors of the University of California as Amici Curiae in Support of Respondents at 31–34, Schuette v. Coal. to Defend Affirmative Action, No. 12-682 (U.S. Aug. 30, 2013) (referring to Sander's unpublished paper and reaching the same result that cuts against Sander's conclusion, namely, that Sander masks a trend in the data); Kidder, *supra* note 15, at 105–08 (same).

88. Peter Arcidiacono et al., *Affirmative Action and University Fit: Evidence from Proposition 209* (Inst. for Study of Labor Discussion Paper Series, Paper No. 7000, 2012), *available at* http://ftp.iza.org/dp7000.pdf. This Review uses the benchmark of five-year graduation rates. *Id.* at 6 n.7.

89. SANDER & TAYLOR, *supra* note 3, at 147–48.

earlier discussion); (2) 30%–45% was attributable to "university response," a residual category including various efforts to promote student success (see our conclusion of this Review for related observations); and (3) the lessening of "mismatch" accounted for 20% of the change in graduation rates.[90] Sander and Taylor's "all eggs in one basket" reliance on the Arcidiacono et al. study is unpersuasive in light of the literature reviewed herein (including the wages studies noted below), and the Arcidiacono et al. study has also been recently criticized by Chingos for ignoring the trend in UC graduation rates.[91] Moreover, Arcidiacono's recent paper with Koedel regarding the Missouri higher education system is in tension with the mismatch hypothesis, as they estimate that African American degree attainment would improve if more African Americans were upwardly shifted to more selective public colleges in Missouri.[92]

Additionally, like the study by Cortes of Texas, a recent study about several UC campuses by Kurlaender and Grodsky took advantage of a natural experiment to address selection effects by looking at a unique set of students who were initially denied straight admission as freshmen to UC because the 2003–2004 budget crisis caused funding cuts, but were then later admitted at Berkeley, UCLA, and UC San Diego late in the summer when the budget modestly improved.[93] Kurlaender and Grodsky utilized additional controls for selection bias (patterned after the "self-revelation" Dale and Krueger method, discussed below) by focusing on those students who had applied to the same UC campuses. They looked at these students' performance over the next four years and found that mismatch "has no reliable or substantively notable bearing on grades, rates of credit accumulation, or persistence."[94] Other recent articles on degree attainment that include, but are not limited to, California and Prop. 209 have found that

90. *See* Arcidiacono et al., *supra* note 88, at 3–4, 29 tbl.8.

91. *See* Matthew M. Chingos, *Are Minority Students Harmed by Affirmative Action?, Brown Center Chalkboard*, BROOKINGS (Mar. 7, 2013, 11:00 AM), http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/03/07-supreme-court-chingos ("A key problem with the before-and-after method is that it does not take into account pre-existing trends in student outcomes.").

92. Peter Arcidiacono & Cory Koedel, *Race and College Success: Evidence from Missouri*, AM. ECON. J.: APPLIED ECON. (forthcoming) (manuscript at 3–4), *available at* http://public.econ.duke.edu/~psarcidi/ak_college.pdf ("[W]e show that differences in enrollment patterns between African Americans and whites across groups of less prestigious colleges are the primary drivers behind the counterfactual sorting gains. In particular, it is moving African Americans out of urban schools and the very bottom schools that result in the graduation gains.").

93. Michal Kurlaender & Eric Grodsky, *Mismatch and the Paternalistic Justification for Selective College Admissions*, 86 SOC. EDUC. 294, 297–98 (2013).

94. *Id.* at 305–07. Initially, budget cuts caused the UC System to scale back admissions to a group of eligible, but less academically competitive, students, who were made the promise of later admission after two years at a community college. *Id.* at 297–98. When funding was partly restored in the summer of 2004, this group of "guaranteed transfer offer" students at UC Berkeley, UCLA, and UC San Diego were offered automatic admission. *Id.* Note that this study had retention data through four years, which is similar to, but not the same thing as, graduation rates.

affirmative action bans have modest negative effects (or modest negative effects nationwide) on URMs' graduation prospects, particularly at the most selective universities.[95]

D. After Graduation: Earnings in the Labor Market

Sander and Taylor's claim about lower post-graduation wages for so-called mismatched minority students also is not supported by the evidence. For example, the part of *Mismatch* that centers on Sander and Taylor's related discussion of earnings—where they argue that the "hard evidence" of earning advantages for attending elite schools is "surprisingly weak"[96]—also has a certain time warp quality.[97] After again deriding Bowen and Bok, Sander and Taylor then discuss a "clever analysis" in Dale and Krueger's 2002 matching study.[98] Despite the fact that Sander had earlier bent the Dale and Krueger study to fit his own critique of affirmative action,[99] the "proof in the pudding" is found in a very recent follow-up article by Dale and Krueger that looked at C&B schools (plus some others) and federal administrative and tax data on earnings.[100] For the 1989 cohort at largely C&B schools (overlapping a lot with the cohort studied by Bowen

---

95. Ben Backes, *Do Affirmative Action Bans Lower Minority College Enrollment and Attainment? Evidence from Statewide Bans*, 47 J. HUM. RESOURCES 435, 437 (2012) (concluding, based upon 1990–2009 Integrated Postsecondary Education Data System data, "All in all, although the effect sizes were modest, estimates show that there were fewer black and Hispanic students graduating from four-year, public universities following the bans, and those who did graduate tended to do so from less prestigious universities"). Peter Hinrichs has also stated:

> I find that overall graduation rates do not change very much when affirmative action is banned. I find that graduation rates for underrepresented minorities at selective universities rise, although I acknowledge that this may be due to the changing composition of students who enroll at such universities. Moreover, the effects are small compared to the number displaced from selective universities due to affirmative action bans. I find that the negative effect on enrollment outweighs the positive effect on graduation from these universities, so that affirmative action bans lead to fewer underrepresented minorities becoming graduates of selective institutions.

Peter Hinrichs, Affirmative Action Bans and College Graduation Rates 5 (Nov. 21, 2012) (unpublished manuscript), *available at* http://www9.georgetown.edu/faculty/plh24/affactionbans-collegegradrates_112112.pdf.

96. SANDER & TAYLOR, *supra* note 3, at 108.

97. *See supra* notes 10–13 and accompanying text.

98. SANDER & TAYLOR, *supra* note 3, at 108, 319 n.108 (citing Stacy Berg Dale & Alan B. Krueger, *Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables*, 117 Q.J. ECON. 1491 (2002)). Likewise, Sander previously called the Dale–Krueger method "the most reliable way of measuring mismatch effects." Sander, *supra* note 36, at 2016.

99. *See* David L. Chambers et al., *supra* note 36, at 1882 & n.101 (critiquing Sander's empirical methodology and asserting that—contrary to the conclusions drawn by Sander—the Dale and Krueger study "has a more nuanced message when read in context").

100. Stacy Dale & Alan Krueger, *Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data*, J. HUM. RESOURCES (forthcoming) (manuscript at 4–5), *available at* http://www.aeaweb.org/aea/2013conference/program/retrieve.php?pdfid=220.

and Bok), Dale and Krueger found that, among matched students, wage premiums were not significant *except that* "the effect of attending a school with a higher average SAT score is positive for black and Hispanic students, even in the selection-adjusted model."[101]

Other recent economic research finds that attending selective colleges is associated with higher economic returns for blacks and Latinos,[102] and earlier studies utilizing the National Longitudinal Survey of Youth (NLSY) reach the same conclusion.[103] Recently, Andrews, Li, and Lovenheim looked at males in Texas who graduated from high school in 1996–2002 to determine the extent to which attending the University of Texas at Austin (UT Austin) and Texas A&M had later effects on earnings, other things being equal, compared to those attending less selective public universities.[104] They found heterogeneous results, with small returns among UT Austin's African Americans and Latinos in the middle of the income distribution, but "quite large" returns elsewhere in the distribution, and for African Americans and Latinos at Texas A&M, the earnings returns were "universally large."[105] Another study by Hoekstra addressed selection bias by comparing students who were *barely* above or below the admission cutoff at one of the Texas flagship universities, and while this study analyzed only white men, the author found a 20% wage premium of attending the "most selective" Texas flagship university by the time the students were in their late twenties and early thirties.[106]

---

101. *Id.* (manuscript at 28).

102. Mark C. Long, *Changes in the Returns to Education and College Quality*, 29 ECON. EDUC. REV. 338, 346 (2010) (concluding that "[f]or annual earnings, the increases in returns to years of education were greatest for men, Blacks, and Hispanics").

103. *See* Kermit Daniel et al., *Racial Differences in the Effects of College Quality and Student Body Diversity on Wages*, *in* DIVERSITY CHALLENGED: EVIDENCE ON THE IMPACT OF AFFIRMATIVE ACTION 221, 222, 229 (Gary Orfield ed., 2001) (finding "strong evidence of a much larger effect of college quality on the later wages of blacks than of nonblacks"); James Monks, *The Returns to Individual and College Characteristics: Evidence from the National Longitudinal Survey of Youth*, 19 ECON. EDUC. REV. 279, 286 (2000) ("In particular, non-white [black and Latino] graduates of highly or most competitive institutions earn a larger premium than whites.").

104. Rodney J. Andrews, Jing Li & Michael F. Lovenheim, *Quantile Treatment Effects of College Quality on Earnings: Evidence from Administrative Data in Texas* 6 (Nat'l Bureau of Econ. Research, Working Paper No. 18068, 2012), *available at* http://www.nber.org/papers/w18068.

105. *Id.* at 4, 26–28. This review also compared graduating from a Texas community college, instead of a non-flagship, public, four-year university, and for black and Latino students there were negative returns for graduating from a community college below the 91st percentiles and the 84th percentile, respectively. *Id.* at 28–29.

106. Mark Hoekstra, *The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach*, 91 REV. ECON. & STAT. 717, 724 (2009) ("The results indicate that attending the flagship state university increases the earnings of 28- to 33-year-old white men by approximately 20%, which suggests significant economic returns to college quality, at least in the context of the most selective public state university.").

All of these studies on graduation rates and wages, nationally and in Texas and California, are reflected in the summary table below.[107] The weight of the overall evidence substantially calls into question the claims made by Sander and Taylor. In Table 2, under the Sander and Taylor column, a superscript question mark follows the name of two studies (Dale and Krueger, 2002; Light and Strayer, 2000) where we believe Sander and Taylor's claims are at variance with the conclusions the authors of those studies reach in related works and refers readers to those related studies, indicated by a superscript asterisk in the right-hand column. We also mark with a single asterisk several studies that shed light on Loury and Garman's 1995 findings.

If one is to read between the lines, Sander and Taylor may be arguing something along the lines of, "We are unsatisfied with the vast majority of scholarly studies; we believe that *if* the research were to reflect controls for selection bias that we deem satisfactory, then we expect the resulting findings *would* conform to our belief that mismatch significantly reduces graduation rates and wages of affirmative action beneficiaries."[108] If that is essentially their position—rather than simply failing to provide sufficient research support for their claims—then the *Mismatch* book is covertly bottomed on dogma rather than data. Either way, Sander and Taylor's claims are not supported by the weight of social-science evidence.

107. *See infra* Table 2.

108. *Cf.* SANDER & TAYLOR, *supra* note 3, at 107–08 ("Taking [selection] bias into account, these studies as a group provide substantial—if not definitive—evidence that mismatch reduces minority graduation rates.").

Table 2: Summary of the Graduation Rate/Wages Literature

SANDER & TAYLOR	THIS REVIEW
Grad. Rates: National[109] Loury & Garman, 1995* Light & Strayer, 2000?**	**Grad. Rates: National (and Texas)** Golann et al., 2013; Cortes, 2010; Bowen et al., 2009; Espenshade & Radford, 2009; Melguizo, 2008; Fisher & Massey, 2007*; Massey & Mooney, 2007; Small & Winship, 2007; Gong, 2006*; Holzer & Neumark, 2006*; Alon & Tienda, 2005; Light & Strayer, 2002**; Bowen & Bok, 1998; Kane, 1998* HBCUs: Flores & Park, 2013; Allen et al., 2007*; Kim & Conrad, 2006; Ehrenberg et al., 1999; Allen, 1992*
Grad. Rates: California Arcidiacono et al., 2012 Sander & Taylor, 2012 (see also Sander, 2010)	**Grad. Rates: California+** Arcidiacono & Koedel, forthcoming; Kurlaender & Grodsky, 2013; Chingos, 2013; Kidder, 2013; Arcidiacono et al., 2012; Backes, 2012; Hinrichs, 2012; Chang & Rose, 2010
Wages Dale & Krueger, 2002?*** Loury & Garman, 1995*	**Wages** Dale & Krueger, 2011 and forthcoming***; Andrews et al., 2012; Long, 2010; Daniel et al., 2001; Monks, 2000; Hoekstra, 2009[110]; HBCUs: Fryer & Greenstone, 2010*; Ehrenberg et al., 1999*; Kane, 1998*

In summary, our review and synthesis of the social science around college graduation rates, labor market earnings, and the mismatch hypothesis, reflected in Table 2, reveals that Sander and Taylor have cherry-picked[111] data to support their conclusions, and they substitute

109. We debated adding Cole and Barber to the Sander and Taylor column. *See generally* STEPHEN COLE & ELINOR BARBER, INCREASING FACULTY DIVERSITY (2003). While that book does address African American and Latino college degree attainment somewhat, *see id.* at 226–30, it is referenced by Sander and Taylor primarily around STEM mismatch and other issues. SANDER & TAYLOR, *supra* note 3, at 44–47, 283.

110. This study analyzed only white men. *See supra* note 106 and accompanying text.

111. *See* Roy L. Brooks, *Helping Minorities by Ending Affirmative Action? A Review of* Mismatch: How Affirmative Action Hurts Students It's Intended to Help, and Why Universities Won't Admit It (San Diego Legal Studies Paper Series, Paper No. 13-133, 2013) (manuscript at 37), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2327713 ("The authors simply do not engage this evidence. Instead, they rely on ridiculously narrow definitions of

questionable *ipse dixit* models and premises (e.g., the cascade effect model purporting to reveal significant harm to minority students who would otherwise go to middle-tier universities) rather than engaging in a real and robust attempt to address the cumulative (and largely peer-reviewed) social-science evidence discussed herein. Not surprising then, in the *Fisher* case nearly a dozen top social scientists and methodologists from various academic disciplines—including Gary King and Donald Rubin, who are members of the National Academy of Science—filed an amicus brief responding to Sander and Taylor's *Fisher* brief. The leading empirical scholars reviewed Sander's prior data and methods and other studies cited in the Sander and Taylor brief, and concluded:

> Whether one finds Sander's conclusions highly unlikely or intuitively appealing, his "mismatch" research fails to satisfy the basic standards of good empirical social-science research. The Sander-Taylor Brief misrepresents the acceptance of his hypothesis in the social-science community and, ultimately, the validity of mismatch. Numerous examples exist of better ways to perform the type of research Sander undertook. Sander's failure to set up proper controls to test his hypothesis and his reliance on a number of contradictory assumptions lead him to draw unwarranted causal inferences. At a minimum, these basic research flaws call into question the conclusions of that research.
>
> . . .
>
> In light of the many methodological problems with the underlying research, *amici curiae* respectfully request that the Court reject Sander's "mismatch" research . . . .[112]

To the extent Sander and Taylor attempt to deflect the searing rebuke in the Empirical Scholars' brief by claiming it was too singularly focused on law

---

academic and professional success . . . and cherry pick the data on the effects Prop 209 has had on black students.").

112. Brief of Empirical Scholars as *Amici Curiae* in Support of Respondents at 27–28, Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411 (Aug. 13, 2012) (No. 11-345). Sander and Taylor attempt to respond to the Empirical Scholars in their *Schuette* brief, and some of their responses are peculiar. They speculate as to the reason that "most of the distinguished signatories" agreed to sign the brief, claim that Empirical Scholars cite but failed to review the details of twenty cited journal articles critical of Sander "since the specific arguments have been answered so decisively as to be discredited," and claim in the accompanying footnote, "Most authors of these critiques have generally made no substantive reply to scholarly responses. Specifically, there has been no further defense of the critiques advanced by Ian Ayres, Richard Brooks, Jesse Rothstein, Albert Yoon, David Wilkins, or Mitu Gulati." Brief for Sander & Taylor, *Schuette* Case, *supra* note 4, at 25 & n.70, 26 . This latter claim about "further defense of the critiques" is a mischaracterization about how social-science scholarship normally works, as there is typically no social-science norm that is the equivalent of a sur-reply legal brief—and one of us (Kidder) is speaking from direct experience as one of the foolhardy minority of scholars who "replied to Sander's reply" by posting a working paper responsive to Sander's *Reply to Critics* piece in the *Stanford Law Review*, *see* Lempert et al., *supra* note 36.

school mismatch and neglected "academic [undergraduate?] mismatch or the stunningly positive effects of Proposition 209 at the University of California,"[113] our Review has a lot to say about academic/undergraduate mismatch and Proposition 209 but little to say about law school mismatch (for space reasons), yet our conclusions in this Review closely parallel the collective judgment rendered by our more esteemed colleagues who authored the Empirical Scholars' brief.

II. The Warming Effect and Stigma?: Keepin' It Real?

Much like they did when discussing graduation rates and post-graduation wages, Sander and Taylor, in their book *Mismatch*, also failed to examine all available data, looking only to studies that they view as supporting their claims and turning a blind eye to the bulk of research on these topics, as well as the overall demographic changes that have occurred in California.

A. Examining the Direct Evidence

Sander and Taylor devote a chapter to the "warming effect" of Prop. 209, which is their rejoinder to the notion that affirmative action bans can result in "chilling effects," whereby URMs perceive university campuses with such bans as less welcoming.[114] Sander and Taylor posit that "[i]t is worth standing back and asking whether a rigorous analysis of all the available data supports" their opposite claim of a warming effect hypothesis about Prop. 209.[115] At the university application stage, Sander and Taylor focus attention on one 2005 study by Card and Krueger, which found application patterns to be unchanged after Prop. 209 among high-credential

---

113. Brief for Sander & Taylor, *Schuette* Case, *supra* note 4, at 24. This claim is also dubious in light of the studies discussed in Brief of Empirical Scholars as *Amici Curiae* in Support of Respondents, *supra* note 112, at 14–16, regarding undergraduate-level mismatch research, including Alon and Tienda; Fischer and Massey; Kane; Long; Small and Winship; Cortes; Melguizo; and Bowen and Bok.

The Sander and Taylor *Schuette* brief even has an amusing tidbit of criticism directed at one of us (Kidder) regarding data transparency. *See* Brief for Sander & Taylor, *Schuette* Case, *supra* note 4, at 23. But the Kidder memo to the State Bar of California cited in the Sander and Taylor brief stakes out a different position than those totally opposed to release of California bar data, recommending: "If the Sander et al. team were to overcome the methodological, data privacy and sample size concerns detailed herein, and the State Bar was then inclined to release the data, this should only be done with a prior agreement that the same access will be granted to other bona fide researchers." Memorandum from Bill Kidder, Special Assistant to the Vice President, Student Affairs, Univ. of Cal. Office of the President, to Gayle Murphy, Senior Exec. for Admissions, Office of Admissions, State Bar of Cal. 2 (Jan. 19, 2007), *available at* http://www.seaphe.org/pdf/bar-proposal/kidder_critique.pdf. *Contra* Brief for Sander & Taylor, *Schuette* Case, *supra* note 4, at 23 n. 60 (citing Memorandum from Bill Kidder, *supra*).

114. *See* SANDER & TAYLOR, *supra* note 3, at 131–42.

115. *Id.* at 135.

URMs in California.[116] However, Sander and Taylor make no mention of (or attempt to distinguish) three studies by Long, Dickson, and Brown and Hirschman that found net declines in applications by URMs after affirmative action bans in California, Texas, and Washington, respectively.[117] Thus, Sander and Taylor fall short of their own benchmark of looking soberly at all available data.[118]

Moreover, to the extent Sander and Taylor might justify their focus on Card and Krueger because of Sander and Taylor's disproportionate policy interest in the behavior of URMs with the highest credentials,[119] we note such a justification is inconsistent with Sander and Taylor's focus on lower credential black and Latino admits to UC campuses as the basis for their claims discussed further below about rates of accepting admission offers (i.e., yield rates) and Prop. 209's supposed "warming effect."

Sander and Taylor then turn to a study of UC yield rates by Antonovics and Sander, which compared yield rates among admitted students at UC campuses in 1995–1997 versus 1998–2000, as their key evidence of a post-Prop. 209 "warming effect."[120] Sander and Taylor then add theoretical embellishment to their findings about "warming effects" by asserting that Prop. 209 may have caused African Americans and Latinos admitted to UC to feel "more intellectually self-confident and less (if at all) stigmatized" and that, conversely, there is little support for the "critical

---

116. *Id.* at 136–37 (discussing David Card & Alan B. Krueger, *Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas*, 58 INDUS. & LAB. REL. REV. 416 (2005)).

117. *See* Susan K. Brown & Charles Hirschman, *The End of Affirmative Action in Washington State and Its Impact on the Transition from High School to College*, 79 SOC. EDUC. 106, 125 (2006) (interpreting the drop in minority applications after Washington State's affirmative action ban as a "discouragement effect" that followed the ban); Lisa M. Dickson, *Does Ending Affirmative Action in College Admissions Lower the Percent of Minority Students Applying to College?*, 25 ECON. EDUC. REV. 109, 116 (2006) (finding a decrease in the number of Hispanic and black applicants applying to college in Texas after the Top Ten Percent Plan, which essentially ended affirmative action, was put into place); Mark C. Long, *College Applications and the Effect of Affirmative Action*, 121 J. ECONOMETRICS 319, 324–25 (2004) (finding that in California, URMs sent relatively fewer applications to colleges after Prop. 209).

118. Another recent study by a coauthor of Sander reached ambiguous results regarding "chilling effects." Kate Antonovics & Ben Backes, *Were Minority Students Discouraged from Applying to University of California Campuses After the Affirmative Action Ban?*, 8 EDUC. FIN. & POL'Y 208, 249 (2013) ("An important issue in the debate surrounding Prop 209 . . . is whether [such bans] lowered the value URMs placed on attending UC schools. . . . Unfortunately, our results do not allow us to make definitive conclusions about this kind of 'chilling effect'. . . .").

119. *See* SANDER & TAYLOR, *supra* note 3, at 136 (suggesting that the Card and Krueger study used only highly qualified applicants because of the belief that those applicants would get into the schools both before and after Prop. 209).

120. *Id.* at 137–38 (discussing Kate L. Antonovics & Richard H. Sander, *Affirmative Action Bans and the "Chilling Effect,"* 15 AM. L. & ECON. REV. 252, 279 (2013)).

mass" hypothesis[121] (a key issue in the remanded *Fisher v. University of Texas* case[122]).

One of us has written in more detail elsewhere about UC yield rates and the problems with Sander et al.'s claims in both a refereed journal and a working paper,[123] so here we simply note a handful of points that have implications for Sander and Taylor's "warming effect" claim in *Mismatch*, and then we move on to a broader discussion of "stigma." First, Sander and Taylor claim that, under Prop. 209 at UC campuses, "it seems that the aura of race-neutrality attracted many, many more black and Hispanic students than it repelled."[124] However, the Antonovics and Sander data show that URM yield rates to the *UC system* went down (in absolute and relative terms) after Prop. 209 *even though* URM yield rates purportedly *went up on individual UC campuses.*[125] Thus, as a claim about numbers, Sander and Taylor's claim makes little sense unless (as occurs elsewhere in *Mismatch*), the authors are relying on extraneous trends to do the "heavy lifting" behind their Prop. 209 claim, such as the increase in total available freshmen "seats" at UC campuses between the mid-1990s and the early 2000s or the growth in Latino, college-going, high school graduates in California during that time.[126]

Second and relatedly, the most straightforward analytical question Antonovics and Sander could have looked at is whether Prop. 209 "warmed" more URMs to choose a UC campus without affirmative action *instead of* selective private institutions with affirmative action. However,

---

121. *Id.* at 153.

122. *See* Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411, 2416 (2013) (describing the University's goal of attaining "critical mass" as the reason behind its decision to include race in the admissions process, suggesting that the "critical mass" theory is a point of contention).

123. Kidder, *supra* note 15, at 71–85; WILLIAM C. KIDDER, CIVIL RIGHTS PROJECT, THE SALIENCE OF RACIAL ISOLATION: AFRICAN AMERICANS' AND LATINOS' PERCEPTIONS OF CLIMATE AND ENROLLMENT CHOICES WITH AND WITHOUT PROPOSITION 209, at 15–32, app. B. at 37–42 (2012), *available at* http://civilrightsproject.ucla.edu/research/college-access/affirmative-action/the-salience-of-racial-isolation-african-americans2019-and-latinos2019-perceptions-of-climate-and-enrollment-choices-with-and-without-proposition-209/Kidder_Racial-Isolation_CRP_final_Oct2012-w-table.pdf.

124. SANDER & TAYLOR, *supra* note 3, at 139.

125. *See* Antonovics & Sander, *supra* note 120, at 273 tbl.4 (finding an overall 1.9% decrease for the UC system but an increase varying between 5.8% and 1.3% for individual UC campuses).

126. *See* Brief of Civil Rights Project/Proyecto Derechos Civiles as *Amicus Curiae* in Support of Respondents Chase Cantrell et al. at 11, 12 & n.14, Schuette v. Coal. to Defend Affirmative Action, No. 12-682 (U.S. July 1, 2013); Brief for the President and Chancellors of the University of California as Amici Curiae in Support of Respondents, *supra* note 87, at 22–23; PATRICIA GÁNDARA, CIVIL RIGHTS PROJECT, CALIFORNIA: A CASE STUDY IN THE LOSS OF AFFIRMATIVE ACTION: A POLICY REPORT 5–8 (2012), *available at* http://civilrightsproject.ucla.edu/research/college-access/affirmative-action/california-a-case-study-in-the-loss-of-affirmative-action; Kidder, *supra* note 15, at 89–90.

Antonovics and Sander did not have data on selective private colleges.[127] Even worse, they claim that their study was the first to investigate pre- and post-Prop. 209 yield rates in a systematic manner, yet they were seemingly unaware of Geiser and Caspary's study (using 1997–2002 data),[128] finding that after Prop. 209, "private selective enrollment of top URM admits to UC jumped by approximately six percentage points in 1999–2000, while the UC enrollment rate for these students fell by almost the same amount."[129] Ten years of post-Prop. 209 data suggest *that relative to a pre-Prop. 209 baseline of 1997*, the gap between URMs enrolling at selective privates widened compared to whites, Asian Americans, or others in both the top and middle thirds of UC's admit pool.[130] Such findings are inconsistent with Sander and Taylor's warming effect hypothesis and are consistent with the chilling-effect hypothesis.

Third, Sander and Taylor claim the warming effect is all the more remarkable given the cessation of race-conscious financial aid after Prop. 209,[131] but they (and Antonovics) again seem unaware of the anomalous situation whereby UC in-state and out-of-state tuition *decreased* by ten percent during the post-Prop. 209 years of their study (1998–2000 versus 1995–1997),[132] while at the same time that tuition *increased* nationwide between 1995 and 1999 by thirteen percent at public universities and eighteen percent at private universities.[133] Thus, UC had an unusually robust, if temporary, market price advantage among research universities in the years right after Prop. 209,[134] and Sander and Taylor fail to consider or account for that.[135]

---

127. *See* Antonovics & Sander, *supra* note 120, at 284 ("While our data do not allow us to directly examine what happened to URMs' relative chances of being admitted to schools outside the UC system after Proposition 209, we can calculate the net drop in the number of URMs enrolled in the UC system after Proposition 209.").

128. This study and its key findings were cited in one of our coauthored critiques of Sander's law school mismatch article, *see* Chambers et al., *supra* note 36, at 1864 n.32, to which Sander published a reply.

129. Saul Geiser & Kyra Caspary, *"No Show" Study: College Destinations of University of California Applicants and Admits Who Did Not Enroll, 1997–2002*, 19 EDUC. POL'Y 396, 401 (2005).

130. KIDDER, *supra* note 123, at 28–29. The same data are in Kidder, *supra* note 15, at 80 tbl.2, 81 tbl.3, but an error was introduced in the editing process so the "difference" row in table 3 is not correct.

131. SANDER & TAYLOR, *supra* note 3, at 139.

132. *See* KIDDER, *supra* note 123, at 39–40; *UC Mandatory Student Charge Levels*, U. OF CAL. OFFICE OF THE PRESIDENT, http://budget.ucop.edu/fees/documents/history_fees.pdf.

133. *See* CHRISTINA CHANG WEI & LUTZ BERKNER, NAT'L CTR. FOR EDUC. STATISTICS, TRENDS IN UNDERGRADUATE BORROWING II: FEDERAL STUDENT LOANS IN 1995–96, 1999–2000, AND 2003–04, at 23, 28 (2008), *available at* http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008179rev. Figures in the text above and this source are not adjusted for inflation.

134. Complementing the broad national trend data by Wei and Berkner are more precise data by Hemelt and Marcotte documenting that public research universities in California (i.e., the University of California) experienced a temporary decline in total tuition costs in the late-1990s

Fourth, the unit-level data obtained by Antonovics and Sander have some advantages, but one disadvantage seems to be a greater propensity for missing data, and another disadvantage is that they were unable to separately analyze African Americans and Latinos even though those two groups exhibit important differences. For example, an exchange with Sander shows that his UC data indicate that URMs in the top third of UCLA's admit-pool yield rates rose from 13.5% in 1995–1997 to 17.3% in 1998–2000,[136] whereas the data we obtained (also from the UC Office of the President, like Antonovics and Sander) indicate that for African Americans and Latinos there was a decline between 1995–1997 (18.5%) and 1998–2000 (17.2%) in the top third of UCLA's admit pool.[137] For African Americans reported separately, there was a more substantial drop in the top third of UCLA's admit pool—from 29% in 1995–1997 to only 8% in 1998–2000.[138] All UC campuses saw disproportionate declines in African American and Latino yield rates in the top thirds of UC campus admit pools, and over a dozen times in the years 1998–2011 there were African American yield rates in the top third of UC campus admit pools

during the same time tuition increased at research universities in Florida and Texas and was flat in New York. *See* Steven W. Hemelt & Dave E. Marcotte, *Rising Tuition and Enrollment in Public Higher Education*, (Inst. for Labor Studies Discussion Paper Series, Paper No. 3827, 2008) (manuscript at 14, 24 fig.23), *available at* ftp://ftp.iza.org/SSRN/pdf/dp3827.pdf.

135. The relationships between financial considerations and student enrollment choice are complex and need to be carefully considered. *Cf.* Laura W. Perna & Marvin A. Titus, *Understanding Differences in the Choice of College Attended: The Role of State Public Policies*, 27 REV. HIGHER EDUC. 501 (2004).

136. Letter from Richard Sander, Professor, UCLA School of Law, to author (July 16, 2013) (on file with author).

137. Reply Memorandum from author to Richard Sander, Professor, UCLA School of Law, (July 29, 2013) (on file with author).

138. *Id.* A partial explanation may be that our data were for California resident applicants, while Antonovics and Sander's data included out-of-state applicants. To the extent nonresident admittees are more affluent—and less likely to be URMs and to have modest yield rates because they are, by definition, greater participants in the "national admissions market" with many good choices across the country—Antonovics and Sander's study may be capturing a spurious correlation associated with demographic differences between in-state and out-of-state candidates in the UC admissions pool. Regarding the meaning and import of the national admissions market, see, for example, Caroline M. Hoxby, *The Changing Selectivity of American Colleges*, 23 J. ECON. PERSP. 95 (2009), documenting the increasingly national admissions market, which increases the policy relevance of attending highly selective colleges vis-à-vis long-term career outcomes).

(including three times at UC Berkeley) that fell to the "inexorable zero,"[139] which is something that never occurred on UC campuses in 1994–1997.[140]

Fifth, the Antonovics and Sander results are being driven by yield rates in the bottom third of the UC admit pool, which is the area least relevant to the analysis of "warming effects" and stigma[141] and is inconsistent with the *Mismatch* book's emphasis on Card and Krueger's study of the most competitive URM applicants. While Sander and Taylor claim the opposite—celebrating "astonishing" gains at UC Berkeley's ability to enroll the most competitive African Americans' in 1998 immediately after Prop. 209, their claim is demonstrably false.[142] Moreover, in the bottom third of the UC admit pool, there are additional confounders not adroitly handled by Antonovics and Sander. At UCLA, for example, the NCAA data indicate that student–athletes receiving scholarships were 7.3% of African American freshmen in 1995–1997 versus 12.8% in 1998–2000.[143] The shift in the

---

139. *See, e.g.*, Johnson v. Transp. Agency, 480 U.S. 616, 656–57 (1987) (O'Connor, J., concurring) (comparing the percentage of available women in the workforce to the fact that zero women were in fact employed, noting that this fact was "sufficient for a prima facie Title VII case brought by unsuccessful women job applicants," and concluding that this statistic was a proper justification to institute an affirmative action program); Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 342 n.23 (1977) (stating that "the company's inability to rebut the inference of discrimination came not from a misuse of statistics but from 'the inexorable zero'").

140. KIDDER, *supra* note 123, at 24–25.

141. *Id.* at 24, app. B at 37–38.

142. Sander and Taylor also focus on the 1998 admissions cycle at Berkeley and claim that the African American yield rate immediately after Prop. 209 in 1998 was "particularly astonishing because the black students admitted that year had, on average, far stronger academic records than their predecessors." SANDER & TAYLOR, *supra* note 3, at 134. We believe Sander and Taylor's claim—or perhaps it is better described as a gossamer chain of statements that give the reader the impression they are making a claim about the credentials of enrolled African American students at Berkeley—used to bolster the "warming effect" hypothesis, is demonstrably false. What was astonishing was the drop in African American freshmen who enrolled at Berkeley in 1998, but the average credentials of those who did enroll that year were similar to other years. The table below on average SAT scores for African American freshmen admits and enrollees shows that the average SAT score for enrolled black freshmen in 1998 actually *dropped* 23 points compared to the prior year with affirmative action.

Year	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*
Admits	1132	1133	1136	1165	1154	1167
Enrollees	1082	1089	1087	1064	1057	1102

Perhaps Sander and Taylor are confusing data about admits and enrollees or African Americans versus Latinos (or there are deeper "missing data" problems on their end). The average SAT score of black admits at Berkeley went up 29 points, but that fact accompanied by the 23 point decline in the black enrollees' SAT averages in 1998 is highly inconsistent with Sander and Taylor's warming effect and is consistent with the studies (Geiser and Caspary, 2005 and Kidder, 2012) pointing to a chilling effect at UC Berkeley. The data in the above table was generated by UC Office of the President's Statfinder in 2012, a query tool that is no longer available, but charts with these SAT data for all racial/ethnic groups at UC Berkeley and UCLA covering 1994 to 2009 are available at Kidder, *supra* note 15, at 95–96.

143. The data show that 54 out of 739 African American freshmen received scholarships in 1995–1997 versus 59 out of 461 in 1998–2000. *See Federal Graduation Rates: University of California, Los Angeles, Education & Research*, NCAA, http://fs.ncaa.org/Docs/newmedia/public/

concentration of recruited student athletes among UCLA's African American, Prop. 209 freshmen population is consequential because while other high school seniors are making up their mind in April about enrolling at UC, recruited athletes commit to a university under an earlier, and very distinct, recruitment process that other researchers try not to confound with the general campus admissions and recruitment cycle.[144] This reinforces the previous point that the "warming effect" data cited in *Mismatch* regarding "blacks and Hispanics"[145] have not been shown to meaningfully apply to African American students specifically.

For all of the aforementioned reasons, Sander and Taylor (and Antonovics and Sander) do not fashion a good test of holding out Prop. 209 as the basis for the stigma-reducing, "warming effect" hypothesis that they advocate.[146] The two of us have written separately about the topic of stigma in the context of affirmative action and have tested the extent of affirmative action's purported causal role by comparing survey data at institutions with and without affirmative action at the law school[147] and undergraduate[148] levels. Unfortunately, the *Mismatch* book by Sander and Taylor participates in a too-familiar political trope of affirmative action critics—including Justice Clarence Thomas[149]—deriding the harmful impact of the

---

rates/index.html. These NCAA federal graduation-rate reports only go back to 1998, but the 1998 report lists four years of data (1995–1998) from which the 1995–1997 data can be obtained by deleting the 1998 totals.

144. *See* Stephen L. DesJardins, *An Analytic Strategy to Assist Institutional Recruitment and Marketing Efforts*, 43 RES. HIGHER EDUC. 531, 534 (2002) ("Recruited athletes are eliminated since the recruitment process for student-athletes is markedly different than for students in general.").

145. *See, e.g.*, SANDER & TAYLOR, *supra* note 3, at 138 & fig.8.1 (indicating that the "announced end of racial preferences at the University of California coincided with a jump in the rate at which blacks and Hispanics accepted offers of admission from UC schools").

146. *See* Antonovics & Sander, *supra* note 120, at 288–90 ("Removing the stigma of being a 'special admit' has both social and economic advantages. Being a URM admitted without a racial preference could increase the signaling value of one's college degree; thus, Proposition 209 may have increased the signaling value of a UC degree for URMs.").

147. *See* Angela Onwuachi-Willig et al., *Cracking the Egg: Which Came First—Stigma or Affirmative Action?*, 96 CALIF. L. REV. 1299, 1304 (2008) (administering a survey related to stigma issues to law students at UC Berkeley, UC Davis, Cincinnati, Iowa, Michigan, Virginia, and Washington).

148. KIDDER, *supra* note 123, at 20–32; Kidder, *supra* note 15, at 57–85.

149. *See* Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411, 2422–32 (2013) (Thomas, J., concurring) ("We acknowledged the possibility of stigma but nevertheless concluded that the reality of private biases and the possible injury they might inflict do not justify racial discrimination." (internal quotation marks omitted)); Grutter v. Bollinger, 539 U.S. 306, 373 (2003) (Thomas, J., concurring in part and dissenting in part) (protesting that African Americans admitted to law schools are "tarred as undeserving" because of affirmative action); Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 240–41 (1995) (Thomas, J., concurring) (objecting to the premise that there is a "racial paternalism exception to the principle of equal protection"); *see also* Angela Onwuachi-Willig, *Just Another Brother on the SCT?: What Justice Clarence Thomas Teaches Us About the Influence of Racial Identity*, 90 IOWA L. REV. 931, 987–96 (2005)

"stigma" supposedly created by race-conscious policies without either a serious theoretical understanding of stigma scholarship or firm data delineating the causal role of affirmative action (as opposed to the longstanding and deep-seated sociological phenomenon of racial stigma that is rooted in America's legacy of racial inequality).[150]

An additional example is in Sander and Taylor's portrayal of the stigma-related study by Sidanius, Levin, van Laar et al., who found that African Americans and Latinos at UCLA in 1996 who *believed* they were admitted due to affirmative action had, controlling for SAT scores, lower self-reported academic performance at the end of their freshmen year.[151] Sander and Taylor acknowledge that Sidanius, Levin, van Laar et al.'s "remarkable finding" about stereotype threat is "probably real," but then they pivot to misappropriate this study under the mismatch banner by claiming that stereotype threat "plausibly will be most severe for students admitted with the largest racial preferences."[152] But Sander and Taylor's "spin" is directly at odds with what the authors of this study (in both a companion article and the book) state: "We do not take our findings to indicate that affirmative action is harmful for ethnic minority students. On the contrary, suspecting that one was a beneficiary of affirmative action impaired ethnic minorities' academic performance *only when it was accompanied by personal or social identity stereotype threat.*"[153] In this study, students' SAT scores explained only 2% of the variance in whether

---

(discussing at length Justice Thomas's views with respect to law school affirmative action and the negative perceptions it can promulgate).

150. *See* Onwuachi-Willig et al., *supra* note 147, at 1308–24 (tracing the storied history of stigma as related to affirmative action); *see also* Christopher A. Bracey, *The* Cul de Sac *of Race Preference Discourse*, 79 S. CAL. L. REV. 1231, 1234 (2006) (suggesting that the debate over affirmative action has "devolve[d] into disengaged moral and ideological posturing"); R.A. Lenhardt, *Understanding the Mark: Race, Stigma, and Equality in Context*, 79 N.Y.U. L. REV. 803, 809 (2004) (arguing that when understood in context, stigma is the cause of many racial harms, and that intentional discrimination and racialized behavior are a function of racial stigma, not vice versa). From the shrewd standpoint of political persuasion, commitments to theoretical coherence and evidence-based argument by such affirmative action critics becomes epiphenomenal. *See* Onwuachi-Willig et al., *supra* note 147, at 1323 (noting that in the political context, "stigma rhetoric is persuasive because of how it impacts the ordering of our national values and political commitments").

151. JIM SIDANIUS ET AL., THE DIVERSITY CHALLENGE: SOCIAL IDENTITY AND INTERGROUP RELATIONS ON THE COLLEGE CAMPUS 287–88 (2008). This study used both self-reported college GPA and self-perceived performance ("How well will you do (are you doing) in school, compared to other students at UCLA?") on a seven-point scale. *See id.* at 255, app. A at 326. For more information, such as the methods used in the companion study, see Colette van Laar et al., *Social Identity and Personal Identity Stereotype Threat: The Case of Affirmative Action*, 30 BASIC & APPLIED SOC. PSYCHOL. 295, 298–99 (2008).

152. SANDER & TAYLOR, *supra* note 3, at 105–06.

153. van Laar et al., *supra* note 151, at 308. Jim Sidanius was not a coauthor of this companion article, but his book similarly states that "affirmative action did not have harmful effects on later academic performance, unless that student was concerned about the negative stereotypes about his or her group." SIDANIUS ET AL., *supra* note 151, at 290–91.

students believed they were admitted because of affirmative action (r = −.15), and two-fifths (41%) of the 54 African Americans in this study did not believe affirmative action was a factor, a combination of facts that is hardly an endorsement of the mismatch hypothesis.[154]

Moreover, Sidanius, Levin, van Laar et al., properly acknowledge that for black and Latino college students, academic stigma and stereotype threat are "part of a larger set of minority status stressors that can undermine minority students' psychological and academic outcomes"; therefore, they recommended that universities communicate to students of all backgrounds that the "institution is committed to maintaining a positive campus racial climate."[155]

B. Campus Climate Survey Data and the "Warming Effect"

The above discussion segues our Review to a key natural-experiment question that Sidanius, Levin, van Laar et al. could not analyze, but that is central to Sander and Taylor's "warming effect" and stigma reduction hypotheses.[156] From the perspective of black and Latino undergraduates, do UC campuses after Prop. 209 have a "warmer" campus racial climate whereby URMs feel more respected and less stigmatized than their peers at comparable leading research universities with affirmative action? Or do UC campuses with low diversity levels because of the affirmative action ban have black and Latino students who feel less respected compared to those at universities with affirmative action or higher diversity levels (i.e., critical mass)? Against this benchmark of student perceptions about campus racial climate (and stigma salience), can Sander and Taylor's claims—that the "size of the warming effect should be, as it is, closely related to the reduction in racial preferences after Prop. 209[;] [P]references fell dramatically at Berkeley and UCLA, and this had particularly impressive warming effects"[157]—still be substantiated?

---

154. van Laar et al., *supra* note 151, at 298–301. Though not definitive, the fact that 41% of African Americans but only 28% of Latinos in this study did not believe that affirmative action was a factor in their UCLA admission, *see id.* at 301, suggests that a student's self-perceptions are important regardless of whether they are objectively accurate or not, which again cuts against the mismatch hypothesis. The authors also eliminated reverse causation (i.e., lower academic performance was not associated with increased identity stereotype threat). *Id.* at 304–05.

155. SIDANIUS ET AL., *supra* note 151, at 291.

156. Sander and Taylor state:

> But, of course, another possibility was at least equally plausible: that students of color would welcome the chance to attend a school without the stigma of being a suspected 'affirmative-action admit.' They may have anticipated that under a race-neutral regime campus life would be easier and that white and Asian students would be less likely to stereotype them as academically weak and more likely to be friends.

SANDER & TAYLOR, *supra* note 3, at 140.

157. *Id.* at 141.

To address these questions relevant to warming effects and stigma, we present data from a campus survey item administered at thirty campus-level data points between 2008 and 2012, which includes twenty-five administrations at UC campuses, two at UT Austin, and three at other leading research universities that were willing to share their data if their institutions were not named (AAU #1 and AAU #2).[158] This survey asked undergraduates if they believed that students of their race or ethnicity were respected on campus, and includes over 3,000 African American and over 17,000 Latino respondents, which is an unusually large sample relative to the campus climate research literature.[159]

In the set of UC campuses on the right side of Figure 2A—Berkeley, Davis, Irvine, UCLA, San Diego, Santa Barbara, and Santa Cruz—African Americans are only 2%–4% of the student body, and on these campuses, only 59.0% of African Americans feel respected (defined as students who responded that they "strongly agree," "agree," or "somewhat agree"). The set of universities on the left side—UT Austin, UC Riverside, UC Merced, AAU #1, and AAU #2—are ones where African Americans are 5% or more of the student body and include cases with affirmative action. At this set of universities, by contrast, the percentage of African American undergraduates who report feeling respected is 79.9%, approximately 21 percentage points higher (or 20 percentage points higher if excluding UC Merced[160]). There is a robust relationship between African American representation in the student body and the percentage of these students who feel respected on their campus ($R^2 = 0.52$).[161] All of this runs contrary to

158. Kidder, *supra* note 15, at 60–61.

159. Our findings here add 2012 data to companion papers by Kidder that provide additional detail about these survey data. *See* KIDDER, *supra* note 123, at 34–37 (using 2008–2011 data but noting that the data do not include 2012 data, which, as noted, is what is added in this subpart); Kidder, *supra* note 15, at 61–63 (using 2008–2011 data). UC Merced did not administer this survey item in 2008 and 2010. KIDDER, *supra* note 123, at 34–35.

160. It can be argued that UC Merced is not as comparable as the other 29 campus data points because UC Merced is a small and new campus that does not yet have a Carnegie classification as a "very high" research university. *See* Kidder, *supra* note 15, at 63 n.22 (indicating that the "much smaller" UC Merced campus was not included in the study); Results of Search for Institutions with a "Very High" Carnegie Classification, CARNEGIE FOUND., http://classifications.carnegiefoundation.org/lookup_listings/srp.php?clq={%22basic2005_ids%22%3A%2215%22}&limit=0,50 (listing doctorate-granting universities classified as having "very high research activity," of which UC Merced is not included). So if UC Merced is excluded, then 79.3% is average for the percentage of African American students who feel respected at the other universities in the left grouping in the chart.

161. These data are not part of a causal model, and we do not have the quantitative data on other contextual factors that may influence this relationship. The wider scholarly literature documents the multifaceted nature of a positive campus racial climate. *See* KIDDER, *supra* note 123, at 7–9 (summarizing several studies); Liliana M. Garces & Uma M. Jayakumar, *Dynamic Diversity: Toward a Contextual Understanding of Critical Mass*, EDUC. RESEARCHER (forthcoming);. Thus, it is very plausible (even expected) that if one were analyzing a broader representation of American public and private research universities that—in contrast to our data here—were not effectively "over-sampling" UC campuses under an affirmative action ban, the

the warming effect hypothesis, including the finding that UC Berkeley (51%–57%) and UCLA (49%–62%) are toward the lower end in terms of having African American students who feel respected, whereas the Sander and Taylor hypothesis is that Berkeley and UCLA should be the campuses where one can most readily see the unbounded Prometheus of ending "racial preferences" after Prop. 209.[162]

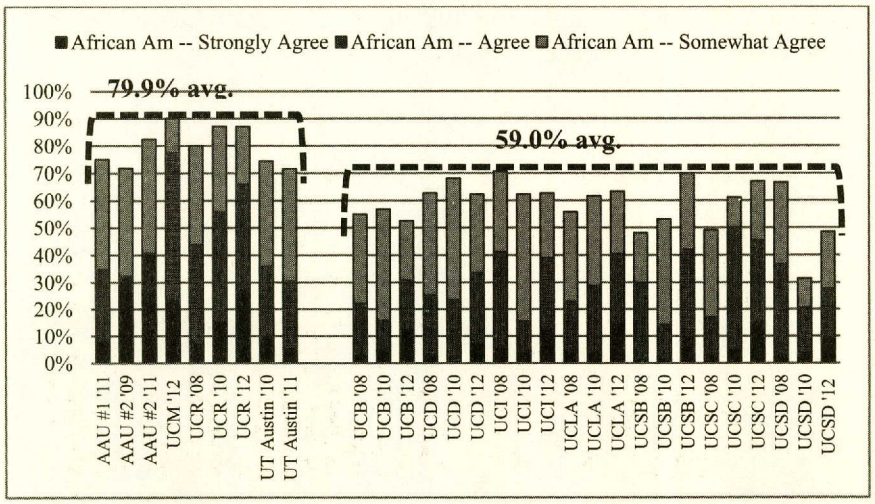
---

strength of the simple correlation between "critical mass" and URM students feeling respected would become more attenuated. Rather, our intent in this part of our Review is to present what lawyers and judges would refer to as "rebuttal evidence" vis-à-vis the claims Sander and Taylor make in *Mismatch* in a context in which the "over-sampling" of UC campuses is relevant and responsive because Sander and Taylor's claims are about Prop. 209.

162. Professor Sander was quoted in connection with the *Fisher* case as dismissive of "research based on surveys of students who felt pressure 'to endorse the diversity ideology' of their college." Peter Schmidt, *444 Scholars Tell Court that Research Supports Race-Conscious Admissions*, CHRON. HIGHER EDUC. (Aug. 10, 2012), https://chronicle.com/article/444-Scholars-Tell-Court-That/133515/. It is difficult to respond to such an abbreviated quote, but to the extent this is relevant to Professor Sander's reply to our Review, we simply note that the above quote appears to be self-referential and lacking in empiricism. It amounts to dodging rather than offering a viable alternative explanation for the core finding of Figures 2A and 2C regarding how comparable research universities exhibit considerable variation in URM students feeling respected and how that is associated with "critical mass" at least for this set of universities. As one distinguished sociologist at UCLA puts it, "[I]t goes without saying that survey research has its limitations: one wants to know, not just what people say, but what they do, though one would have to endorse a very strong view of the mind/body split to insist that what people say is of no value at all." Roger Waldinger, *The Bounded Community: Turning Foreigners into Americans in Twenty-first Century L.A.*, 30 ETHNIC & RACIAL STUD. 341, 367 (2007). Second, the studies by Park, Sax & Arredondo, and Edwards discussed later in this subpart, *see infra* notes 174–75 and accompanying text, all rely on CIRP freshmen surveys taken *just before students enrolled* in college, yet show a consistent pattern that black and Latino students have substantially more favorable attitudes about affirmative action in college admissions than white students. *See also* WALTER R. ALLEN ET AL., BLACK UNDERGRADUATES FROM *BAKKE* TO *GRUTTER*: FRESHMEN STATUS, TRENDS AND PROSPECTS, 1971–2004, at 23 (2005), *available at* http://www.heri.ucla.edu/PDFs/pubs/TFS/Special/Monographs/BlackUndergraduatesFromBakkeToGrutter.pdf ("In 2004, 50 percent of incoming freshmen felt affirmative action should be abolished, as compared to 25 percent of Black freshmen."). These studies are not the same as campus racial climate, obviously, but large pre-existing differences in students' attitudes by race/ethnicity are a reminder to readers that the "college indoctrination" hypothesis suggested by the Sander quote should—like so many claims in the *Mismatch* book—be regarded with strong skepticism.

"Students of my race/ethnicity are respected on this campus" Surveys in 2008–2012 (% strongly agree, agree, or somewhat agree)[163]

Figure 2A: African-American Undergraduates



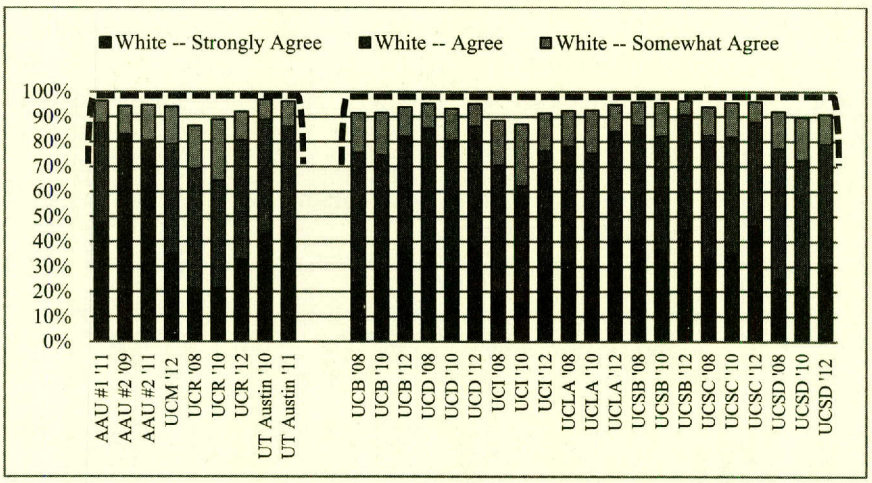
Juxtaposing Figure 2A on African Americans (above) with Figure 2B (below) for white students at the same set of universities provides additional confirmation that Sander and Taylor's working hypothesis—that affirmative action *qua* affirmative action is primarily or entirely causing the stigma that African American students face on college campuses—is shallow and poorly theorized.[164] In the set of UC campuses on the right side, only 34.5% of African Americans either "strongly agree" or "agree" that they are respected, which confirms that these students perceive their

163. Regarding statistical significance, the two-tailed *P* value is less than 0.0001 when comparing the respected/not respected totals for the two groups of universities. The group of universities on the left includes 1,421 African American respondents, and the group on the right includes 1,768 African Americans. Note that for AAU #2 there are no middle bars in Figures 2A–C for "agree" because the institution providing the data already combined "strongly agree" and "agree."

164. *See, e.g.*, Brenda Major & Laurie T. O'Brien, *The Social Psychology of Stigma*, 56 ANN. REV. PSYCHOL. 393, 412 (2005) ("[O]ne of the major insights . . . on stigma is the tremendous variability across people, groups, and situations in responses to stigma. The emerging understanding of . . . stigma and identification of effective coping strategies for dealing with identity-threatening situations holds some promise for improving the predicament of the stigmatized."); *see also* John F. Dovidio et al., *Stigma: Introduction and Overview*, *in* THE SOCIAL PSYCHOLOGY OF STIGMA 1, 16 (Todd F. Heatherton et al. eds., 2000) (explaining that the "oversimplification" of a "topic as broad and complex as stigma" may "obscure critical distinctions or exclude important points"); Cheryl R. Kaiser, *Dominant Ideology Threat and the Interpersonal Consequences of Attributions to Discrimination*, *in* STIGMA AND GROUP INEQUALITY: SOCIAL PSYCHOLOGICAL PERSPECTIVES 45, 45–64 (Shana Levin & Colette van Laar eds., 2006) (observing that social psychologists have just recently begun examining the interpersonal consequences of perceptions of prejudice).

world in a decidedly different manner than their white classmates, of whom 81.8% either "strongly agree" or "agree" that students of their race are respected on these same UC campuses.[165] Even the "outlier" data among white students are consistent with the "critical mass" hypothesis.[166] Moreover, the black–white student gaps in feeling respected are considerably worse at the UC campuses on the right side of the two charts.

Figure 2B: White Undergraduates



Likewise, other reports using the same UC Undergraduate Experience Survey (UCUES) data show that, at UC Berkeley for the 2008–2012 UCUES combined, on this "respect" survey item there is a gulf separating African Americans (52% at least somewhat agree they are respected) from more privileged and even other traditionally marginalized student affinity groups on campus, including students identifying as heterosexual (98%), white (93%), Asian (91%), bisexual (85%), Christian (83%), gay/lesbian (83%), Muslim (81%), and Jewish (75%).[167] And if excluding those who

165. Regarding statistical significance, the two-tailed *P* value is less than 0.0001 when comparing the totals for the African American and white respondents at the UC campuses on the right side of the two charts. This comparison includes 1,768 African American and 28,213 white respondents.

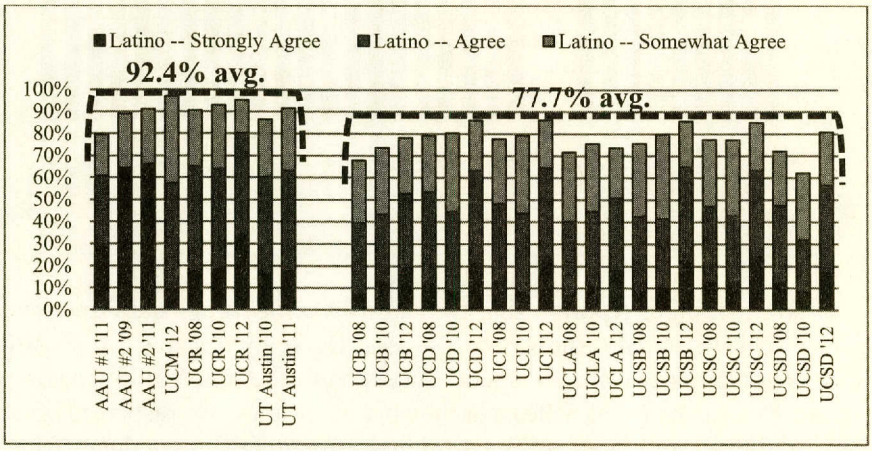
166. Where there is modest softening of whites' high percentage of feeling respected, at UC Riverside and UC Irvine, it is on those campuses where white students are a smaller percentage of the student body (under 20% in 2012), consistent with what one would predict a priori based on the "critical mass" hypothesis, other things being equal.

167. ANDREW EPPIG & SEREETA ALEXANDER, ASSESSING UNDERGRADUATE CAMPUS CLIMATE TRENDS AT UC BERKELEY 7–8 (2012), *available at* http://www.cair.org/conferences/cair2012/pres/32_Eppig.pdf (presentation at the 2012 California Association for Institutional Research conference). For this UC Berkeley report, the African American UCUES sample was 484, and all of the groups mentioned above had larger samples (e.g., Christian n = 6,544) except for Muslim students (n = 277). *Id.*

"somewhat agree," the proportion of African American UC Berkeley students who strongly agree or agree about feeling that students of their race are respected drops to half or below the levels for all the other above-mentioned groups.[168]

The data for Latinos are shown below in Figure 2C. While not as dramatic as the data for African Americans, there is a fifteen point difference (92.4% versus 77.7%) between Latinos feeling respected in the group of campuses on the left side of Figure 2C versus the group of UC campuses on the right side.[169] As with African Americans, Latinos have a lower sense of feeling respected at UC Berkeley and UCLA—the opposite of what would be predicted by the "warming effect" hypothesis advanced by Sander and Taylor.

Figure 2C: Latino Undergraduates[170]



Our results are consistent with other studies. Using the new *Diverse Learning Environments* survey,[171] Hurtado and Guillermo-Wann find:

---

168. *Id.* at 8.

169. To facilitate consistent comparisons, the campuses are clustered in 2C in the same order as in 2A and 2B. If arrayed in terms of Latinos' percentage of the student body, the results are only slightly different, and the big picture is the same.

170. Regarding statistical significance, the two-tailed P value is less than 0.0001 when comparing the respected/not respected totals for the two groups of universities. The group of universities on the left includes 5,405 Latino respondents and the group on the right includes 13,027 Latinos. As noted earlier, for AAU #2, the institution provided the data that already combined "strongly agree" and "agree."

171. SYLVIA HURTADO & CHELSEA GUILLERMO-WANN, HIGHER EDUC. RESEARCH INST., DIVERSE LEARNING ENVIRONMENTS: ASSESSING AND CREATING CONDITIONS FOR STUDENT SUCCESS (2013), *available at* http://heri.ucla.edu/dle/DiverseLearningEnvironments.pdf. This survey covered a broader set of colleges and universities and included 218 African Americans and 959 Latinos in the sample. *See id.* at 59–60. A concise policy brief with the key findings can be found in SYLVIA HURTADO & ADRIANA RUIZ, THE CLIMATE FOR UNDERREPRESENTED GROUPS

"While underrepresented minority students experience less frequent discrimination at more compositionally diverse institutions, negative climates still persist, especially for African American students and for students underrepresented in their major departments."[172] Likewise, Deirdre Bowen's finding that a higher proportion of the URM students from four states with affirmative action bans feel "[p]ressure to prove themselves academically because of race" compared to URM students from nearly two dozen states with affirmative action (74% versus 41%).[173] And the longstanding CIRP (Cooperative Institutional Research Program) freshmen surveys consistently show that African Americans on predominantly white campuses express by far the highest levels of support for (i.e., disagree with abolishing) "affirmative action in college admissions;"[174] that pattern held for the freshmen attending four-year universities in California who took the CIRP survey shortly before and after Prop. 209.[175] Not only do the results of all of these surveys run counter to Sander and Taylor's claims, but they must be understood in the context of our earlier point that Antonovics and Sander did not have separate data on African Americans. More broadly, reviewers of the wider literature (mostly on the employment sector) find that any negative stigma of being an affirmative action beneficiary is highly context-dependent and the negative effects can fade from relevancy under the right conditions.[176]

Conclusion

In conclusion, despite claims of rigor, Sander and Taylor failed throughout their book to look beyond the miniscule number of studies that support their claims and, in so doing, neglected to respond to mountains of

---

AND DIVERSITY ON CAMPUS fig.2 (2012), *available at* http://heri.ucla.edu/briefs/urmbriefreport.pdf.

172. *Id.* at 32.

173. Deirdre M. Bowen, *Brilliant Disguise: An Empirical Analysis of a Social Experiment Banning Affirmative Action*, 85 IND. L.J. 1197, 1222 tbl.2, 1223–24 (2010).

174. *See, e.g.*, Julie J. Park, *Taking Race into Account: Charting Student Attitudes Towards Affirmative Action*, 50 RES. HIGHER EDUC. 670, 675–76, 678 tbl.1 (2009); Linda J. Sax & Marisol Arredondo, *Student Attitudes Toward Affirmative Action in College Admissions*, 40 RES. HIGHER EDUC. 439, 443, 445 tbl.1 (1999).

175. *See* William A. Edwards, Student Attitudes Toward Affirmative Action in College Admissions and Racial Diversity Before and After Proposition 209, at 71–73, 87, 130 app. A (2008) (unpublished Ph.D. dissertation, Michigan State University) (on file with author) (analyzing 1996 and 2000 CIRP freshmen surveys of students at over thirty four-year colleges and universities in California).

176. *See* Faye J. Crosby et al., *Affirmative Action: Psychological Data and the Policy Debates*, 58 AM. PSYCHOLOGIST 93, 106 (2003) (suggesting that in everyday work conditions—where the competence of an affirmative action beneficiary's peers can be observed—there is less negative association with the affirmative action label); Linda Hamilton Krieger, *Civil Rights Perestroika: Intergroup Relations After Affirmative Action*, 86 CALIF. L. REV. 1251, 1261 (1998) ("Subsequent studies have demonstrated that the self-denigrating effects of affirmative action are highly sensitive to contextual variables and, under certain conditions, disappear entirely.").

research by many of the world's top social scientists that have found such claims about mismatch to be empirically groundless. Moreover, the few studies that Sander and Taylor examined and cited in support of their arguments about mismatch were either based on outdated data or their own or others' flawed empirical analyses.

Indeed, the one-sided nature of Sander and Taylor's arguments—the very way in which the two authors seem to pay no attention to white students with grades and scores that are comparable to those of allegedly "mismatched" students of color—exposes a fatal flaw about claims in their research. After all, if mismatch were such a problem, why would Sander and Taylor specifically link their analyses predominantly to race and affirmative action?[177] They could, for example, add gender and affirmative action, particularly in the sciences, to their discussion. Or better yet, they could make broader claims that include legacies—nearly all white students who find themselves "mismatched" at their institutions.[178] Indeed, consider the fact that Sander and Taylor supported and urged the Supreme Court to review the lawsuit by Abigail Fisher.[179] Had Fisher been admitted to the University of Texas at Austin, she, too, would have been a "mismatched" student. As the University proclaimed in its Supreme Court brief, Abigail Fisher (who had an Academic Index score of 3.1), "would not have been admitted to the Fall 2008 freshman class even if she had received a 'perfect' [Personal Achievement Index (PAI)] score of 6" (and her actual PAI was, in fact, lower than that).[180] In fact, Ms. Fisher was also denied admission to UT Austin's 2008 summer freshmen admissions program in which 168 African Americans and Latinos were also denied admission with AI/PAI scores equal to *or higher* than Fisher's (versus only a handful of African Americans or Latinos offered summer admission with lower AIs/PAIs than Fisher).[181] Moreover, while comparing students based solely on SAT

177. *See also* Kurlaender & Grodsky, *supra* note 93, at 294 ("Although the logic of the mismatch argument is color-blind, we have not been able to find an instance in which the mismatch argument has been deployed by advocates out of concern for white or Asian students.").

178. In their brief supporting Supreme Court review of the *Fisher* case, Sander and Taylor begin a discussion of mismatch by briefly noting that "admissions preferences — regardless of whether these are based on race, 'legacy' considerations, or other factors" cause lower grades, Brief Amicus Curiae for Richard Sander and Stuart Taylor, Jr. in Support of Petitioner at 4, Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411 (Oct. 19, 2011) (No. 11-345), but this is a rhetorical pivot and the thrust of their book and Supreme Court briefs focus on race/ethnicity.

179. *See, e.g.*, SANDER & TAYLOR, *supra* note 3, at 274–75 (asserting that the "Supreme Court case of *Fisher v. University of Texas* provides an opportunity for the Court to start us down this better path").

180. Brief for Respondents at 15–16, Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411 (2013) (No. 11-345) (internal quotation marks omitted). Fisher's exact PAI is in a sealed brief. *Id.* at 15.

181. For example, the *Fisher* Brief stated:

> Although one African-American and four Hispanic applicants with lower combined AI/PAI scores than petitioner's were offered admission to the summer program, so were 42 Caucasian applicants with combined AI/PAI scores identical to or lower than

scores is simplistic and not how college admissions really works, Fisher's SAT score of 1180 would have placed her below at least 84% of the summer-program students at UT Austin in 2008.[182] Yet, despite the fact that Abigail Fisher herself would have been subject to the purported harms of mismatch, Sander and Taylor praise her lawsuit as a critical intervention, noting that the "mismatch is bound to be a serious problem for the racially preferred at UT."[183] In fact, Sander and Taylor argued in the *Fisher* case that at UT, "Hispanics who are admitted due to preferences typically enter with markedly less academic preparation," and they cited as their supporting evidence that in 2009 Latinos admitted outside the Ten Percent Plan had SAT scores at the 80th percentile nationally in 2009, compared to the 89th percentile for whites and 93rd percentile for Asian Americans.[184] While Sander and Taylor argue that "*Fisher* does not directly pose the problem of mismatch. . . . But the mismatch issue lurks in the background,"[185] Abigail Fisher's SAT score was equivalent or lower to the Latino SAT mean score that Sander and Taylor cited as primary evidence of "markedly less academic preparation."[186] Notwithstanding the poor

---

> petitioner's. In addition, 168 African-American and Hispanic applicants in this pool who had combined AI/PAI scores identical to or *higher* than petitioner's were *denied* admission to the summer program.

Id. at 15–16. This since-discontinued summer program bears some resemblance to the "mismatched" students from the provisional University of California program that Kurlaender and Grodsky studied, though the latter was a one-time occurrence. *See supra* notes 87, 94–95 and accompanying text.

182. *Compare* Brief for Respondents, *supra* note 180, at 15 (identifying Fisher's SAT score of 1180), *with* UNIV. OF TEX. AT AUSTIN OFFICE OF ADMISSIONS, THE PERFORMANCE OF STUDENTS ATTENDING THE UNIVERSITY OF TEXAS AT AUSTIN AS A RESULT OF THE COORDINATED ADMISSION PROGRAM (CAP): STUDENTS APPLYING AS FRESHMEN 2008, at 4 tbl.5 (2011), *available at* http://www.utexas.edu/student/admissions/research/CAPreport-CAP08.pdf (demonstrating that a sum of 84% of the 2008 summer-program freshmen at UT Austin had SAT scores of 1200 or higher).

183. Kali Borkoski, *Ask the Author: Richard Sander and Stuart Taylor, Jr. on* Mismatch, SCOTUSBLOG (Oct. 16, 2012, 9:39 AM), http://www.scotusblog.com/2012/10/ask-the-author-richard-sander-and-stuart-taylor-jr-on-mismatch/.

184. Brief for Sander & Taylor, *Fisher* Case, *supra* note 4, at 3–4. A similar claim appears in SANDER & TAYLOR, *supra* note 3, at 288. It is unclear why they rely on 2009 data when Abigail Fisher applied in 2008.

185. SANDER & TAYLOR, *supra* note 3, at 289.

186. Sander and Taylor are referencing SAT percentile ranks for scores on the 2400-point scale that includes the writing section, but the sparse record in *Fisher* only seems to report her SAT of 1180 on the 1600-point scale (500 on critical reading; 680 on math). Joint Appendix at app. C at 41a, Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411 (2013) (No. 11-345). We say "or lower" in the text because it is unclear if Abigail Fisher took the SAT writing test, but if she did, then her discrepant scores between reading and math suggest that it may be too optimistic to assume that her SAT score on a 2400-point scale (i.e., including writing) was at the 80th percentile (1780) nationally. *See* COLLEGE BOARD, SAT PERCENTILE RANKS FOR MALES, FEMALES, AND TOTAL GROUP: 2008 COLLEGE BOUND SENIORS—CRITICAL READING + MATH + WRITING (no date), *available at* http://professionals.collegeboard.com/profdownload/sat_percentile_ranks_2008_composite_cr_m_w.pdf.

empirical support for the mismatch hypothesis that we documented earlier in this Review, for adherents like Sander and Taylor who believe that mismatch is prevalent and deeply harmful, the case of Abigail Fisher was one where mismatch was hardly lurking in the background—it was staring them directly in the face. The extent to which Abigail Fisher and a Latina applicant with equivalent qualifications (let's call her "Abigaíl Pescadora") are being marked in decidedly different ways in the affirmative action debate by Sander and Taylor[187] reveals a form of "doubletalk"[188] different than the type their book purports to expose.

Such gaps in analysis reveal the malleability of standards for admission for many critics of affirmative action, like supporters of Abigail Fisher's case. For many of these critics, their concerns are not so much about merit and consistency but rather about whom they view (whether consciously or unconsciously) as belonging and not belonging at selective institutions, about whom they presume as properly having a claim to seats at certain schools. In their book *Mismatch*, Sander and Taylor consistently argued for alterations to affirmative action that would push minority students (yet not the Abigail Fishers of the world) into less elite institutions. While doing so, the two authors presumed the neutrality of the approaches being used to teach students and never questioned the curriculum in any programs, despite the many questions being raised about the exclusivity in topics and the practicality of, and approaches to, education today. Sander and Taylor also imagined, through all their arguments about why less (or non-) selective schools are better options than selective schools for minority students, a nonexistent world in which an institution's resources play no role in a school's ability to offer programs that enable students to both survive and thrive within their hallways (but see Table 1 graduation rates).[189] Furthermore, Sander and Taylor pretended that students receive their education and important lessons only from the books and classroom learning, and not at all from interactions and other kinds of nonacademic resources and programs.[190] Yet, as Abigail Fisher herself once explained in

187. "[P]references on the scale used by UT are almost certain to backfire on the students they purport to help." SANDER & TAYLOR, *supra* note 3, at 289.

188. *Id.* at xiv.

189. For example, at the top thirty or so American private universities—members of the AAU—the endowment per alumni in 2012 was $54,959, compared to $5,852 at the approximately thirty public universities in the AAU and $6,710 at the University of California. *See Indicator 12.3.5*, U. CAL. ACCOUNTABILITY REP. 2013, http://accountability.universityofcalifornia.edu/index.php?in=12.3.5&source=uw.

190. For example, they state: "The general claim that boosting blacks and Hispanics up to more elite institutions is essential for their long-term success relies on outdated assumptions and falls apart on close examination." SANDER & TAYLOR, *supra* note 3, at 277. Then, in their discussion and comparison of law school mismatch to undergraduate mismatch, they argue that studies like Loury and Garman's "strongly suggest that the same thing is true for undergraduates: Performance trumps elite credentials." *Id.* at 278.

a newspaper interview, education is not only about academic performance; it is also about relationships, broadened experiences, and cultural capital. Speaking about what she believes she "lost" when the University of Texas at Austin denied her admission, Abigail Fisher proclaimed: "The only thing I missed out on was my post-graduation years. . . . Just being in a network of U.T. graduates would have been a really nice thing to be in. And I probably would have gotten a better job offer had I gone to U.T."[191]

We cannot help but notice the striking similarities between this quote by Abigail Fisher and parts of the Supreme Court's rationale in *Sweatt v. Painter*,[192] another decision that involved the University of Texas at Austin, only more than sixty-three years ago. In *Sweatt*, the Supreme Court responded to the legal challenge from Heman Marion Sweatt, a man whom the University's law school refused to admit because he was African American, against the University of Texas's policies of racial segregation.[193] Finding that the educational opportunities offered to black and white students at the University were not substantially equal, the Court held that the Equal Protection Clause of the Fourteenth Amendment entitled Sweatt to the law school admission he would have earned had he not been African American.[194] In so holding, the Supreme Court, like Abigail Fisher, highlighted many benefits of education, detailing how such benefits always extend beyond what professors lecture about in the classroom. The Court declared:

> Whether the University of Texas Law School is compared with the original or the new law school for Negroes, we cannot find substantial equality in the educational opportunities offered white and Negro law students by the State. In terms of number of the faculty, variety of courses and opportunity for specialization, size of the student body, scope of the library, availability of law review and similar activities, the University of Texas Law School is superior. *What is more important, the University of Texas Law School possesses to a far greater degree those qualities which are incapable of objective measurement but which make for greatness in a law school. Such qualities, to name but a few, include reputation of the faculty, experience of the administration, position and influence of the alumni, standing in the community, traditions and prestige.* It is difficult to believe that one who had a free choice between these law schools would consider the question close.

---

191. Adam Liptak, *Race and College Admissions, Facing a New Test by Justices*, N.Y. TIMES, Oct. 8, 2012, http://www.nytimes.com/2012/10/09/us/supreme-court-to-hear-case-on-affirmative-action.html?pagewanted=all.

192. 339 U.S. 629 (1950).

193. *Id.* at 631.

194. *Id.* at 633–36.

> Moreover, although the law is a highly learned profession, we are well aware that it is an intensely practical one. The law school, the proving ground for legal learning and practice, *cannot be effective in isolation from the individuals and institutions with which the law interacts. Few students and no one who has practiced law would choose to study in an academic vacuum, removed from the interplay of ideas and the exchange of views with which the law is concerned. The law school to which Texas is willing to admit petitioner excludes from its student body members of the racial groups which number 85% of the population of the State and include most of the lawyers, witnesses, jurors, judges and other officials with whom petitioner will inevitably be dealing when he becomes a member of the Texas Bar. With such a substantial and significant segment of society excluded, we cannot conclude that the education offered petitioner is substantially equal to that which he would receive if admitted to the University of Texas Law School.*[195]

Access via affirmative action to the leadership, educational, and career opportunities associated with attending the most selective and elite institutions in the United States matters for America's future; this observation applies to a variety of settings including the University of Texas School of Law, where the combined proportion of black and Latino J.D. students enrolled today (20.4%) is miles ahead of UC Berkeley Law (12.4%) and UCLA Law (11.6%);[196] science and engineering doctoral education, where nationally 62% of African Americans and 73% of Latinos earn their Ph.Ds. at universities with "very high" research profiles;[197] America's military academies and officer corps;[198] and undergraduate education, in light of all the graduation-rate and wage studies summarized in our Review.[199] Yet, in their book *Mismatch*, Sander and Taylor repeatedly discount, for minority students, these very kinds of interactive experiences and benefits of education, contending over and over that what truly matters are the mere books and classroom learning, and not the eliteness and resources of an institution and its alumni network. And, they

---

195. *Id.* at 633–34 (emphasis added).

196. LAW SCHOOL ADMISSIONS COUNCIL, 2014 ABA-LSAC OFFICIAL GUIDE TO ABA-APPROVED LAW SCHOOLS (2013), *available at* https://officialguide.lsac.org/release/officialguide_default.aspx (reporting 2012 JD enrollments).

197. *Women, Minorities, and Persons with Disabilities in Science and Engineering, Data Tables*, NAT'L SCI. FOUND. tbl.7-18, *available at* http://www.nsf.gov/statistics/wmpd/2013/pdf/tab7-18.pdf; *see also* Liliana M. Garces, *Understanding the Impact of Affirmative Action Bans in Different Graduate Fields of Study*, 50 AM. EDUC. RES. J. 251, 274 (2013) (summarizing data showing that affirmative action bans in four states were associated with a decline, controlling for other factors, of 26% in engineering and 19% in the natural sciences).

198. *See, e.g.*, Brief for the United States as Amicus Curiae supporting Respondents at 5–6, 10–15, Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411 (Aug. 13, 2012) (No. 11-345).

199. Regarding leadership specifically see, for example, BOWEN & BOK, supra note 11, at 160–75.

do so while also implicitly accepting that such benefits are worth it for mismatched white students. In the end, Sander and Taylor are right in one sense, at least. There is a mismatch, but the mismatch is not with the students of color they discuss in the book and the institutions that those students attend. Instead, it is in the cherry-picked data and flawed analyses that Sander and Taylor employ as support for their arguments and in the sad, sad fact that we still find ourselves trying to convince individuals such as Sander and Taylor to understand important points that the Supreme Court made very clearly sixty-four years ago in that other Texas decision, *Sweatt v. Painter*. Indeed, the rationale in *Sweatt* applies with similar force for white students today, who, without affirmative action, would be attending colleges and universities with substantially less interaction with huge segments of the rapidly growing and diversifying population within the United States.[200]

200. For an example of meta-analytic studies of the benefits of diversity where college diversity experiences are positively related to cognitive skills and development, see Nicholas A. Bowman, *College Diversity Experiences and Cognitive Development: A Meta-Analysis*, 80 REV. EDUC. RES. 4, 20 (2010). For a similar meta-analytic study where greater intergroup contact is associated with lower levels of prejudice, see Thomas F. Pettigrew & Linda R. Tropp, *A Meta-Analytic Test of Intergroup Contact Theory*, 90 J. PERSONALITY & SOC. PSYCHOL. 751, 766 (2006). And for a meta-analysis study that found cross-group friendships promote positive intergroup attitudes, see Kristin Davies et al., *Cross-Group Friendships and Intergroup Attitudes: A Meta-Analytic Review*, 15 PERSONALITY & SOC. PSYCHOL. REV. 332, 345 (2011).

As should be clear from our discussion of Figures 2A–C, *supra*, affirmative action is an important tool in protecting URM students from the educational harms of racial isolation, and our Review of graduation rates and earnings are leading to the point about affirmative action fostering the training of future minority leaders. Both of these concerns were acknowledged by the Court in *Grutter v.* Bollinger, 539 U.S. 306, 331–33 (2003). Thus, the above paragraph and footnote should not be misinterpreted as an argument that constitutionally permissible affirmative action results in benefits primarily for white students—our view is much broader than that.

* * *

Corruption, Governance, and Morality

CAPTURED BY EVIL: THE IDEA OF CORRUPTION IN LAW. By Laura S. Underkuffler. New Haven, Connecticut: Yale University Press, 2013. 344 pages. $55.00.

Reviewed by Edward L. Rubin*

Introduction

Corruption is a matter of serious concern throughout the world. It often determines the decisions of voters in democratic regimes and of insurgents in nondemocratic ones; it guides investment decisions by business firms; and it influences the scope and design of assistance programs, both governmental and nongovernmental, for developing nations. Many policy makers and scholars have advanced theories about why it occurs and what we can do to prevent it. Laura Underkuffler's new book[1] does something different: it explains why we are so concerned about the subject. Corruption, she argues, is regarded as the essence of evil, and to be corrupt is to succumb to the temptations of evil. This is not the way we typically think about issues of governance in the modern world, of course, and that is Underkuffler's point. Our attitude toward corruption is a holdover from an earlier mode of thought, she argues, a worldview that saw the legal and social rules guiding behavior in moralistic rather than instrumental terms. The Devil was a real presence in that world; walking the Earth, he subverted political regimes, fomented crime and other social misbehavior, took possession of individuals, and tempted them to sell their souls for material advantages. Having replaced the Devil with political science, sociology, and psychology, we regard material advantages as their own reward. But, Underkuffler argues, we continue to think about corruption in premodern terms, as a process of being "captured by evil."

This is an important book. It should be read by anyone who is interested in the problem of corruption, not because of the information it provides—it is not an empirical study—but because of the insight it offers into the way to think about such information and about the subject in general. For scholars, it offers the promise of what Rawls describes as reflective equilibrium, the revision of existing attitudes in light of principled reconsideration,[2] which in this case would be reconsideration of the role of

---

* University Professor of Law and Political Science, Vanderbilt University.

1. LAURA S. UNDERKUFFLER, CAPTURED BY EVIL: THE IDEA OF CORRUPTION IN LAW (2013).

2. JOHN RAWLS, A THEORY OF JUSTICE 48–50 (1971). Rawls describes this as being "reached after a person has weighed various proposed conceptions and he has either revised his

moralism in the contemporary legal treatment of corruption. For policy makers, it offers the promise of what Giddens describes as institutional reflexivity, the use of increased knowledge to revise an institution's objectives in a profound way that can alter that institution's basic role,[3] the institutions in this case being all those that collectively determine anticorruption policy. But Underkuffler's book also has value beyond the ambit of its admittedly important topic. It provides a case study of the way that preexisting cultural attitudes control current policies. Such attitudes are often so deeply embedded in our concepts, and in the language we use to describe those concepts, that it is only by directly confronting and reflecting on their prior use that we can decide whether we want to perpetuate their ongoing effects.[4]

Part I of this Review will summarize Underkuffler's arguments in *Captured by Evil*. Part II will then place these arguments in their historical context. While Underkuffler's theory is intrinsically historical, more can be said about the specific features of Western European and American history that support it. This Review will then proceed to consider the central issue that Underkuffler's book addresses—the relationship between corruption, as a political and legal issue, and the social morality that supports or undermines the effort to control it. Part III discusses Underkuffler's argument that the moralistic approach to corruption has deleterious effects on this effort, and Part IV discusses her argument that it is nonetheless necessary to maintain an attitude of moral condemnation toward corruption. I agree with both of Underkuffler's positions on these issues, but I argue in Part IV that we are disserved by relying on traditional morality to generalize our social sense of condemnation. Instead, we should recognize that a new morality has arisen in conjunction with the advent of administrative governance and that this morality provides a more coherent and pragmatic basis for condemning and combatting political corruption.

I. Underkuffler's *Captured By Evil*

Underkuffler begins her book by considering existing theories of political corruption. She divides them into three categories: shell theories,

---

judgments to accord with one of them or held fast to his initial convictions (and the corresponding conception)." *Id.* at 48. Furthermore, Rawls states that "[m]oral philosophy is Socratic: we may want to change our present considered judgments once their regulative principles are brought to light." *Id.* at 49.

3. ANTHONY GIDDENS, MODERNITY AND SELF-IDENTITY: SELF AND SOCIETY IN THE LATE MODERN AGE 20–21, 149–55 (1991). Giddens defines this as "the regularised use of knowledge about circumstances of social life as a constitutive element in its organisation and transformation." *Id.* at 20.

4. *See* EDWARD L. RUBIN, BEYOND CAMELOT: RETHINKING POLITICS AND LAW FOR THE MODERN STATE 6–7 (2005) (arguing that many of the basic concepts used to describe current governmental practices "bear the indelible imprint of the prior era when they took shape and control our current controversies in ways that we neither desire nor expect" (footnote omitted)).

substantive theories, and economic theories.[5] Shell theories identify corruption as wrongful on some independent, normative basis;[6] substantive theories identify evils that corruption itself engenders;[7] and economic theories apply that analysis to either approach.[8] She argues that none of these theories fully explains the meaning we attach to the term "corruption" in the political context,[9] and I think she is right.

The two leading shell theories she considers view corruption as either illegality or breach of duty,[10] but neither corresponds to the way we use the applicable terms. Many violations of law or obligation, however reprehensible, are not typically described as corrupt, even when perpetrated by a public official. If the head of an agency improperly denies benefits to eligible recipients, she is violating the law, and if she frequently fails to come into work as a result of being drunk, she is violating her duty, but we would not describe either as corrupt or attach the same level of opprobrium to her malefactions.[11]

The substantive theories that Underkuffler considers—betrayal, inequality, and abuse of power[12]—suffer from a similar defect. Betrayal is a much broader category than corruption. Although Underkuffler relies on more limited examples,[13] it can be plausibly said that Earl Warren and William Brennan betrayed President Eisenhower's trust by turning as liberal as they did[14] or that Clarence Thomas has betrayed his own ethnic group by being so conservative,[15] but no one would describe their behavior as corrupt. Inequality is even broader. It is certainly true that elected officials who engage in the paradigmatically corrupt practice of taking

---

5. UNDERKUFFLER, *supra* note 1, at 8–9.

6. *Id.* at 8.

7. *Id.*

8. *Id.* at 8–9.

9. *Id.* at 3–4.

10. *Id.* at 9, 14.

11. *See id.* at 9–20.

12. *Id.* at 20, 24, 37.

13. *See id.* at 22 (differentiating between garden-variety betrayals that could be characterized merely as "annoying" and other betrayals that warrant being labeled "corrupt").

14. *See* JIM NEWTON, JUSTICE FOR ALL: EARL WARREN AND THE NATION HE MADE 328, 354–55 (2006) (describing the events that lead to a tenuous relationship between Warren and Eisenhower and describing the President as "livid" about the Court's decisions during certain points of his administration); SETH STERN & STEPHEN WERMIEL, JUSTICE BRENNAN: LIBERAL CHAMPION 71–95 (2010) (suggesting Brennan was regarded by Eisenhower as a major mistake).

15. *See Justice Clarence Thomas, A Classic Example of an Affirmative Action Baby*, J. BLACKS HIGHER EDUC., Winter 1997–1998, at 35, 35 (noting that Thomas has been described as a traitor to his race); Stephen F. Smith, *The Truth About Clarence Thomas and the Need for New Black Leadership*, 12 REGENT U. L. REV. 513, 513 (2000) (reporting—with disproval—that "at the 1995 convention of the National Association for the Advancement of Colored People ('NAACP'), Justice Thomas was repeatedly called a 'pimp' and a 'traitor'"); Jack E. White, *Uncle Tom Justice*, TIME, June 26, 1995, at 36, 36 ("The maddening irony is that Thomas owes his seat to precisely the kind of racial preference he goes to such lengths to excoriate.").

bribes often grant the person who bribed them greater access and influence than others in return for the bribe, a clear case of inequality.[16] But these officials also grant greater access and influence to those who agree with them ideologically. This may be seen as partisan, but it is rarely regarded as corruption; the more common argument is that it is entirely proper for elected officials to give unequally favorable treatment to those groups in their constituency who supported them.[17] Responding to Franklin Zimring and David Johnson's substantive theory that corruption can be regarded as an abuse of power, and more specifically institutional power,[18] Underkuffler argues that the characterization is overinclusive. Public officials who mistreat their subordinates or use some punitive authority in a racially discriminatory way are certainly abusing their institutional power, but once again, we generally do not regard such misbehavior as a form of corruption.[19]

Economic analyses of corruption, such as Susan Rose-Ackerman's, generally define it as legally prohibited, rent-seeking behavior by public officials.[20] While such behavior can have beneficial effects, it is generally inefficient and should be prohibited on that ground.[21] In response, Underkuffler points out that the formulation depends on the concept of illegality and thus adds little to the shell theory she has already discussed.[22] In addition, it is once again too broad, a problem that would be exacerbated if the concept of illegality were omitted in favor of purely economic criteria.[23] Economists regularly condemn the ordinary process of economic regulation as rent-seeking behavior, after all, and while they are ready to argue that it is inefficient, it would be somewhat hysterical, even from their

16. *See* UNDERKUFFLER, *supra* note 1, at 24–25.

17. *Id.* at 27. Underkuffler also considers the substantive theory that corruption consists of betraying the public interest and rejects it on similar grounds. *Id.* at 35–37.

18. Franklin E. Zimring & David T. Johnson, *On the Comparative Study of Corruption*, 45 BRIT. J. CRIMINOLOGY 793, 799 (2005).

19. UNDERKUFFLER, *supra* note 1, at 38–40.

20. *See* SUSAN ROSE-ACKERMAN, CORRUPTION AND GOVERNMENT: CAUSES, CONSEQUENCES, AND REFORM 2 (1999) ("[T]he interaction between productive economic activity and unproductive rent seeking [can be elucidated] by focusing on the universal phenomenon of corruption in the public sector."). *See generally* Anne O. Krueger, *The Political Economy of the Rent-Seeking Society*, 64 AM. ECON. REV. 291 (1974) (exploring the distinction between legally acceptable rent-seeking and other forms of rent-seeking such as corruption and bribery); Johann Graf Lambsdorff, *Corruption and Rent-Seeking*, 113 PUB. CHOICE 97 (2002) (analyzing corruption from a standpoint of welfare economics); Kevin M. Murphy, Andrei Shleifer & Robert W. Vishny, *Why Is Rent-Seeking So Costly to Growth?*, 83 AM. ECON. REV. (PAPERS & PROC.) 409 (1993) (suggesting that rent-seeking by public officials is one of two reasons why rent-seeking in general impedes economic growth).

21. *See* ROSE-ACKERMAN, *supra* note 20, at 15–16, 21–26 (challenging the view espoused by economists that bribery can achieve efficiency).

22. UNDERKUFFLER, *supra* note 1, at 43–44.

23. *See id.* at 44–45 (questioning whether economic efficiency justifications, alone, can separate corrupt from noncorrupt behavior).

perspective, to treat this as corruption.[24] Moreover, as Underkuffler says, "[a]ctions by government actors can be incompetent, wasteful, market thwarting, or otherwise inefficient, and not be necessarily corrupt."[25]

The reason these theories all fail, and do so through the particular defect of establishing overinclusive categories, is that they do not recognize the essential concept of corruption that we employ in our society. Corruption, Underkuffler argues, is an explicitly moral notion: to say that certain people are corrupt is not simply to say that they have acted wrongly but that they are basically and intrinsically evil, that they are sinners.[26] Furthermore, corruption is regarded as an external force that lies in wait for all of us, and thus for all who are vested with public authority.[27] Those who are weak or immoral are "captured" by this evil and become moral pariahs.[28] As Underkuffler says, corruption "confers *a status*. A person, now corrupt, has changed. Evil has captured his being, his essence, his soul."[29]

Thus, unlike other crimes, which we regard as acts, corrupt behavior is regarded as a status, a revelation of the offending official's basic nature. That status is regarded as irrevocable and all-consuming; even those committed to rehabilitative theories of punishment want to permanently exclude corrupt officials from holding any further office.[30] This extreme and categorical approach to corruption, which contrasts with the modern attitude toward other crimes, results from the association of corruption with evil, a concept that modern society rarely invokes in other contexts.[31] It draws upon religious modes of thought that our secular society has generally rejected[32] and thus incorporates the supernatural, eschatological machinery of an earlier era.[33]

Is this concept of corruption, so obviously at odds with other aspects of

---

24. *See* GEORGE W. DOUGLAS & JAMES C. MILLER III, ECONOMIC REGULATION OF DOMESTIC AIR TRANSPORT: THEORY AND POLICY (1974); Sam Peltzman, *Toward a More General Theory of Regulation*, 19 J.L. & ECON. 211, 211 (1976); Richard A. Posner, *The Social Costs of Monopoly and Regulation*, 83 J. POL. ECON. 807, 818 (1975); George J. Stigler, *The Theory of Economic Regulation*, 2 BELL J. ECON. & MGMT. SCI. 3, 12 (1971).

25. UNDERKUFFLER, *supra* note 1, at 45.

26. *Id.* at 90, 104.

27. *Id.* at 64.

28. *See id.* at 81.

29. *Id.* at 69.

30. *See id.* at 80–81 (recognizing the pervasive nature of this desire).

31. *Id.* at 97–100.

32. *See* STEVE BRUCE, GOD IS DEAD 204 (2002); STEVE BRUCE, RELIGION IN THE MODERN WORLD 25–62 (1996); RONALD INGLEHART, CULTURE SHIFT IN ADVANCED INDUSTRIAL SOCIETY 186–87 (1990); THOMAS LUCKMANN, THE INVISIBLE RELIGION 36–37 (1967); DAVID MARTIN, A GENERAL THEORY OF SECULARIZATION (1978); CHARLES TAYLOR, A SECULAR AGE 1–3 (2007).

33. *See* UNDERKUFFLER, *supra* note 1, at 58–60 (tracing, throughout Western history, the depiction of corruption as evil).

our legal system, desirable or not, Underkuffler asks. The great problem with it is that it generates a high-pitched emotionalism to the legal, as well as the public, response to charges of corruption.[34] While modern scholars often point out that the role of emotion in law should not be ignored, their purpose in doing so is often to warn us that it should not necessarily be indulged.[35] In the case of prosecutions for corruption, it leads to the introduction of character evidence that would be excluded in other criminal prosecutions, an all-or-nothing approach that contrasts with the typical grading of offenses, and a reliance on impressionistic, standardless decision making.[36] On the other hand, Underkuffler notes, a moralistic attitude toward corruption may be a proper and necessary response to the damage that it does to the political system in its entirety and the universality of its appeal.[37] Economic arguments that corruption can be beneficial under certain circumstances only confirm, for Underkuffler, the importance of articulating general and definitive condemnations of this offense.[38]

The categorical, Manichean approach to corruption that prevails in the United States and other Western nations helps explain the reason they differ from many developing nations. In these nations, Underkuffler suggests, there is a culture of corruption, a sense that corrupt behavior, while recognized as wrong, is nonetheless acceptable because "everybody does it."[39] Such a culture often arises in developing nations because the self-interest encouraged by an emerging market economy undermines the preexisting moral ethos of the nation without having generated a countervailing morality of governmental honesty.[40] In an effort to balance the disadvantages of treating corruption as capture by evil with the moral force that this conception deploys, Underkuffler concludes that this conception should not be used in the legal system but that it is a valuable part of a civil society's belief system.[41]

---

34. *See id.* at 107–13 (discussing how even the legal terms associated with corruption engender emotional responses, which are potentially dangerous in their influence).

35. *See, e.g.*, Susan Bandes, *Empathy, Narrative, and Victim Impact Statements*, 63 U. CHI. L. REV. 361, 365 (1996) (noting that the narrative deployment of emotion is not always helpful in a legal context); Terry A. Maroney, *Emotional Regulation and Judicial Behavior*, 99 CALIF. L. REV. 1485, 1502 (2011) (warning that the value of emotion "does not signify that it must be allowed free rein"); Toni M. Massaro, *Show (Some) Emotions*, *in* THE PASSIONS OF LAW 80, 104 (Susan A. Bandes ed., 1999) (acknowledging the need for emotion yet stressing the importance of its measured use). *But see* Terry A. Maroney, *Angry Judges*, 65 VAND. L. REV. 1207 (2012) (arguing that judicial anger serves a valid purpose, but only when harnessed in the cause of righteousness).

36. *See* UNDERKUFFLER, *supra* note 1, at 121–28.

37. *See id.* at 54–56 (noting that the idea of corruption is an "explicitly moral notion" and that its moral nature actually helps clarify the term as used in law).

38. *Id.* at 129–40.

39. *See id.* at 131–32.

40. *Id.* at 231–36.

41. *Id.* at 248–51.

II. Underkuffler's Theory in Historical Context

As suggested above, *Captured by Evil* is not a book about the legal doctrine, economic analysis, or public policy of corruption. It is about the concept of corruption. Its thoughtful and, in my view, convincing theory explains the source of our current attitudes about the subject and the way that those attitudes affect, and in some cases distort, legal doctrine, economic analysis, and public policy in this area. In addition, its theory, which is that our current attitudes derive from a premodern conceptual framework, makes the book a case study of the way that the past controls the present, particularly in the conceptual realm. It is fairly easy to see that a dirt road needs to be paved, somewhat less easy to see that an old building needs to be redesigned, but often quite difficult to recognize that familiar concepts need to be rethought.

Corruption is far from the only concept in our legal system that embeds the attitudes of prior times in ways that are not necessarily applicable to current circumstances. Words can, of course, be redefined, but Humpty Dumpty notwithstanding,[42] they are not as malleable as the definition section of a statute might suggest. As Derrida observed, they derive their meaning from a complex linguistic system, generated by society and beyond the direct control of any individual writer.[43] Thus, when we use terms that evolved in earlier times, we may well get a good deal more than we bargained for. In some cases, the baggage that the word brings with it can be readily discarded; the modern chemical theory of the elements is robust enough to be uninfluenced by the term's original reference to water, air, earth, and fire. In other cases, the continuities are sufficient to overcome the cultural associations; despite vast changes in its prevalence, pragmatic use, and cultural significance, a horse is basically the same thing now as it was in the Middle Ages. The question, therefore, is whether there have been intervening social changes that render a term's continued usage problematic and whether those changes have been so extensive that they overcome the problem and place contemporary users in control.

These considerations suggest that Underkuffler's argument is best evaluated by situating it in its historical context. Underkuffler hints at this

---

42. Mr. Dumpty explains as follows:

> "When *I* use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean neither more nor less."
>
> "The question is," said Alice, "whether you *can* make words mean so many different things."
>
> "The question is," said Humpty Dumpty, "which is to be master——that's all."

LEWIS CARROLL, ALICE'S ADVENTURES IN WONDERLAND & THROUGH THE LOOKING-GLASS 178 (Bantam Classics reissue ed. 2006) (1871).

43. *See* JACQUES DERRIDA, OF GRAMMATOLOGY 101–07 (Gayatri Chakravorty Spivak trans., Johns Hopkins Univ. Press 1976) (1967).

issue,[44] but there is somewhat more that can be said about it. My own view, which I have elaborated elsewhere with respect to a range of issues,[45] is that the event that casts doubt on many of the terms that we use to describe political and legal issues is the advent of the administrative state, which I date to the end of the eighteenth and beginning of the nineteenth centuries.[46] Terms such as legitimacy, power, discretion, law, and property are all holdovers from a prior period and embed concepts from that period that are largely inapplicable to modern government.[47] Legitimacy, for example, referred to whether the monarch's biological progeny were the products of a Christian marriage and could thus succeed to the throne.[48] The dichotomous standard it implies is inaccurate in the modern context, where the real question is not whether a law will be recognized as an act of the established government but what level of compliance it will achieve in practice.[49] The term law was used in the premodern era to describe a system of rules that fit together in a coherent ("law-like") manner.[50] In a

---

44. *See* UNDERKUFFLER, *supra* note 1, at 1–2 (discussing corruption as a concept that has historical roots and has been subject to repeated attempts to redefine it over the course of time).

45. RUBIN, *supra* note 4, at 22–25.

46. *See id.* at 29. The process was somewhat delayed in the case of the American federal government. *See infra* notes 82–85 and accompanying text.

47. *See* RUBIN, *supra* note 4, at 74–91 (power and discretion); *id.* at 144–60 (legitimacy); *id.* at 191–203 (law); *id.* at 227–37 (legal rights); *id.* at 260–68 (human rights); *id.* at 296–308 (property).

48. *Id.* at 144–45; *see also* 2 MARC BLOCH, FEUDAL SOCIETY 383–89 (L.A. Manyon trans., 1961) (describing the centrality of legitimate inheritance as a principle of succession and also describing the way that it was integrated with elective mechanisms); ERNST H. KANTOROWICZ, THE KING'S TWO BODIES: A STUDY IN MEDIAEVAL POLITICAL THEOLOGY 330–34 (1957) (describing the way legitimate inheritance replaced coronation as the theoretical basis for royal succession); FRITZ KERN, KINGSHIP AND LAW IN THE MIDDLE AGES 13–14 (S.B. Chrimes trans., 1956) (discussing how the shift from election to hereditary right was probably based on a shift in thinking to the belief that the divine right of a ruler can be inherited by members of his bloodline); Jean Dunbabin, *Government*, *in* THE CAMBRIDGE HISTORY OF MEDIEVAL POLITICAL THOUGHT, C. 350–C. 1450, at 477, 496–98 (J.H. Burns ed., 1988) [hereinafter CAMBRIDGE HISTORY OF MEDIEVAL POLITICAL THOUGHT] (describing the centrality of legitimate inheritance as a pragmatic approach to questions of succession). The issue was taken seriously throughout the premodern period. For a description of two titanic battles between monarchs who wanted an heir and the Catholic Church, which insisted on the principle of legitimacy, see FRANCES & JOSEPH GIES, MARRIAGE AND THE FAMILY IN THE MIDDLE AGES 88–94 (1987), dealing with the conflict over the second marriage of Lothair II, king of Lotharingia, or Lorraine, which determined the current political map of Europe, and DIARMAID MACCULLOCH, REFORMATION: EUROPE'S HOUSE DIVIDED, 1490–1700, at 198–204 (2003), dealing with the conflict over the second marriage of Henry VIII, which resulted in England becoming a Protestant country.

49. RUBIN, *supra* note 4, at 160–61.

50. *Id.* at 193; *see also* OTTO GIERKE, POLITICAL THEORIES OF THE MIDDLE AGE 75–80 (Frederic William Maitland trans., 1900) (describing the evolution of natural law into a highly structured system of rules); K. Pennington, *Law, Legislative Authority, and Theories of Government 1150–1300*, *in* CAMBRIDGE HISTORY OF MEDIEVAL POLITICAL THOUGHT, *supra* note 48, at 424, 424–30 (tracing the evolution of law and the reconciliation of three modes of thought into a coherent system based on reason). For the classic statement of the connection between law and reason, see 1 ST. THOMAS AQUINAS, SUMMA THEOLOGICA pt. I-II q. 90, art. 4,

modern administrative state, the relevant question about legislative enactments is not whether or not they fit together in this manner but whether each one, considered separately, represents a democratically established policy and how effectively that policy will be implemented by the administrative agency to which it is assigned.[51]

Underkuffler's theory acquires additional explanatory force when placed within this context. One question that can be asked about it is why political corruption was so closely associated with pure evil during the premodern period; why, as Underkuffler suggests, it acquired the aura of at least semi-Satanic possession.[52] To be sure, this was an era that tended to delineate its moral judgments with the intense coloration of sin and salvation. But it was nonetheless necessary to decide which issues would be subject to this supernatural moralism. Despite its religiosity, premodern society adopted a genially forgiving attitude to many defalcations that we now view with concern, such as marital rape, the rape of civilians by an occupying army, sexual harassment of female workers, the physical abuse of children, assault, and even, in some cases, murder, when committed by a member of the elite against slaves, serfs, or other members of the lower classes.[53] It took many other offenses seriously, of course, and regarded some, such as blasphemy or suicide, as equally associated with an evil nature and hence unforgivable, but most of these had direct and apparent connections to religious issues.[54] Political corruption is, by definition, a

---

at 995 (Fathers of the English Dominican Province trans., 1947); *id.* q. 95, art. 3, at 997–98; and *id.* q. 95, art. 2, at 1014–15.

51. *See* RUBIN, *supra* note 4, at 205.

52. *See* UNDERKUFFLER, *supra* note 1, at 64–67.

53. *See* PHILIPPE ARIÈS, CENTURIES OF CHILDHOOD: A SOCIAL HISTORY OF FAMILY LIFE 252–68 (Robert Baldick trans., Vintage Books 1962) (1960) (corporal punishment of children); JAMES A. BRUNDAGE, LAW, SEX, AND CHRISTIAN SOCIETY IN MEDIEVAL EUROPE 297–313, 444–53 (1987) (concubinage, prostitution, and marital rape); EUGENE D. GENOVESE, ROLL, JORDAN, ROLL: THE WORLD THE SLAVES MADE 413–31 (1974) (master's sexual relations with slaves); BARBARA A. HANAWALT, GROWING UP IN MEDIEVAL LONDON: THE EXPERIENCE OF CHILDHOOD IN HISTORY 185–87 (1993) (master's physical and sexual abuse of servants); Pamela DeLargy, *Sexual Violence and Women's Health in War*, *in* WOMEN AND WARS 54, 55 (Carol Cohn ed., 2013) (sexual violence in war). Because the Catholic Church declared that twelve was the legal age of consent, sexual relations between a grown man and a twelve-year-old girl were not even considered an abuse in the premodern era. *See* BRUNDAGE, *supra*, at 357. When a grown man had sex with an even younger girl to whom he was engaged, Pope Alexander III declared that it was quite alright, as long as the man followed through with the marriage. *See id.* at 335.

54. *See* 1 DANTE ALIGHIERI, THE DIVINE COMEDY: INFERNO 65–69 (John Ciardi trans., 1961) (condemning the suicidal to the seventh ring of hell); *id.* at 71–75 (condemning the blasphemers—those violent against God—to the seventh ring as well); 2 AQUINAS, *supra* note 50, pt. II-II q. 13, art. 3, at 1231–32 (treating blasphemy as a more serious sin than murder because it is a sin committed against God); *id.* q. 64, art. 5, at 1468–70 (deeming it unlawful to commit suicide). These prohibitions were taken seriously. *See* 2 ALEXANDER MURRAY, SUICIDE IN THE MIDDLE AGES: THE CURSE ON SELF-MURDER 33–76 (2000) (describing the practice of "executing" or otherwise mutilating the body of a person who had committed suicide); DAVID

secular act; thus, the question about why it was treated in such starkly moralistic terms remains.

The nonadministrative nature of the premodern state provides an answer. Prior to the development of bureaucratic government, the public and the private realms were intricately intertwined. This was most dramatically and completely true for the monarch and constitutes the essential feature of what Weber described as a patrimonial state.[55] In a few cases, such as William the Conqueror's England, the entire realm was literally regarded as the monarch's personal property.[56] In most cases, the royal property or domain was limited, albeit still extensive, and the king was regarded as the owner of the rest as overlord of subordinate aristocrats who in turn controlled their territory as private property and owed their royal overlord their personal loyalty.[57] It was understood, and well-accepted, that the monarch could use money raised through taxation and fees—public money, according to our current view—for his personal expenses.[58] What we now regard as a classic case of corruption, appropriating money from the public fisc to build or improve one's personal residence, was the universal practice for premodern European monarchs. It was not only regarded as acceptable, but near obligatory; people believed, quite consciously, that the magnificence of the monarch's personal residence and equipage was necessary to the maintenance of his authority and the stability of the regime.[59]

NASH, BLASPHEMY IN THE CHRISTIAN WORLD: A HISTORY 189 (2007) (discussing how blasphemers suffered physical punishment).

55. A patriomonial state, as Weber defines it, is one where the entire polity is viewed as the monarch's household (*oikos*). 2 MAX WEBER, ECONOMY AND SOCIETY 1010–15 (Guenther Roth & Claus Wittich eds., Ephraim Fischoff et al. trans., Univ. of Cal. Press 1978) (1922).

56. *See* R.H.C. DAVIS, A HISTORY OF MEDIEVAL EUROPE: FROM CONSTANTINE TO SAINT LOUIS 284–85 (2d ed. 1988) (describing William the Conqueror's procedure for claiming ownership of nearly the entire kingdom); DAVID C. DOUGLAS, WILLIAM THE CONQUEROR: THE NORMAN IMPACT UPON ENGLAND 289–96 (1967) (detailing William the Conqueror's consolidation of control over the realm, including the end of the earldom system in favor of increased monarchial control).

57. *See, e.g.*, 1 BLOCH, *supra* note 48, at 145–75 (describing systems of French vassalage and fiefdom in the medieval era, which centered on loyalty to the king); F.L. GANSHOF, FEUDALISM 69–155 (Philip Grierson trans., Harper Torchbooks 3d rev. ed. 1964) (describing the personal nature of vassals and the operation of the the fief system).

58. *See* S.B. CHRIMES, AN INTRODUCTION TO THE ADMINISTRATIVE HISTORY OF MEDIAEVAL ENGLAND 135–47 (3d ed. 1966) (tracing the evolution of the king's "wardrobe" from the place to store the king's treasure headed by a person who was attendant to the king's needs into the financial administration of the king's for the entire realm); JOSEPH R. STRAYER, ON THE MEDIEVAL ORIGINS OF THE MODERN STATE 61–62 (Princeton Classic ed. 2005) (1970) (discussing the process of collecting revenue to expand the king's domain and power); 1 T.F. TOUT, CHAPTERS IN THE ADMINISTRATIVE HISTORY OF MEDIAEVAL ENGLAND: THE WARDROBE, THE CHAMBER AND THE SMALL SEALS 179 (rev. ed. 1937) (stating that the wardrobe—the financial department that controlled the household accounts of the king—"depended on the exchequer for a large part of its income"); W.L. WARREN, THE GOVERNANCE OF NORMAN AND ANGEVIN ENGLAND, 1086-1272, at 73–78 (1987) (discussing the king's

The same imbrication of the public and private realms occurred with respect to appointed rather than hereditary offices of government. It was typical in the premodern era, as Weber famously noted, to treat public offices as private possessions.[60] That is, a person obtained an office and then kept the revenues that the office generated, with the office itself becoming one's personal, often heritable, property.[61] This obvious extension of feudal land grants to the operational tasks of Early Modern monarchies was used for raising taxes—significantly called tax farming—recruiting troops, and regulating various areas of the economy.[62] Unlike feudal fiefdoms, however, these offices were sold in many cases, most notoriously in France, and the resulting revenue was a major source of funding for the monarchy.[63]

When offices were appointive rather than purchased, they were typically given to members of the elite through a system of personal connections and relations. Because character, rather than credentials, was typically regarded as the essential qualification, there could be no higher recommendation than to say that someone is "my dear and trusted friend."[64] Although Early Modern regimes were rapidly solidifying their control over their territory, there remained enough local autonomy and social agitation to make personal loyalty to the existing ruler an essential qualification. Even the Early Modern absolute monarch par excellence, Louis XIV, spent a

---

financing practices and suggesting that, at points, his coffers were overflowing with funds obtained from taxation).

59. *See* G.R.R. TREASURE, SEVENTEENTH CENTURY FRANCE 286–87 (Anchor Books 1967) (1966); MALCOLM VALE, THE PRINCELY COURT 169 (2001). This was true for leading nobles as well. HEINRICH FICHTENAU, LIVING IN THE TENTH CENTURY 50–51 (Patrick J. Geary trans., Univ. of Chi. Press 1991) (1984).

60. *See* WEBER, *supra* note 55, at 1028–29 (describing the king's patrimonial officialdom as lacking any distinction between personal and official).

61. When the Spanish Armada was approaching England, Sir Francis Drake, in command of his own ship and three others, captured a crippled Spanish flagship, the *Nuestra Señora del Rosario*. GARRETT MATTINGLY, THE ARMADA 283, 293 (1959). Drake immediately ordered one of his other ships to take the captured vessel into port as his personal prize. *Id.* at 293–94. As Mattingly comments: "No one . . . seems to have blamed Drake's behavior in this extraordinary episode. No one . . . spoke of it with the slightest disparagement except Martin Frobisher, and his quarrel was rather over the division of the *Rosario's* spoils than over the manner of their acquisition." *Id.*

62. *See* THOMAS ERTMAN, BIRTH OF THE LEVIATHAN 76 (1997); DENYS HAY, EUROPE IN THE FOURTEENTH AND FIFTEENTH CENTURIES 100–11 (1966); WEBER, *supra* note 55, at 965–66.

63. *See* JAMES B. COLLINS, THE STATE IN EARLY MODERN FRANCE 50–51 (1995) (describing the practice of forced annual loans from officers and the resulting boon to revenue in France); ERTMAN, *supra* note 62, at 17, 100–01 (describing the "feudalized" practice of selling offices and the fact that laws to the contrary were disregarded in places like France); TREASURE, *supra* note 59, at 53–55.

64. *See* KEITH THOMAS, THE ENDS OF LIFE: ROADS TO FULFILMENT IN EARLY MODERN ENGLAND 192 (2009); *see also* EVA ÖSTERBERG, FRIENDSHIP AND LOVE, ETHICS AND POLITICS 50–60, 74–81 (2010) (noting the importance of friendship in developing countries and how it overlaps with cronyism).

significant portion of his youth being chased out of Paris by the rebellion of aristocrats known as the Fronde, an experience he of course never forgot.[65]

In these circumstances, the concept of corruption could not be defined as the private use of public funds. To say that there was no distinction between the public and the private realm would be incorrect, but there was obviously no clear separation of the two, no defined boundary that public officials were forbidden to cross. The charge of corruption, therefore, could only be leveled against the most extreme behavior, the acquisition of such excessive wealth that it passed beyond the boundaries of the personal enrichment that often accompanied a purchased or appointed office.[66] As such, it was understood as a betrayal of the monarch, in effect an act of treason.[67] Thus, although the overlap of public and private functions in some sense reiterated the patrimonial character of the monarchy, the pattern had its limits. Because the king was in effect the owner of the entire polity, subordinate officials who purchased their offices, and were permitted to enrich themselves at public expense, intruded on the monarch's patrimonial status if they violated some vaguely defined sense of proportion.

Thus, dismissing someone for the use of public funds for private gain in premodern times was not analogous to dismissing a person for that reason today but more similar, instead, to dismissing a civil servant for incompetence. We accept the fact that everyone makes mistakes sometimes and, in fact, that there is a wide range of acceptable performance. In the absence of any clearly defined boundaries, it is only colossal incompetence, something that would need to be described in more emotive terms, that would lead to a dismissal on these grounds.

There were a few exceptions. The English legal system had evolved to the point that a judge was expected to be neutral and forbidden to accept monetary gifts from litigants. Underkuffler refers to the well-known case of Sir Francis Bacon, dismissed from his post as Lord Chancellor of England after he admitted to a parliamentary investigation that he had accepted bribes in his capacity as the presiding judge of the High Court of

65. *See* COLLINS, *supra* note 63, at 65–76; DAVID OGG, EUROPE IN THE SEVENTEENTH CENTURY 213–14 (8th ed. 1960); TREASURE, *supra* note 59, at 200–13. At the same time, England experienced a full-scale revolution, complete with the decapitation of the king, and there were disturbances of various sorts in other European regimes. *See* H.R. Trevor-Roper, *The General Crisis of the 17th Century*, 16 PAST & PRESENT 31, 31–34 (1959).

66. *See* Maryvonne Génaux, *Early Modern Corruption in English and French Fields of Vision*, *in* POLITICAL CORRUPTION: CONCEPTS AND CONTEXTS 107, 110 (Arnold J. Heidenheimer & Michael Johnston eds., 3d ed. 2002) (arguing that corruption must extend beyond the social norms of the day); Koenraad W. Swart, *The Sale of Public Offices*, *in* POLITICAL CORRUPTION: CONCEPTS AND CONTEXTS, *supra*, at 95, 95 (noting that public offices were regarded as property and that their sale was routine).

67. *See* Génaux, *supra* note 66 (noting how the "failure to acquit a personal commitment went through the lexicon of treason").

Chancery.[68] As she points out, his defense that he was unaffected by the gifts, and always decided cases fairly, proved unavailing.[69] But even in these circumstances, some caution is required before generalizing. The impeachment of Bacon was highly political; it was spearheaded by the formidable Edward Coke and involved the controversy about the boundaries of royal authority that would lead to the independence of the English judiciary, the Civil War, and the execution of the king in the next generation.[70] Thus, Bacon's impeachment may be comparable to the impeachment of Bill Clinton, who was charged with a recognizable offense but not one that would necessarily lead to condemnation in our present moral system had the accusation not served obvious political purposes.

This historical contextualization of Underkuffler's theory provides additional support for her claim that that the emotive approach to political corruption does not fit with our contemporary legal system. Underkuffler's argument, as summarized above, is that treating political corruption as a character-revealing sin, as allowing oneself to be "captured by evil," violates the modern notions that we punish people for their actions, not their characters; that we do so according to clearly defined behavioral standards; and that we vary the amount of punishment according to the severity of the offense. The additional argument is that we no longer need any of the moral machinery of condemnation because we now have clearly defined legal standards.[71] Patrimonial monarchs are completely gone from Western nations and their successors relegated to the gossip pages. The elected officials who have replaced them are generally regarded as public servants, with no claim to the resources of the state beyond their salaries and expenses.[72] Below them, and more relevant to the topic of corruption, the advent of the administrative government has separated the previously intertwining realms of public and private behavior, not in any theoretically justifiable way, perhaps, but in a pragmatically manageable one.

Weber's classic account of bureaucracy can be read as an instruction manual for the legal separation of the public and private realms and the consequent definition of corruption as a legal rather than a moral category. Offices are no longer personal possessions but predefined positions that

68. UNDERKUFFLER, *supra* note 1, at 78–79.

69. *Id.* at 79.

70. *See* CATHERINE DRINKER BOWEN, FRANCIS BACON: THE TEMPER OF A MAN 177–204 (1963) (describing the political backdrop in which Bacon was impeached); LOREN EISELEY, THE MAN WHO SAW THROUGH TIME 44–47 (1973) (showing that Bacon's downfall can be attributed to political forces). For the proposition that the behavior for which Bacon was impeached was typical for officeholders at the time, see BOWEN, *supra*, at 179–80. For an imaginative, novelistic account of Bacon's public life, see DAPHNE DU MAURIER, THE WINDING STAIR: FRANCIS BACON, HIS RISE AND FALL (1976).

71. *See* UNDERKUFFLER, *supra* note 1, at 249–50.

72. For the U.S. general rule, see IDA A. BRUDNICK, CONG. RESEARCH SERV., RL30064, CONGRESSIONAL SALARIES AND ALLOWANCES 1–2 (2013).

appointed individuals fill in succession.[73] The economic benefits that these individuals receive are no longer the income that the position generates but a fixed salary and a pension on retirement.[74] They are no longer chosen because of their social status or their personal relations with higher officials but rather on the impersonal basis of socially defined qualifications, an examination, or a combination of the two.[75] Their responsibilities, salaries, and selections are all monitored and recorded by permanent, carefully maintained, and readily examined files.[76]

The United States fully conforms to this general pattern, although it was a bit delayed in reaching it. Western European nations entered the modern era with centralized royal regimes already in place so that the transition to administrative government involved reorganizing, redefining, and expanding an established structure.[77] The practice of treating public offices as private property, for example, was ended by restructuring existing positions.[78] This is not to say that the process was an easy one—in France it required a revolution,[79] in Austria it was implemented by a benevolent despot,[80] and in Britain, which had neither, the offices often had to be bought out from their owner, in some cases for enormous sums.[81] But however expensive or traumatic this shift to bureaucratic governance may have been, it can be plausibly described as a reform. The United States lacked any premodern inheritance of this sort, having been formed from thirteen colonies that were separately administered by Britain. With no unified colonial government in North America, it was required to concoct its central government from scratch as it achieved independence and simultaneously entered the administrative era. The result was a rather

---

73. *See* WEBER, *supra* note 55, at 962.

74. *See id.* at 958–59 ("[E]ntrance into an office . . . is considered an acceptance of a specific duty of fealty to the purpose of the office . . . in return for the grant of a secure existence.").

75. *See id.* at 958–62.

76. *See id.* at 957.

77. *See* ERNEST BARKER, THE DEVELOPMENT OF PUBLIC SERVICES IN WESTERN EUROPE, 1660–1930, at 2 (Oxford Univ. Press ed. 1966) (1944) (observing how the philosopher–king movement of the eighteenth century compelled existing monarchial sovereigns to reform and "rationalize" existing administrative systems).

78. *See id.* at 8–10 (noting that, in France, although the proprietary offices of the nobility persisted, the offices of the *intendants*, comprised of members of the legal profession, developed so as to supersede the local authority of these hereditary positions).

79. *Id.* at 12.

80. *See* T.C.W. BLANNING, JOSEPH II 65–67, 92 (1994) (tracing how the Hapsburg Emperor Joseph II abolished preexisting privileges and customs to establish a rationalized, bureaucratic state).

81. *See* BARKER, *supra* note 77, at 32 (describing the English system of offices in the seventeenth and eighteenth centuries as one in which offices were bought and sold for consideration, essentially as investments); NORMAN CHESTER, THE ENGLISH ADMINISTRATIVE SYSTEM, 1780–1870, at 18–20 (1981) (showing that in order to eliminate a wasteful office, the English system had to worry about violating property rights and therefore offices could not be eliminated without paying substantial sums to the holder).

diminutive central government—although not quite as diminutive as previously thought[82]—that was staffed by a narrow, quasi-aristocratic elite. Andrew Jackson, the first president from the western part of the nation, replaced this with a political patronage system based on the premodern principle of loyalty.[83] Having thus regressed, the United States then had to work its way back to the Weberian rationality that European nations had already achieved. The first success of this effort was the Pendleton Civil Service Reform Act of 1883,[84] but the process took at least another fifty years.[85]

In both the United States and Europe, the resulting separation of public service from private property made possible, and was in turn facilitated by, a well-developed set of rules. Corruption could now be defined as receiving any money from one's exercise of office from any source other than one's official employer or as hiring anyone for an administrative position on any basis other than the meritocratic criteria established for the position. There are of course ambiguities in these rules, and borderline cases that demand an exercise of judgment, but that is true for any set of legal rules. The basic point is simply that these are legal rules; they can be stated and enforced without the need to invoke notions of evil or to draw upon the emotive force of moral condemnation. In other words, because we are now able to define illegality, we can dispense with sin. The idea that corruption consists of being captured by evil not only violates our ways of defining crimes, as Underkuffler suggests, but also our ways of organizing and controlling our public officials.

III. Getting Morality out of Political Corruption

The aspects of corruption considered thus far can be properly described as pragmatic. They involve the way the crime of corruption is defined so that people who fall within the defined category can be

82. *See* JERRY L. MASHAW, CREATING THE ADMINISTRATIVE CONSTITUTION: THE LOST ONE HUNDRED YEARS OF AMERICAN ADMINISTRATIVE LAW 11 (2012) (analyzing former views among academics and constitutional scholars that the Constitution failed to recognize or vest the government with direct authority for establishing administrative powers); *id.* at 34–35 (describing how Congress established the Departments of War, Foreign Affairs, and the Treasury, as well as the Post Office, national bank, and federal judiciary, all within ten years of the Constitution's effective date).

83. *See* DANIEL WALKER HOWE, WHAT HATH GOD WROUGHT: THE TRANSFORMATION OF AMERICA, 1815–1848, at 331–34 (2007); MASHAW, *supra* note 82, at 175–78; GLYNDON G. VAN DEUSEN, THE JACKSONIAN ERA, 1828–1848, at 31–37 (1959).

84. *See* Pendleton Civil Service Reform Act, ch. 27, 22 Stat. 403 (1883) (codified as amended at 5 U.S.C. § 3304 (2012)) (establishing a merit system and mandating, among other requirements, that all officers and clerks for both new and preexisting positions must pass an examination that tests their "relative capacity and fitness" and that no more than two members of a single family can hold positions covered by the Act).

85. *See* RONALD N. JOHNSON & GARY D. LIBECAP, THE FEDERAL CIVIL SERVICE SYSTEM AND THE PROBLEM OF BUREAUCRACY 12–13 (1994).

criminally prosecuted. Underkuffler's conclusion, which is that this pragmatic definition is anachronistically moralistic because it clashes with our current ways of defining crime, can be amplified by an historically conceptualized analysis, as suggested above, which suggests that the pragmatic definition's moralism is also anachronistic because it was devised for a premodern state where public and private functions overlapped. But an historically contextualized analysis of corruption also reveals a more conceptual way in which the moralistic definition of corruption is out of date. This further consideration supports Underkuffler's conclusion as well, but I think it also suggests a modification of her theory.

To say that someone who is corrupt has been captured by evil suggests that people have an essential nature of some sort, a core identity that is subject to a plenary characterization. By itself, this does not exactly correspond to the Christian concept of a soul, since there is no necessary implication of immortality. But it is certainly a necessary component of the Christian concept, which assumes a unitary essence on which final judgment can be rendered and does not include the idea that a multifaceted individual might spend part of the afterlife in Heaven and part in Hell. The idea, then, is that becoming corrupt is equivalent to selling one's soul to Satan, an abandonment of one's basic status as a moral being.

As Underkuffler points out, there is an essential ambiguity about the idea of being captured by evil.[86] On the one hand, it suggests that certain individuals are evil at their essential core.[87] On the other hand, it suggests that evil is a powerful temptation that lies in wait for all of us and can capture anyone at a moment of weakness or vulnerability.[88] She never quite resolves this ambiguity, and it may be unresolvable. Quite possibly, however, it acknowledges the existence of free will (and is thus Catholic[89] or Arminian,[90] rather than Lutheran[91] or Calvinist[92]). The idea is that we are

86. UNDERKUFFLER, *supra* note 1, at 4–6.

87. *See id.* at 74 ("[C]orruption is a *dispositional concept*, which describes the deep, moral character of the accused.").

88. *See id.* at 64 ("[C]orruption is portrayed as an *external* evil, which attacks and undermines better human impulses. It is, furthermore . . . something to which we are all potentially susceptible . . . ." (footnote omitted)).

89. *See* AQUINAS, *supra* note 50, pt. I q. 83, art. 1, at 417–18; Erasmus, *A Discussion or Discourse Concerning Free Will (1524)*, *in* ERASMUS AND LUTHER: THE BATTLE OVER FREE WILL 1, 29 (Clarence H. Miller ed., Clarence H. Miller & Peter Macardle trans., 2012); Erasmus, *The Shield-Bearer Defending: A Discussion Part 1 (1526)*, *in* ERASMUS AND LUTHER: THE BATTLE OVER FREE WILL, *supra*, at 127, 128.

90. *See* W. STEPHEN GUNTER, ARMINIUS AND HIS DECLARATION OF SENTIMENTS 140 (2012) (outlining Arminius's theological position on free will, which features the doctrine that humans are endowed with a free will directed by divine grace); MACCULLOCH, *supra* note 48, at 375 (suggesting that Arminius modified the traditional Calvinist doctrine by adding that people are subject to damnation by their own fault). Wesleyanism, and modern American evangelism in general, is derived from Arminian theology. *See id.* at 699–701.

all capable of becoming evil, at our essential core, but that we must choose to do so. What makes corruption different from other defalcations, according to Underkuffler, is that we view it as a total surrender to evil rather than a single, separable action that can be repented.[93]

This totalistic conception of human nature is at odds with our modern understanding, and specifically with the notion of a personally and socially constructed self. We tend to view the self as a narrative structure, an ongoing creation that changes over the course of one's life. Notions of the formative role of child, the development of one's capabilities, and the cumulation of experience are distinctive features of our contemporary sense of human personality. Similarly, the idea that people are the products of their particular society, rather than standing in relation to a universal, supernatural order, is central to the worldview of an era that is widely identified as a secular age. According to this view, there is really nothing for evil to capture; it can become an event or factor in one's ongoing existence, but it cannot seize control of one's total being because that totality does not truly exist.

The modern conception of a personally and socially constructed self has important political implications. It means that the process of self-formation is a crucial aspect of political autonomy as well as personal development. This is, in effect, Isaiah Berlin's notion of negative liberty.[94] The state is forbidden to impose a general pattern of behavior on its citizens or even to urge such a pattern on them.[95] According to Berlin, the effort to create "Socialist Man" or "Aryan Man" was the essence of twentieth-

91. *See* Luther, *The Enslaved Will (1525)*, *in* ERASMUS AND LUTHER: THE BATTLE OVER FREE WILL, *supra* note 89, at 32, 110–11.

92. *See* JOHN CALVIN, INSTITUTES OF THE CHRISTIAN RELIGION 62–63 (Elise Anne Mckee trans., Wm. B. Eerdmans Publ'g Co. 2009) (1541); MAX WEBER, THE PROTESTANT ETHIC AND THE SPIRIT OF CAPITALISM 55–57 (Stephen Kalberg trans., Routledge 2012) (2001) (outlining Calvin's doctrine of predestination, which depends on the absence of free will). For a contemporary discussion of the competing views, see generally JOHN MARTIN FISCHER ET AL., FOUR VIEWS ON FREE WILL (2007).

93. *See* UNDERKUFFLER, *supra* note 1, at 80–81 (noting that corruption is a dispositional concept that confers an "irrevocable moral status" whereby an individual is deemed to be "completely and thoroughly consumed by [corruption]" (emphasis in second quotation omitted)).

94. *See* ISAIAH BERLIN, *Two Concepts of Liberty*, *in* FOUR ESSAYS ON LIBERTY 118, 121–22 (1969) (discussing negative liberty as the ability for humans to construct a sense of self within a defined area and without outside influences and giving as an example of negative freedom the idea that "[p]olitical liberty in this sense is simply the area within which a man can act unobstructed by others"); *id.* at 141 (discussing self-realization and stating that "[t]he only true method of attaining freedom . . . is by the use of critical reason, the understanding of what is necessary and what is contingent").

95. *See id.* at 122 ("You lack political liberty or freedom only if you are prevented from attaining a goal by human beings."); *id.* at 129–31 (suggesting that negative freedom questions how much a government should be able to limit individual liberty, whereas positive freedom questions who should impose those limits).

century totalitarianism.[96] The idea of self-formation as a human right is further linked to the theory of representative democracy. If the government is to be truly chosen by the people, then that government, once chosen, should not be able to exercise a significant impact upon its citizens' personalities and thus affect their future choices. Once again, this means that the government must not impose or even convey a view about its citizens' essential natures. It may only judge, and attempt to control, their specific actions.

Thus, the government's basic stance toward its citizens is that they do not have an essential nature, that there is nothing to be captured. To regard a particular offense as affecting people's basic natures—to transfer the premodern idea of corruption into the modern context—thus represents a conceptual mismatch. This does not necessarily create an immediate problem in pragmatic terms; as Oliver Wendell Holmes observed, there are many ways in which practical governance does not obey the rules of logic.[97] But when the time comes to construct operational rules, the reliance on underlying concepts that conflict with our basic mode of thought can readily lead us in the wrong direction.

Oddly, Underkuffler herself succumbs to the analytic version of this same mismatch. In discussing the deficiencies of the existing theories of corruption, she states repeatedly that these theories fail to "capture" the meaning of the term.[98] Just as the use of this term to define the offense of corruption implies that people have an essential nature, Underkuffler's use of it to assess those definitions implies that words have an essential nature. This is as outdated an approach to language as the definition she is criticizing is to law. The modern view, most closely associated with Derrida and Wittgenstein, is that words represent a range of usages that depend on the totality of our language or our life.[99] Thus, Wittgenstein points out that the word *game* refers to a range of activities; each one shares some characteristics with some of the other uses, but there is no core meaning shared by all these usages.[100] Instead, the usages display nothing more than "family resemblances" that reflect our forms of life.[101]

---

96. *See id.* at 131 (noting that the "great clash of ideologies" resulted from the misapplication of positive freedom, which adherents of the negative notion call "no better than a specious disguise for brutal tyranny").

97. O.W. HOLMES, JR., THE COMMON LAW 1 (Bos., Little, Brown & Co. 1881). For a general discussion of the views that Holmes's famous observation represents, see WILLIAM P. LAPIANA, LOGIC AND EXPERIENCE 89–91, 113–20 (1994).

98. *See* UNDERKUFFLER, *supra* note 1, at 8, 11, 16–17, 19–21, 37, 43, 46, 53.

99. *See* DERRIDA, *supra* note 43, at 6. *See generally* LUDWIG WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS (G.E.M. Anscombe trans., 3d ed. 1972).

100. WITTGENSTEIN, *supra* note 99, §§ 66–78.

101. *Id.* § 67. Underkuffler refers to this notion that the varying usages of a term display only family resemblances with one another in identifying a possible argument favoring theories that explain only certain uses, but not all uses, of the term corruption. UNDERKUFFLER, *supra* note 1,

This suggests a modification of Underkuffler's argument, but not a basic defect. It is not necessary for her to argue that the term *corruption* possesses some verbal essence—some core meaning—that a theory of the concept needs to "capture." It is enough to say that many contemporary uses of the term embed the outdated idea that the corrupt official has been captured by evil. Her observations about the discourse of contemporary prosecutions for corruption, and their connection to premodern concepts, are sufficiently convincing by themselves to establish the importance of her theory without the further assertion that they reveal the essential meaning of the term.

At the same time, the modification frees Underkuffler's theory from the obligation of explaining every situation to which the concept of political corruption currently applies. Underkuffler seems correct in arguing that the idea of evil, and of being captured by evil, animates most contemporary usages and that the prevailing shell theories, substantive theories, and economic theories are not able to distinguish misbehavior that we describe as corrupt from other misbehavior that we do not place in this category. But her characterization, despite its explanatory power, does not necessarily apply to all our uses of the term or concept of corruption. Consider for example the U.S. Foreign Corrupt Practices Act, which forbids American companies from bribing public officials in other nations.[102] Do we really regard those who violate this law as evil, or do we see them as rational decision makers accommodating themselves to the realities of the situation in many developing countries? The Act seems motivated by the policy-based recognition that some American companies are large enough to disrupt small foreign nations and the policy-based aspiration that these companies should abide by domestic standards in their behavior overseas rather than by the sense of moral condemnation that we feel when someone tries to bribe our own officials.

Modifying Underkuffler's theory in this way places greater emphasis on its historical contextualization. The reason that an outmoded, moralistic approach to the topic of political corruption has prevailed in contemporary times is not because the word itself necessarily implies this meaning and that it automatically transports the idea of being captured by evil into every context where it is employed. Rather, as suggested above, it is because the concept of corruption as being captured by evil is part of our historical inheritance. The advent of the administrative state has transformed our concepts of government and law, but the terms that we have inherited from the past will continue, and retain their premodern meanings, unless we consciously rethink them and revise those meanings in recognition of the

---

at 50–53. She rejects this, however, in favor of a theory that "captures" the whole range of uses. *Id.* at 53.

102. 15 U.S.C. § 78dd-1(a) (2012).

altered circumstances. We have failed to do so for basic terms such as legitimacy or law, and we have also failed to do so for the term corruption.

The idea that the term *corruption* does not have an identifiable essence, that it is merely a useful designation for a range of related actions, is connected with the political position that government may not treat people as possessing an essential nature. If there is nothing that can be captured in its totality, then there is no unified condition that can be described as corruption per se. Rather, people must be judged on the basis of their specific and separable acts, and those acts are likely to vary in their character and culpability. In other words, the idea that corruption is a status, not an action, depends on the equally outdated idea that people have an essential nature to which a status such as "corrupt" can be attached. In this way as well, the inheritance of our past conflicts with, and distorts, our present.

IV. Bringing Morality Back In

Moral considerations are not absent from modern law, however. While we accept the existence of *malum prohibitum* offenses, we are not necessarily comfortable with this concept. We prefer that actions punished as crimes—the most serious sanction our society deploys—are also condemned by our system of morality. Underkuffler draws upon and endorses this inclination. While she believes that the concept of corruption as capture by evil does us a disservice when it is used as the basis for legal rules, she also believes that it serves a real purpose as a component of our moral system.[103] It induces us to be vigilant about political corruption, to condemn it in all its forms, and to punish its perpetrators electorally, whether or not they are also punished legally.[104]

There is a great deal to be said in favor of this view. What can be said against it, however, is that a premodern concept—one that depends upon the notions that both individuals and corruption have an identifiable essence—is unlikely to provide any better guidance for our moral inclinations than for our legal rules. We are better served, according to the countervailing argument, by dispensing with the buttressing effects of morality and focusing more fully on the legal rules themselves. A number of economists have adopted a particularly strong version of this position by arguing that what we call corruption is economically efficient under certain circumstances.[105] As a result, the moral condemnation that we have

103. *See* UNDERKUFFLER, *supra* note 1, at 136–38.

104. *See id.*

105. *See* Daron Acemoglu & Thierry Verdier, *The Choice Between Market Failures and Corruption*, 90 AM. ECON. REV. 194, 209 (2000) (suggesting that government intervention with partial corruption could be optimal "when corruption is relatively rare and the market failure it is trying to correct is relatively important"); Bruce L. Benson & John Baden, *The Political Economy of Governmental Corruption: The Logic of Underground Government*, 14 J. LEGAL STUD. 391,

historically attached to these acts is not only unnecessary but counterproductive.

Most people, among them most economists, have found this view problematic. As Susan Rose-Ackerman argues, once corruption is tolerated, there is no guarantee that it will be limited to inefficient regulations.[106] The corporate executive who is willing to bribe government officials to relax a burdensome restriction on fair competition will also be willing to bribe those officials to relax essential health and safety regulations. To tolerate corruption on the grounds that some or even many official requirements are inefficient delegates the authority to set compliance policy to dishonest private parties. A dreary litany of collapsing bridges and buildings in developing countries attests to the dangers of that policy.

In addition, a tolerant approach to political corruption undermines the integrity of government in general and decreases its ability to carry out any of its functions. We cannot rely on public officials to behave like Orwell's Burmese judge, who "would never sell the decision of a case, because he knew that a magistrate who gives wrong judgments is caught sooner or later. His practice, a much safer one, was to take bribes from both sides and then decide the case on strictly legal grounds."[107] That justification of corrupt behavior was rejected in Francis Bacon's case, as Underkuffler notes,[108] and has been rejected ever since; here again, once corruption is tolerated, there is no way of restricting it to undesirable or excessive regulations. Rather, it will place the entire administrative and judicial apparatus under the influence of dishonest private parties and beyond the control of the officials who have been chosen by the citizenry.[109]

We are thus presented with a dilemma. On the one hand, morality seems to disserve us, distorting public policy and inserting discordant consideration into modern legal rules. On the other hand, we seem to need it as a means of generating the sense of condemnation necessary to combat corruption and to draw a clear boundary around all its manifestations. This dilemma is hardly unique to the subject of corruption of course. As a number of observers have noted, modern law in general has jettisoned the moralistic component of the premodern era and become "positivized," a social instrumentality designed to serve specific and identifiable public

---

410 (1985) (proposing the use of a cost–benefit analysis regarding the prevention of corruption associated with regulations); Nathaniel H. Leff, *Economic Development Through Bureaucratic Corruption*, AM. BEHAV. SCIENTIST, Nov. 1964, at 8, 11–12 (noting that in some circumstances corrupted officials can produce "more effective policy than the government").

106. ROSE-ACKERMAN, *supra* note 20, at 22–23.

107. GEORGE ORWELL, BURMESE DAYS 6 (1934).

108. UNDERKUFFLER, *supra* note 1, at 78–79.

109. *See* ROSE-ACKERMAN, *supra* note 20, at 22–23.

purposes.[110] But this process raises concerns about the ability of law to serve those purposes, given the crucial role of morality in inducing commitment to the law and compliance with its particular provisions.[111] Thus morality, however outdated, seems difficult to dispense with; this concern brings to mind the old adage that "you can't beat something with nothing."

In my view, however, the apparent dilemma is illusory. It arises from the assumption that premodern morality is morality itself and that modern administrative government is essentially amoral. This assumption, however, is implausible because virtually every society has a moral system of some sort. Perhaps some small, dysfunctional groups in the process of collapse can be described as lacking any operative moral system, like Colin Turnball's Ik.[112] But a large, complex society such as ours, that maintains such high levels of social order,[113] cannot function without a well-

110. *See* JÜRGEN HABERMAS, BETWEEN FACTS AND NORMS: CONTRIBUTIONS TO A DISCOURSE THEORY OF LAW AND DEMOCRACY 132–93 (William Rehg trans., Mass. Inst. Tech. Press 1996) (1992) [hereinafter HABERMAS, BETWEEN FACTS AND NORMS] (describing law as collective will formation); 1 JÜRGEN HABERMAS, THE THEORY OF COMMUNICATIVE ACTION 243–71 (Thomas McCarthy trans., Beacon Press 1984) (1981) (describing law as the instrumentally rational implementation of policy); 2 JÜRGEN HABERMAS, THE THEORY OF COMMUNICATIVE ACTION 361–73 (Thomas McCarthy trans., Beacon Press 1987) (1981) (describing the conversion of socially integrated contexts into the medium of law); NIKLAS LUHMANN, LAW AS A SOCIAL SYSTEM 76–141 (Fatima Kastner et al. eds., Klaus A. Ziegert trans., Oxford Univ. Press 2004) (1993) (describing law as a self-contained system of posited rules); WEBER, *supra* note 55, at 880–95 (describing law as the professionalized administration of justice). For my agreement with the view, see RUBIN, *supra* note 4, at 191–226.

111. *See* HABERMAS, BETWEEN FACTS AND NORMS, *supra* note 110, at 67 ("The addressees of a norm will be sufficiently motivated to comply with norms on the average only if they have internalized the values incorporated in the norms."); BRIAN Z. TAMANAHA, LAW AS A MEANS TO AN END: THREAT TO THE RULE OF LAW 215 (2006) (describing the perils of a legal system without "a built-in integrity, a core of good and right"). This is an animating concern of modern natural rights scholarship. *See* RANDY E. BARNETT, THE STRUCTURE OF LIBERTY: JUSTICE AND THE RULE OF LAW 21 (1998) (contending that only human laws in compliance with natural rights are obligatory); JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 354–62 (2d ed. 2011) (arguing that morals necessarily play a role in obtaining compliance with law); ROBERT P. GEORGE, IN DEFENSE OF NATURAL LAW 49 (1999) (discussing broadly the role of moral principles in structuring and guiding human choice); HEIDI M. HURD, MORAL COMBAT 1 (1999) (recognizing three foundational moral principles that guide our legal and political system); LLOYD L. WEINREB, NATURAL LAW AND JUSTICE 6 (1987) (proposing an account of moral responsibility); Michael S. Moore, *Law as a Functional Kind*, *in* NATURAL LAW THEORY: CONTEMPORARY ESSAYS 188, 198 (Robert P. George ed., 1992) (endorsing the view that "the justness of a norm is *necessary* to its status as law").

112. COLIN M. TURNBULL, THE MOUNTAIN PEOPLE 265–86 (1972). The Ik's problems, according to Turnbull, arose when they were forcibly removed from their home territory and restricted to a smaller area where they were unable to obtain adequate sustenance. *See id.* at 24–32.

113. Regarding the orderliness and functionality of modern society, see generally NORBERT ELIAS, STATE FORMATION AND CIVILIZATION, *reprinted in* THE CIVILIZING PROCESS 257 (Edmund Jephcott trans., 1994); ANTHONY GIDDENS, THE CONSEQUENCES OF MODERNITY 4 (1990); and STEVEN PINKER, THE BETTER ANGELS OF OUR NATURE: WHY VIOLENCE HAS DECLINED 175–86, 592–611 (2011).

established and largely internalized system of beliefs about proper and improper behavior.

As I argue in a forthcoming work,[114] we in fact possess a moral system, but it is different from the moral system of the premodern era. Just as government was transformed by the advent of the administrative state, so morality was transformed in the transition to the modern era. The new morality does not urge people to achieve salvation, or to serve their nation, but rather to seek their own personal self-fulfillment. Its primary rule is that all people are entitled to define their own idea of fulfillment and to the basic material, social, and political conditions that would allow them to pursue those self-defined goals. Its basic constraint on people's behaviors is that they must not interfere with the self-fulfillment of others. This produces the familiar pattern of modern moral prohibitions, including prohibitions against the condemnation of other people's lifestyles, of intolerance, of coercive sex in any form, of nonnurturing or nonsupportive parenting, and so forth. If one thinks of premodern morality as morality itself, then the modern world will naturally look amoral—and immoral—because it differs, in distinctive ways, from its premodern predecessor. But if one recognizes that morality can vary from one culture to the next, then the new morality of our modern culture becomes rapidly apparent.

This new morality is organically linked to the modern state and, in fact, both generates and is generated by modern governmental structures. In its political aspect as a representative democracy, the modern state, as described above, is forbidden to base its actions on a theory of people's essential nature. Rather, it must allow them to define their own views about their advantages and aspirations, and its representational mechanisms are consciously designed to induce it, if not compel it, to respond to those independently defined desires. In its administrative aspect, the purpose of the modern state is to provide services to people, to facilitate their efforts to achieve personal self-fulfillment. In addition to establishing internal order and external security, where its goals overlap with premodern European monarchies, the modern administrative state provides the basic necessities of life—subsistence, housing, health, and education—so that its citizens can pursue their individual life plans. Moreover, it enforces the general moral prohibition against discrimination and other forms of intolerance for the purpose of ensuring that all its citizens, regardless of race, religion, or gender, have relatively equal opportunities.

Underkuffler is correct in arguing that the premodern morality of sin and evil essences conflicts with modern legal notions and consequently produces distorting and deleterious effects when imported into our modern legal system. She is also correct in arguing that a system of rules designed

114. EDWARD L. RUBIN, THE NEW MORALITY: SELF-FULFILLMENT AND THE MODERN STATE (forthcoming 2014).

to curb potentially widespread abuses requires support from a system of morality in civil society to achieve widespread compliance. But the dilemma is not unresolvable—the new morality enables us to support and intensify a modern, act-based approach to political corruption. It suggests that the service-oriented aspect of government is not merely a matter of efficiency but a true moral imperative. Government is morally required to provide services to its citizens—services that facilitate their efforts to achieve personal self-fulfillment—and it must do so evenhandedly, without discrimination. Political corruption is immoral because it impedes the government in implementing this essential task.

This modern concept of morality informs both the way a government institution interacts with the public and the way that it is staffed. The requirement that the agency be service oriented and its obligation to treat people equally preclude its staff members from taking bribes, which is obvious, but also from exercising favoritism on some other basis. In other words, the institution's staff must not grant privileges to any member of the public for any reason related to their personal advantage. Distinctions among members of the public are permissible and, indeed, essential, but they must be based on the policy established by the legislature or by publicly adopted regulations or enforcement strategies designed to advance that policy.[115] In addition, the institution must choose its officials on the basis of merit, not on the basis of bribery, favoritism, patronage, or other such considerations.[116] Here again, the purpose is to ensure that the institution is acting to advance governmental policy, not for its members' private benefit.

As is apparent, modern morality fully supports our prevailing legal rules against political corruption. There is no need to rely on outmoded moral concepts from an earlier era; we can replace the concept of sin with the concept of disservice, and instead of viewing corrupt officials as captured by evil, we can view them as undermining democratically established public policy. The failure of modern corruption law, therefore,

---

115. *See* EUGENE BARDACH & ROBERT A. KAGAN, GOING BY THE BOOK: THE PROBLEM OF REGULATORY UNREASONABLENESS 87 (1982) (giving examples of Occupational Safety and Health Act (OSHA) enforcement officers rigidly enforcing violations of regulations because OSHA has an enforcement "philosophy that compels inspectors to adhere to the official view of what is feasible" instead of what is practical for each individual company in each instance). Bardach and Kagan recommend that inspectors enforcing OSHA avoid imposing fines for technical violations on firms that are basically cooperative, thus granting minor indulgences in exchange for voluntary compliance. *See id.* at 123–24. Such indulgences could be granted for corrupt purposes, of course, but their recommendation avoids this possibility since the agency would be using this device to advance the goals of the legislation, not for the personal benefit of its staff. *See id.* at 143–50 (noting that such flexibility can aid employers who share the agency's goals).

116. *See* MASHAW, *supra* note 82, at 239 (describing one purpose of the Pendleton Act as instituting competitive examinations for hiring and promotion within the government).

is not that it must detach itself from morality in order to employ contemporary legal standards but rather that it has failed to understand and incorporate the morality that has in fact arisen in the modern world and that conforms to those legal standards.

Even in developed nations, where the problem of corruption, although certainly significant, is manageable,[117] the failure to recognize the new morality as an antidote to its predecessor presents some notable difficulties. Underkuffler offers the example of Rod Blagojevich, the brazenly corrupt Governor of Illinois, who turned out to be surprisingly difficult to convict because of his boyish charm.[118] "Blagojevich and his antics seemed almost too good-natured, too good-hearted, *too ordinary*, to signal the kind of dangerous and evil hypocrite that [the] idea of corruption [as captured by evil] assumes."[119] But he ultimately was convicted, and even if he had miraculously escaped punishment (the evidence against him was overwhelming),[120] his good fortune would have been perceived as a fluke rather than as a signal that other corrupt officials would be immune. Our inability to rethink the basis of our legal and moral condemnation of corruption thus causes certain problems, as Underkuffler argues,[121] but we seem to have sufficient social and legal resources to keep these problems under control.[122]

In many developing countries, however, the problem is considerably more severe. These nations suffer from what has been called a "culture of

---

117. What may not be manageable is the more general problem of private money's distorting effect on our political process. *See* Frank J. Sorauf, *Politics, Experience, and the First Amendment: The Case of American Campaign Finance*, 94 COLUM. L. REV. 1348, 1366–67 (1994) (decrying a broken campaign finance system and lack of effective regulation); David A. Strauss, *Corruption, Equality, and Campaign Finance Reform*, 94 COLUM. L. REV. 1369, 1389 (1994) (noting various ambiguities that make campaign finance reform a daunting task); Zephyr Teachout, *The Anti-Corruption Principle*, 94 CORNELL L. REV. 341, 384–85 (2009) (critiquing *Buckley v. Valeo*, 424 U.S. 1 (1976), the seminal Supreme Court case on campaign finance, expenditures, and corruption). The Supreme Court's recent decision in *Citizens United v. Federal Election Commission*, 558 U.S. 310 (2010), represents a virtual abandonment of any effort to deal seriously with this issue. *See generally* Richard L. Hasen, Citizens United *and the Illusion of Coherence*, 109 MICH. L. REV. 581 (2011). But these legally permissible inequalities, however much they may distort the democratic process, are not the kinds of problems to which Underkuffler's book is addressed. If the term corruption is applied to them, as it is by the above-cited authors, it is being used in a distinctly different sense.

118. UNDERKUFFLER, *supra* note 1, at 145–55.

119. *Id.* at 154.

120. *Id.* at 145–46, 152–53.

121. *See id.* at 107–28.

122. *See, e.g.*, Norman Abrams, *The Distance Imperative: A Different Way of Thinking About Public Official Corruption Investigations/Prosecutions and the Federal Role*, 42 LOY. U. CHI. L.J. 207, 209 (2011) (using the term *distance imperative* to describe the social and legal mechanisms that should be employed to solve the problem of allegedly corrupt officials meddling in, or otherwise interfering with, their own prosecutions and noting that federal prosecutors can obtain convictions in cases where corrupt state officials refuse to act).

corruption."[123] Public officials regularly demand bribes for performing their regular responsibilities, such as issuing licenses, and they create new problems for people that require bribes to resolve, such as false accusations of illegality.[124] Public positions are filled by demanding bribes as well or on the basis of illegal patronage or nepotism.[125] Jean-François Médard describes the situation as neo-patrimonial.[126] The idea is that a modern administrative structure was imposed by colonial powers on traditional, kinship-based societies.[127] The structure demands Weberian rationality, but the underlying society, which resembles premodern Europe, then takes control of the structure and distorts it in accordance with its own norms and values, thus producing a pattern of endemic corruption.[128]

---

123. *See* VICTOR T. LE VINE, POLITICAL CORRUPTION: THE GHANA CASE 12 (1975) (noting that bribery, graft, nepotism, and favoritism "at all levels of officialdom" signaled the establishment of such a culture of corruption in Ghana); *see also* MARTIN MEREDITH, THE FATE OF AFRICA 172–75 (2005) (describing how bribery and corruption quickly became a "way of life" in post-colonial West Africa); JAMES C. SCOTT, COMPARATIVE POLITICAL CORRUPTION 2–3 (1972) (stating that the particular political system of a given country will help determine the amount of corruption that is tolerated, which suggests that a culture of corruption could arise as a result of the political system of a particular country). For a well-known fictional portrayal of this situation, see CHINUA ACHEBE, NO LONGER AT EASE (1960).

124. *See* LE VINE, *supra* note 123, at 25–26, 32–33 (discussing the practice of taking bribes for import licenses in Ghana); ROSE-ACKERMAN, *supra* note 20, at 18 (discussing Mexican and Kenyan officials who take bribes for licenses and permits); SCOTT, *supra* note 123, at 81–82 (detailing rampant corruption in Indonesia whereby groups bribe officials to receive state-guaranteed economic privileges); Tom Lodge, *Political Corruption in South Africa: From Apartheid to Multiracial State*, *in* POLITICAL CORRUPTION: CONCEPTS AND CONTEXTS, *supra* note 66, at 403, 409–10 (documenting evidence of extortion, bribes in exchange for permits and licensing, and other forms of corruption in South Africa); Andrei Shleifer & Robert W. Vishny, *Corruption*, 108 Q.J. ECON. 599, 608 (1993) (discussing literal roadblocks in India that prevent vehicles from passing through without paying a bribe).

125. *See, e.g.*, PAUL D. HUTCHCROFT, BOOTY CAPITALISM: THE POLITICS OF BANKING IN THE PHILIPPINES 73 (1998) (chronicling the corruption surrounding the allocation of banking licenses in the Philippines); LE VINE, *supra* note 123, at 64–71 (chronicling the rampant corruption in a state-run betting pool for soccer games in Ghana); ROSE-ACKERMAN, *supra* note 20, at 104–10 (giving examples of industries, such as the credit and banking sectors, that suffer from corruption and patronage in developing countries); Mushtaq H. Khan, *Patron-Client Networks and the Economic Effects of Corruption in Asia*, *in* POLITICAL CORRUPTION: CONCEPTS AND CONTEXTS, *supra* note 66, at 467, 468 (discussing patronage networks in Thailand); Donatella della Porta & Alberto Vannucci, *Corrupt Exchanges and the Implosion of the Italian Party System*, *in* POLITICAL CORRUPTION: CONCEPTS AND CONTEXTS, *supra* note 66, at 717, 724–25 (discussing how control of political power was bought and sold in Italy).

126. Jean-François Médard, *Corruption in the Neo-Patrimonial States of Sub-Saharan Africa*, *in* POLITICAL CORRUPTION: CONCEPTS AND CONTEXTS, *supra* note 66, at 379, 380.

127. *See* BASIL DAVIDSON, THE BLACK MAN'S BURDEN: AFRICA AND THE CURSE OF THE NATION-STATE 38–51 (1992) (describing the struggles associated with fitting foreign forms of government into Africa's culture and the problem of having foreigners draw the political boundary lines with no concern as to how Africa had previously been divided and governed before European rule).

128. *See* Médard, *supra* note 126, at 382–84 (suggesting that the particular pervasiveness of corruption in African nations is due in part to the fact that corrupt acts in those nations are designed to entrench or reorient particular social structures).

Transitional economies in Eastern Europe and Central Asia also fit within this pattern. During the Communist Era, their official governmental structure and economy, although certainly not free of corruption, was organized according to the modern pattern that separates the public and private realms and imposes a bureaucratic, rationalized structure on the public one.[129] The problem is that the economic system that this bureaucracy implemented was so poorly conceived that much of the society's ordinary and necessary commerce was driven underground, into the black market or second economy.[130] There, patrimonial patterns prevailed; when these nations then established a market economy, the people who were best positioned to control or thrive in that economy were those who had been active in the second economy, and they brought their patterns of behavior with them into the post-Communist setting.[131]

Efforts to reform endemically corrupt or neo-patrimonial nations often employ the moralistic approach that Underkuffler describes. They regard corruption as an irremediable sin and the corrupt official as having been captured by evil. "Integrity initiatives" are combined with prosecutorial task forces and enhanced criminal punishment.[132] Officials convicted of corruption tend to be permanently banned from government.[133] Even though many of these nations did not have the Christian cultural background that seems to have generated the perspective Underkuffler describes, they have adopted its strictures—in part because they see this moralistic approach as intrinsic to the process of Westernization and industrialization, and in part because it is being imposed on them by the World Bank and other development organizations.[134]

A preferable approach might be to jettison premodern morality in its entirety and fashion the effort to reduce corruption on the basis of the new

---

129. *See* JÁNOS KORNAI, THE SOCIALIST SYSTEM: THE POLITICAL ECONOMY OF COMMUNISM 36–39 (1992).

130. SHEILA FITZPATRICK, EVERYDAY STALINISM 59 (1999).

131. *See* Mark Levin & Georgy Satarov, *Corruption and Institutions in Russia*, 16 EUR. J. POL. ECON. 113, 128 (2000); Jasmine Martirossian, *Russia and Her Ghosts of the Past*, *in* THE STRUGGLE AGAINST CORRUPTION: A COMPARATIVE STUDY 81 (Roberta Ann Johnson ed., 2004).

132. *See, e.g.*, Médard, *supra* note 126, at 393–97; *Anti-Corruption and Public Integrity Background, Rule of Law Initiative*, AM. BAR ASS'N, http://www.americanbar.org/advocacy/rule_of_law/thematic_areas/anti_corruption/background.html (giving an example of such an initiative and outlining five aspects of the ABA's anticorruption program).

133. *See* UNDERKUFFLER, *supra* note 1, at 80–81 (noting that only three of the ninety-seven high-ranking public officials convicted of corruption in the United States resumed public careers).

134. *See* Thomas M. Callaghy, *Africa and the World Political Economy: Still Caught Between a Rock and a Hard Place?*, *in* AFRICA IN WORLD POLITICS 39, 43 (John W. Harbeson & Donald Rothchild eds., 4th ed. 2009); David F. Gordon, *Debt, Conditionality, and Reform: The International Relations of Economic Restructuring in Sub-Saharan Africa*, *in* HEMMED IN 90, 104 (Thomas M. Callaghy & John Ravenhill eds., 1993); Reginald Herbold Green, *The IMF and the World Bank in Africa: How Much Learning?*, *in* HEMMED IN, *supra*, at 54, 57; David Simon, *Debt, Democracy and Development: Sub-Saharan African in the 1990s*, *in* STRUCTURALLY ADJUSTED AFRICA 17, 19 (David Simon et al. eds., 1995).

morality of self-fulfillment and administrative governance. Rather than being regarded as sin, corruption could be seen as a failure to provide adequate service to the citizens. The focus on the moral status of the corrupt public official would be replaced by considerations of the official's specific actions as a part of an administrative system. Instead of speaking of public officials as having become corrupt—a reflection of the captured-by-evil idea, as Underkuffler notes—we would speak of them as having taken a bribe, or hired a subordinate, on the basis of patronage. Instead of banishing them from government and imposing severe sanctions, we would consider the gravity of their offense and punish them accordingly. This approach, which is certainly milder but perhaps better described as more modulated, does not represent an abandonment of morality but rather the substitution of a new morality, based on government service, for an older one based on sin, Satan, and damnation.[135]

In developing nations where corruption is such a widespread problem, it is a natural instinct to conclude that penalties for corruption should be more, rather than less, severe, but there are many reasons why the more modulated approach of modern morality might be preferable in this situation. To begin with, if corruption is endemic to the society, it seems hard to argue that each individual who engages in corrupt practices has been captured by evil. This would lead to the disparaging and implausible conclusion that developing nations are filled with evil people, whereas more settled, prosperous societies are paragons of social virtue. Moreover, it is often the case in developing societies that public officials are not paid wages commensurate with their training, or that they are not paid even a living wage, or that they are not paid at all when financial crises occur.[136] In these cases, it is often understood that they will survive by taking bribes and selling favors, activities which thus become more similar to legal, proprietary offices, rather than true corruption. This practice needs to be stopped of course, but to do so by means of severe punishment seems both excessively and unnecessarily harsh.

---

135. *See* Susan Rose-Ackerman, *Introduction: The Role of International Actors in Fighting Corruption*, in ANTI-CORRUPTION POLICY: CAN INTERNATIONAL ACTORS PLAY A CONSTRUCTIVE ROLE, 3, 34 (Susan Rose-Ackerman & Paul D. Carrington eds., 2013) ("Reductions in corruption are not ends in themselves, but are part of the global focus on improving human well-being and governmental functioning."); *see also* Joost Pauwelyn, *Different Means, Same End: The Contribution of Trade and Investment Treaties to Anti-Corruption Policy*, in ANTI-CORRUPTION POLICY: CAN INTERNATIONAL ACTORS PLAY A CONSTRUCTIVE ROLE, *supra*, 247, 248–50 (suggesting that corrupt practices can be challenged as barriers to trade, apart from any finding of criminal liability).

136. *See* ROSE-ACKERMAN, *supra* note 20, at 71–73 (discussing the problem of low civil service pay and stating that incentives to pay bribes are high when high-skill civil service employees are underpaid); Richard C. Crook, *Rethinking Civil Service Reform in Africa: 'Islands of Effectiveness' and Organisational Commitment*, 48 COMMONWEALTH & COMP. POL. 479, 481 (2010); Barbara Nunberg & Robert R. Taliercio, *Sabotaging Civil Service Reform in Aid-Dependent Countries: Are Donors to Blame?*, 40 WORLD DEV. 1970, 1970, 1974 (2012).

As a pragmatic matter, moreover, it is simply not possible to address the problem of endemic corruption by imprisoning the entire police force or all the members of a particular administrative office that is regulating some sector of the economy. In a developed nation, the dismissal and imprisonment of a corrupt official rarely creates a staffing problem, given the supply of human resources,[137] but this is not necessarily the case in a developing one, particularly given the scale of corruption. Inevitably, therefore, efforts to end corruption in these settings must proceed by making examples of targeted individuals. Doing so not only raises general questions of fairness but also may cause pragmatic difficulties in obtaining convictions. Modern morality suggests that widespread corruption, being viewed as failure of service, might be punished by administrative rather than criminal means, such as a year's suspension without pay for the first offense and dismissal or prison only for recidivism. Delimited sanctions of this sort could be more swiftly and widely imposed and would probably seem more fair as well.

Shifting to modern morality as a support for an anticorruption campaign also has the advantage of avoiding standardless rules of behavior, another of the problems that Underkuffler identifies with the captured-by-evil approach.[138] If we see specific acts of political corruption as evidence that the person accused has become intrinsically evil, we must necessarily seek evidence of a sort that would otherwise be deemed irrelevant and has no obvious boundaries or limitations. Shifting to modern morality not only avoids this problem but also provides positive assistance in defining and clarifying the offense. By recognizing that the government's provision of services to the citizens is itself a moral imperative, it focuses attention on the precise nature of the action that we want to condemn. Thus, receiving a gift from a person who was granted a license on valid grounds may be treated as noncriminal (depending on social practices), whereas granting a license to an undeserving friend, even without receiving any financial benefit oneself, might count as corruption. Similarly, the shift to modern morality attaches a normative value to merit-based hiring and may help in clarifying when a discretionary appointment should be regarded as corrupt and when it is based on the desire to create an effective work environment.

Conclusion

Laura Underkuffler's *Captured By Evil* is a major contribution to the

137. In the Chicago Greylord investigation, fifteen Cook County judges were convicted of taking bribes, one of the most widespread scandals in modern American history. *See* Maurice Possley, *Operation Greylord: A Federal Probe of Court Corruption Sets the Standard for Future Investigations*, CHI. TRIB., Aug. 5, 1983, http://www.chicagotribune.com/news/politics/chi-chicagodays-greylord-story,0,4025843.story. It revealed many problems about local government, but there was no difficulty replacing the judges.

138. UNDERKUFFLER, *supra* note 1, at 124–27.

literature on political corruption. This issue will never fully vanish from even the most settled, well-organized nations, and it has become an increasing focus of attention as an explanation for the vicissitudes of the most unsettled, poorly organized ones. In addition, the book explores the enormously complex and important relationship between morality and governance. Underkuffler convincingly argues that our prevailing attitude toward corruption is infused with a premodern moral perspective that regards political corruption as a process of being captured by evil, an only partially secularized version of seduction by Satan. She further argues, again convincingly, that importing such concepts into a modern, administrative legal system creates a range of problems involving modes of proof, ungraded offenses, and the lack of coherent legal standards.

Underkuffler further notes, however, that it is difficult to dispense with the moral underpinnings of the law and view legal rules as purely positive enactments. In order to retain this normative component of law, however, we need not return to the religiously based morality of premodern times. The culture of our own High Modern Era and the system of administrative governance associated with it, generates a morality of its own. This is in fact the morality to which we subscribe, as denizens of this era, even if we do not recognize it as such. It is based on the normative position that government is expected to serve people's interests and that those interests are defined by the people themselves, as an effort to achieve personal self-fulfillment, rather than by some external force or authority. As such, it provides a basis for condemning political corruption, but doing so in a strategic, modulated fashion rather than in the quasi-hysterical righteousness of a bygone time.

Corporate Governance and Social Welfare in the Common-Law World

CORPORATE GOVERNANCE IN THE COMMON-LAW WORLD: THE POLITICAL FOUNDATIONS OF SHAREHOLDER POWER. By Christopher M. Bruner. New York, New York: Cambridge University Press, 2013. 309 pages. $110.00.

Reviewed by David A. Skeel, Jr.*

Introduction

Sometime around 1990, American corporate governance scholars discovered that corporate law and corporate governance do not work the same way everywhere in the world. This was not the first time American corporate governance scholars had made such a discovery. Comparative corporate governance scholarship flourished for a few years in the 1970s, and earlier generations had done their own comparative spadework.[1] But the new wave of scholarship is not shaped in discernible ways by its predecessors and has brought new tools and perspectives to bear.

Many of the articles at the beginning of this wave used the comparisons primarily to shed light on American corporate governance, often to demonstrate that features of American governance are not inevitable. Perhaps most prominently, Mark Roe contrasted governance patterns in Germany and Japan in connection with his political theory of American corporate governance.[2] In Roe's account, the separation between ownership and control in America's "Berle–Means" corporations was not simply caused by economic imperatives, as corporate law scholars often

---

* S. Samuel Arsht Professor of Corporate Law, University of Pennsylvania Law School. I am grateful to John Armour, Brian Cheffins, David Garland, Jennifer Hill, Dionysia Katelouzou, and Mark Roe for helpful comments and conversations, and to the editors of the *Texas Law Review* for inviting me to write this Review.

1. *See, e.g.*, Edward B. Rock, *America's Shifting Fascination with Comparative Corporate Governance*, 74 WASH. U. L.Q. 367, 378 (1996) (noting that in the 1970s American legal scholars compared the German approach to legal and institutional arrangements in corporate governance, including the two-tier board and codetermination).

2. *See, e.g.*, Mark J. Roe, *A Political Theory of American Corporate Finance*, 91 COLUM. L. REV. 10, 59–61 (1991) [hereinafter Roe, *Political Theory*] (comparing bank involvement in German and Japanese firms with bank involvement in the United States); Mark J. Roe, *Some Differences in Corporate Structure in Germany, Japan, and the United States*, 102 YALE L.J. 1927, 1929 (1993) [hereinafter Roe, *Corporate Structure*] (explaining that differences between the corporate governance structures of Germany and Japan on the one hand and the United States on the other can be explained by an economic model that takes into account political histories, cultures, and paths of economic development).

tended to assume.[3] Politics also played a starring role, with populist pressures first forcing the fragmentation of American financial channels in the nineteenth century and then forcing the fragmentation of American corporate ownership at key junctures when institutions had begun to acquire major stakes in American corporations.[4] But were outcomes other than the American one possible? Germany and Japan offered a startling contrast. In each country, banks held or controlled major stakes in the nation's largest corporations and exerted much more direct influence over corporate governance.[5]

A few years later, additional theories of comparative corporate governance began to emerge, some challenging Roe's political theory, or at least providing additional explanations for the American outcome. In a series of much-criticized and highly influential empirical articles, economists Rafael La Porta, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny identified a variety of factors which, they argued, shape the quality of a nation's corporate governance.[6] Among other things, countries that protect minority shareholders are likely to see more diffuse ownership than those that do not[7] and countries with common law origins fare better than their civil law counterparts.[8] Turning his sight more fully to Europe, Roe offered an alternative account of corporate governance differences, one not tied to legal origin. In Roe's account, corporations in countries with social democracies tend to have more stakeholder-oriented corporate governance, with greater solicitude for employees, whereas countries that lack long-term social democratic control are more shareholder oriented.[9] Operationally, when labor was able to make strong claims on large firms' cash flow, as he said was common in the post-World War II decades, concentrated ownership with blockholders and controlling shareholders was more likely to preserve value for shareholders than diffuse ownership and managerial-centered firms.[10] In a similar vein, more recent

3. Roe, *Political Theory*, *supra* note 2, at 10; Roe, *Corporate Structure*, *supra* note 2.

4. MARK J. ROE, STRONG MANAGERS, WEAK OWNERS: THE POLITICAL ROOTS OF AMERICAN CORPORATE FINANCE 39–42 (1994). These political determinants are a central theme of Roe's influential book. *See id.* at xiv–xv.

5. *Id.* at 186.

6. *See, e.g.*, Rafael La Porta et al., *Investor Protection and Corporate Governance*, 58 J. FIN. ECON. 3, 24 (2000) [hereinafter La Porta et al., *Investor Protection*] ("[S]trong investor protection is associated with effective corporate governance . . . ."); Rafael La Porta et al., *Legal Determinants of External Finance*, 52 J. FIN. 1131, 1132 (1997) [hereinafter La Porta et al., *Legal Determinants*] (listing the origin of a country's laws, its investor protections, and quality of law enforcement as influencing a country's external finance).

7. Rafael La Porta et al., *Agency Problems and Dividend Policies Around the World*, 55 J. FIN. 1, 4 (2000).

8. La Porta et al., *Legal Determinants*, *supra* note 6, at 1137, 1149.

9. *See* MARK J. ROE, POLITICAL DETERMINANTS OF CORPORATE GOVERNANCE 29–37 (2003).

10. *See id.* at 17, 199.

work on varieties of capitalism contrasts liberal market economies, on the one hand, with coordinated markets on the other.[11]

These comprehensive theories tend to draw a sharp distinction between corporate governance in the United States and United Kingdom as compared to governance elsewhere in Europe and Japan.[12] The distinction is sensible, given that U.S. and U.K. corporations do not seem to have controlling shareholders,[13] governance in the two countries is more shareholder oriented than governance elsewhere in the world,[14] and the nations share a common history.[15]

But a strange thing happens if, after conducting these comparisons at 30,000 feet, we make our way down to the actual details of U.S. and U.K. corporate governance: at close range, they do not look so similar at all.[16] In the United Kingdom, shareholders can call a shareholders' meeting and displace the directors, or effect a major change, in any corporation at any time.[17] In the United States, by contrast, shareholders cannot replace the directors of a firm that has an effective staggered board without cause, and they cannot initiate fundamental changes on their own.[18] Similarly, if a hostile bidder makes an offer to the shareholders of a U.K. firm, the board of directors cannot interfere with the bid,[19] as the directors of Manchester United[20] and, more recently, Cadbury's are (from their perspective, at least) all-too-well aware.[21] In the United States, by contrast, target directors have considerable flexibility to fend off unwanted suitors.[22] What are we to make of these very significant differences, which suggest that U.K.

---

11. *See* Peter A. Hall & David Soskice, *Introduction* to VARIETIES OF CAPITALISM: THE INSTITUTIONAL FOUNDATIONS OF COMPARATIVE ADVANTAGE 1, 8 (Peter A. Hall & David Soskice eds., 2001).

12. *See* Ruth V. Aguilera et al., *Corporate Governance and Social Responsibility: A Comparative Analysis of the UK and the US*, 14 CORP. GOVERNANCE 147, 147 (2006) (noting the tendency in scholarship to separate Anglo-American forms of corporate governance from Continental European and Japanese forms).

13. ROE, *supra* note 9, at 16; *see also* La Porta et al., *supra* note 7, at 3 (stating that in the United States and United Kingdom, large firms are generally controlled by managers).

14. *See* Aguilera et al., *supra* note 12.

15. John Armour, Brian R. Cheffins & David A. Skeel, Jr., *Corporate Ownership Structure and the Evolution of Bankruptcy Law: Lessons from the United Kingdom*, 55 VAND. L. REV. 1699, 1715 (2002).

16. Each of the distinctions noted in this paragraph is described in detail in the book under review, CHRISTOPHER M. BRUNER, CORPORATE GOVERNANCE IN THE COMMON-LAW WORLD: THE POLITICAL FOUNDATIONS OF SHAREHOLDER POWER (2013).

17. *Id.* at 29.

18. *Id.* at 39.

19. *Id.* at 33.

20. Jere Longman, *American Not Welcome in Manchester*, N.Y. TIMES, Oct. 26, 2004, http://www.nytimes.com/2004/10/26/sports/soccer/26united.html?pagewanted=print&position=.

21. BRUNER, *supra* note 16, at 253–56.

22. *Id.* at 41.

governance is more truly shareholder oriented, whereas the United States protects managerial discretion?

One response to this kind of puzzle is to tease out the distinctive features of each system without attempting to reconcile them. Alternatively, one might simply chalk up the differences to the idiosyncracies of the United Kingdom, which confounds nearly every general account of corporate governance.[23] Another response is to take an opposite tack. Rather than emphasizing the unique features of each system, as contextualists might, functionalist accounts often resolve tensions at a high level of generality. The preeminent functionalist account of corporate law, *The Anatomy of Corporate Law*, identifies three agency cost problems—conflicts of interest between shareholders and managers, conflicts between controlling and minority shareholders, and conflicts between shareholders and creditors and other third parties—as the central concern of corporate governance in every developed nation.[24] From this perspective, the United States and United Kingdom (like the other developed countries considered by *The Anatomy of Corporate Law*[25]) are addressing the same problems in similar, but not identical, ways.

Like La Porta et al. and Roe, Christopher Bruner seeks in *Corporate Governance in the Common-Law World* to claim a middle ground by identifying a key feature—an independent variable or variables, as an empirical scholar might say—that explains enough of the differences to be worth highlighting—as did political economy for Roe and legal origin for La Porta et al. For Bruner, differences between the United States and United Kingdom that confound other governance theories can be understood by focusing on a new variable, each nation's social welfare system.[26] If a country has a robust social welfare system, he argues, corporate law can and will focus more narrowly and more strongly on the interests of shareholders.[27] If the country's social welfare system is weak, by contrast, corporate governance is likely to fill in the gaps by inviting managers to take the concerns of employees into account rather than attending solely to the shareholders' interest.[28] This explains why corporate governance is so shareholder oriented in the United Kingdom, which has universal healthcare and generous unemployment benefits,[29] while shareholders' powers are more attenuated in the United States, with its

23. For discussion of the difficulties of fitting the United Kingdom within current theories of comparative corporate governance, see Armour, Cheffins & Skeel, *supra* note 15, 1714–20.

24. REINIER KRAAKMAN ET AL., THE ANATOMY OF CORPORATE LAW: A COMPARATIVE AND FUNCTIONAL APPROACH 36 (2d ed. 2009).

25. *Id.* at 3.

26. BRUNER, *supra* note 16, at 4.

27. *Id.* at 4–5.

28. *See id.* at 5.

29. *See id.* at 145, 169.

much weaker social welfare protections.[30] Turning to the two other major common law countries, Australia and Canada, Bruner concludes that they fall in between but are more similar to the United Kingdom in their welfare protections and shareholder orientation.[31]

Bruner's social welfare account of corporate governance is a partial theory rather than a theory of everything. The thesis applies only to countries whose corporations tend to have dispersed ownership and does not speak to the many countries whose corporations tend to have a controlling shareholder or shareholders.[32] It suggests, in a sense, that Roe's social democracy insights invert when stock ownership is dispersed, with robust social welfare protections traveling together with highly-shareholder-oriented governance.[33] Or it at least qualifies Roe's argument that labor's capacity to make strong claims on a firm's cash flow is complementary with concentrated ownership. Unlike Roe, Bruner has a refined sense of social democracy: a social democracy that has powerful labor inside the firm might lead to concentrated ownership, but a social democracy that has government satisfy the social democratic demands outside the firm will lead to fewer social democratic demands inside the firm and, hence, will open up political and social space for a more intense shareholder orientation of the firm. Britain fits this latter case, Bruner argues.[34]

After describing Bruner's theory and evidence in more detail in the first Part of this Review, I poke at it from several angles in the two Parts that follow. In Part II, I consider whether there is a mechanism that adequately explains the connection between social welfare and shareholder orientation; interestingly, despite the book's title, Bruner does not suggest that the common law plays any particular role.[35] In Part III, I consider whether shareholders in the United States may have more power than their limited formal rights suggest, and in Part IV I ask whether the United States (rather than the United Kingdom, as is conventionally assumed) may simply be an outlier, due in large part to federalism and as reflected by the United States' weak social welfare system. I then conclude.

Although I will be playing devil's advocate throughout this Review, Bruner's insights are a revelation. As I emphasize by way of conclusion, he has identified a critical, new dimension of our understanding of corporate law.

---

30. *See id.* at 166–69.

31. *Id.* at 176, 200.

32. *Id.* at 3–4.

33. *Id.* at 120–23.

34. *See id.* at 143–66.

35. *See id.* at 117–19 (critiquing arguments suggesting that relative levels of shareholder-centrism depend on the systems of law of particular countries).

I. Corporate Governance and Social Welfare: The Central Thesis

Corporate Governance in the Common-Law World is no mystery story. Bruner states his thesis about shareholder-centeredness and social welfare at the outset then avers to it repeatedly throughout the book. He writes:

> My core claim is that greater regard for the interests of employees in other regulatory domains has tended to insulate certain corporate governance systems from political pressure to show regard for employees and other "stakeholders," permitting more exclusive focus on shareholders without precipitating backlash – a key political determinant of the relatively higher degree of shareholder-centrism exhibited in Australia and the United Kingdom, and to a lesser (but nevertheless substantial) degree in Canada as well.[36]

In the United States, by contrast:

> [W]eaker regard for the interests of employees in other regulatory domains has tended to result in greater political pressure being brought to bear on corporate governance to do so, inhibiting exclusive focus on shareholders – a key political determinant of the relatively lower degree of shareholder-centrism exhibited in the United States.[37]

Bruner's claim—that the robustness of these four countries' social welfare systems determines how shareholder-centered the corporate governance of each is (and thus that "the political foundations of shareholder power effectively lie outside corporate law itself"[38])—is elegant, though substantiating it is inevitably more of a slog. In the first of the two heftiest chapters of the book, Bruner outlines the extent of shareholder-centeredness of the corporate governance in each of his four countries.[39] Bruner begins by describing the shareholder-centric features of U.K. governance and the limits on U.S. shareholder influence—including directors' ability to defend against takeovers and the shareholders' inability to adopt bylaws that constrain directors' discretion.[40] Australian corporate governance is more shareholder oriented than the United States, as evidenced by the outcry that attended the decision by Rupert Murdoch to move the incorporation of his News Corp. empire from Australia to the United States.[41] Canada is a more complicated case. Although its securities

---

36. *Id.* at 22–23.

37. *Id.* at 23.

38. *Id.* at 27.

39. *See id.* at 28–107.

40. *Id.* at 29–65.

41. *Id.* at 71–73. Bruner's discussion of News Corp. draws extensively from a careful study by Jennifer Hill. *Id.* (citing Jennifer G. Hill, *Subverting Shareholder Rights: Lessons from News Corp.'s Migration to Delaware*, 63 VAND. L. REV. 1 (2010)).

and takeover law, which is regulated by the provinces, is quite proshareholder, Canada's corporate law makes important reference to the interests of other constituencies.[42] Based in part on an analysis of the high-profile *BCE Inc. v. 1976 Debentureholders*[43] case, Bruner concludes that its effect is more shareholder oriented than the law on the books seems to suggest.[44] Here and elsewhere in the book, Bruner illustrates the comparisons with helpful charts highlighting the salient corporate governance features of the four countries.[45]

In the book's other major chapter, Bruner conducts a similar exercise with social welfare, while keeping the parallel history of corporate governance in each country continuously in view.[46] Although the United Kingdom first began to construct its social welfare system in the early twentieth century, its full flourishing came after World War II, which brought universal healthcare and substantial unemployment benefits.[47] In the United States, by contrast, social welfare benefits come primarily through an employee's corporate employer, and healthcare coverage is much more spotty.[48] The contrast is reflected in labor's different response to the rise of takeovers in the two countries. After initially opposing shareholder-friendly takeover rules, British labor accepted them in the 1960s,[49] whereas American labor continued to call for limits on takeovers.[50] Australia is in some respects the most interesting case study because its stance on takeovers shifted. After initially seeming to permit directors to defend against takeovers, Australia adopted the U.K. shareholder-centric approach in the early 2000s.[51] Bruner argues that the shift was made possible by Australia's expansive, new, social welfare framework put in place between the 1970s and 1990s.[52]

In the book's final chapter, Bruner begins by contrasting the dispersed shareholder regimes of the four countries under discussion with the concentrated share ownership in countries like Germany.[53] "[I]n a country where blockholding predominates," he writes, "the principal regulatory issue remains how to counteract the blockholders' innate power over

42. *Id.* at 77–78, 84–89.
43. [2008] 3 S.C.R. 560 (Can.).
44. BRUNER, *supra* note 16, at 89–92.
45. *See, e.g.*, *id.* at 53, 68, 78, 83.
46. *Id.* at 143–220.
47. *See id.* at 145–47.
48. *Id.* at 167–71.
49. *Id.* at 153.
50. *See id.* at 166–67 (describing how U.S. corporate law favored promanagement takeover rules through the 1980s).
51. *Id.* at 192–93.
52. *See id.*
53. *Id.* at 223, 225–29.

corporate affairs through various forms of stakeholder-oriented protections."[54] The interests of shareholders are secondary. Bruner also considers three countries that might seem to confound his theory: China, Japan, and the Netherlands.[55] In each of these countries, blockholder share ownership is even lower than in the United Kingdom or United States, he notes, "yet these countries have not historically exhibited the forms of shareholder orientation that I associate with a high degree of ownership dispersal."[56] In China, the government has effective control over most large corporations, and in the Netherlands shareholder influence is stymied by a standard trust arrangement that holds voting control of the corporation.[57] In Japan, crossholdings by lenders traditionally neutralized the influence of ordinary shareholders, although they now appear to be breaking down.[58] Bruner speculates that "deeply entrenched historical, cultural, and political commitments" pull China, the Netherlands, and Japan toward effective concentrated ownership, whereas a different set of historical, cultural, and political commitments tug Australia and Canada in the opposite direction.[59] In the second half of the chapter, Bruner considers shifting shareholder patterns in the United Kingdom and postcrisis developments in the United Kingdom and United States, which he describes, drawing on a framework developed by Peter Gourevich,[60] in terms of shifting coalitions among shareholders, managers, and employees.[61] The United Kingdom has seen increasing shareholder power, while employees' power and welfare protections have declined,[62] whereas shareholder and employee power have both increased in the United States, thanks to a handful of shareholder-centered provisions in the Dodd-Frank Act and the enactment of healthcare legislation.[63]

Along the way, Bruner manages to work in succinct discussions of each of the major corporate governance debates in the American scholarly literature. In addition to a chapter critiquing each of the main comparative

54. *Id.* at 228.

55. *Id.* at 229–36.

56. *Id.* at 229.

57. *Id.* at 230–32.

58. *Id.* at 233.

59. *Id.* at 236.

60. *See id.* at 131 (locating the notion of shifting coalitions of stakeholders in a prior Gourevich work).

61. *Id.* at 242–86.

62. *See id.* at 286 (noting that "the postcrisis response[] of the United Kingdom" has "strengthened shareholders yet weakened protections for stakeholders").

63. *See id.* at 280–81, 283–84, 286 (explaining that postcrisis reform efforts in the United States "have tended to promote both shareholder-centric corporate governance and greater social welfare protection for working families" and referencing both the Dodd-Frank Act and health insurance reforms as specific examples of these efforts). For an earlier argument about the increasing influence of shareholders, see Marcel Kahan & Edward Rock, *Embattled CEOs*, 88 TEXAS L. REV. 987 (2010).

corporate governance theories currently on offer,[64] Bruner discusses the debate over Delaware's status as the leading state of incorporation[65] and the limitations of the passivity thesis of corporate takeovers,[66] as well as the Platonic guardian,[67] team production,[68] and shareholder primacy theories of the proper role of directors.[69] For anyone who is interested in sampling the concerns of recent corporate governance scholarship, *Corporate Governance in the Common-Law World* is a helpful primer.

Given the book's title, one surprise is that the common law tradition plays no direct role in Bruner's thesis. Bruner does speculate that ongoing and historical ties among the four countries may help to explain why all have wound up on the dispersed-ownership side of the governance map and that pulls toward the United States or United Kingdom may shape Australian and Canadian governance. But unlike with La Porta et al., who argue that the nature of the common law process has had a formative influence on the corporate governance of common law nations,[70] the common law does not figure in Bruner's theory. In fact, Bruner's theory posits that factors other than the common law—the nature of social welfare policy and its implementation—are more powerful determinants of shareholder orientation than legal origin. In this sense, his social welfare-based theory can be seen as implicitly rejecting claims that the common law is a key determinant of corporate governance.

To my knowledge, almost the only other recent work arguing that employee-oriented legislation outside of corporate law is essential to understanding the contours of American corporate governance is an

---

64. BRUNER, *supra* note 16, at 111–42.

65. *Id.* at 276–77.

66. *See, e.g.*, BRUNER, *supra* note 16, at 55 (discussing Easterbrook and Fischel, who suggest that weak shareholder rights are acceptable "because management can be sufficiently disciplined through the market for corporate control, which depends critically on free capacity to accept hostile tender offers"); *see also* Frank H. Easterbrook & Daniel R. Fischel, *The Proper Role of a Target's Management in Responding to a Tender Offer*, 94 HARV. L. REV. 1161, 1194 (1981) (proffering the passivity thesis that "managers of target companies should acquiesce when confronted with a tender offer").

67. *See, e.g.*, BRUNER, *supra* note 16, at 55; *see also* Stephen M. Bainbridge, *Director Primacy: The Means and Ends of Corporate Governance*, 97 NW. U. L. REV. 547, 550–51 (2003) (explaining that under the Platonic guardian theory, corporate boards of directors are not "mere agent[s]" of shareholders but are "*sui generis*" bodies "serving as the nexus for the various contracts comprising" a corporation).

68. *See, e.g.*, BRUNER, *supra* note 16, at 57–59; *see also* Margaret M. Blair & Lynn A. Stout, *A Team Production Theory of Corporate Law*, 85 VA. L. REV. 247, 249 (1999) (indicating that team production problems "arise in situations where a productive activity requires the combined investment and coordinated effort of two or more individuals or groups").

69. BRUNER, *supra* note 16, at 60–64.

70. *See* La Porta et al., *Investor Protection*, *supra* note 6, at 8–9 (explaining that "[c]ommon law countries have the strongest protection of outside investors" for various reasons).

insightful article by Adam Winkler.[71] Responding to laments by advocates of stakeholder governance that American corporate governance is too shareholder oriented, Winkler pointed out that corporations are subject to a wide range of noncorporate regulations that have stakeholders in mind.[72] The employment laws regulate collective bargaining and provide protections for employees, for instance, and the environmental laws impel firms to take the environment into account.[73] If we expand our frame of reference beyond corporate law, Winkler argues, the overall system is quite stakeholder oriented.[74]

Bruner's theory can, in a sense, be seen as a dynamic account of some of the same noncorporate laws that Winkler drew attention to. Bruner suggests not only that we need to include noncorporate law in our thinking about corporate governance but that there is a feedback effect between two key areas of regulation: corporate governance and the social welfare system.

II. What Is the Mechanism?

Any theory that makes causal claims about regulatory evolution must be prepared to address two questions. The first is whether the connections that the scholar asserts are real. For Bruner's theory, the question is whether shareholder orientation does indeed vary with the scope of a country's social welfare system. Second, if the relationship is real, the scholar must also marshal evidence that the connection is causal rather than simply a potentially unexplained correlation—here, that some mechanism links an increase or decrease in shareholder orientation with the robustness of a country's social welfare system.

On the first issue, whether the relationship between social welfare and shareholder orientation genuinely does exist, Bruner assembles an impressive amount of evidence. Not surprisingly, given that Bruner is a corporate law scholar, the evidence tilts toward the corporate side of the equation. But he gives an impressive survey of the emergence of a robust social welfare system in the United Kingdom and the absence of comparable protections in the United States, which has less unemployment protection and, until recently, lacked a national healthcare system.[75] The

---

71. Adam Winkler, *Corporate Law or the Law of Business?: Stakeholders and Corporate Governance at the End of History*, 67 LAW & CONTEMP. PROBS. 109 (2004).

72. *See id.* at 111 (expounding that "progressive principles of stakeholder protection" that are "outside of corporate law" are "powerful forces shaping the choices available to corporate management concerning basic operational and organizational decisions").

73. *See id.* (observing that corporate decisions "are made under the mandatory legal rules embodied in employment and labor law, workplace safety law, environmental law, consumer protection law, and pension law").

74. *Id.*

75. *See supra* notes 46–48 and accompanying text.

story in Australia and Canada is more complicated and less stark, but generally consistent. Overall, the connections seem to be real.

A few nagging doubts remain. Several of the most important shareholder-empowering U.K. rules predate the post-World War II expansion of the United Kingdom's social welfare system.[76] The principal development in the past generation was the United Kingdom's adoption of a similarly shareholder-oriented stance toward hostile takeovers starting in the 1960s.[77] These doubts suggest that the relationship between social welfare and shareholder orientation is not pristine, but overall Bruner's case seems strong.

The second issue—the claim that there is a causal relationship between social welfare and corporate governance—is much trickier. If social welfare and shareholder orientation are related, as Bruner claims, we ideally would want to see a direct cause-and-effect relationship between the two. Perhaps after the implementation of strong employee protections we would see a sudden shift in the shareholder orientation of corporate law or perhaps even a direct bargain: in return for strong unemployment benefits, labor leaders agree to strong, shareholder-oriented corporate law reforms. Or, in return for a weakening of collective bargaining protections, investor interests might drop their objections to reforms that would weaken the powers of shareholders in corporate law. Alternatively, the weakening of collective bargaining protections might prompt a backlash against shareholder-centric corporate law rules.[78] How does Bruner fare in providing this kind of evidence?

In some respects, surprisingly well, especially with the United Kingdom. In the United Kingdom, the highly proshareholder takeover rules emerged under the Labour government of Prime Minister Harold Wilson.[79] Labour's embrace of proshareholder rules followed both an initial period of opposition and a steady strengthening of the United Kingdom's social welfare system.[80] Bruner points out, for instance, that "the Redundancy Payments Act of 1965," which significantly enhanced unemployment

76. Arguably more consistent with Bruner's thesis, U.K. governance does not appear to have been especially shareholder oriented (at least in practice) early in the twentieth century. *See, e.g.*, BRIAN R. CHEFFINS, CORPORATE OWNERSHIP AND CONTROL: BRITISH BUSINESS TRANSFORMED 33–40 (2008) (claiming that a shareholder's right to call meetings was established in 1900 but a right to remove directors was not available until 1948).

77. This history is recounted in John Armour & David A. Skeel, Jr., *Who Writes the Rules for Hostile Takeovers, and Why?—The Peculiar Divergence of U.S. and U.K. Takeover Regulation*, 95 GEO. L.J. 1727, 1756–64 (2007).

78. At several points in the book, Bruner suggests that the possibility of backlash may play a causal role in the patterns he detects. *See, e.g.*, BRUNER, *supra* note 16, at 22–23 (describing "backlash" as a key component of the "relatively higher degree of shareholder-centrism exhibited in Australia and the United Kingdom").

79. *Id.* at 151, 160.

80. *See id.* at 147–49.

benefits, "predates by three years the Labour government's reinforcement of an extremely shareholder-centric takeover regime through the creation of the City Code and the City Panel."[81] Although Prime Minister Margaret Thatcher sought to cut back on welfare protections, the combination of fulsome social welfare protections and strong shareholder protections endured;[82] Bruner quotes a recent study finding that "U.K. welfare expenditures remained 'remarkably stable' between 1973 and 1996."[83]

Bruner's survey of developments in Australia is similarly suggestive. When takeovers first emerged, Australia experimented with a U.S.-style approach, which gave the directors of a target corporation discretion to defend against takeovers under some circumstances.[84] But in the early 2000s, Australia shifted direction, creating a Takeover Panel modeled on the United Kingdom and forbidding takeover defenses.[85] The shift, in Bruner's telling, came after the Labour government had put in place a full panoply of welfare protections over a twenty-year period.[86] I will leave it to others to assess whether Bruner's historical description is accurate,[87] but it appears to nicely support his core thesis about social welfare and shareholder orientation.

With Canada, however, the story muddies considerably. Canada has far more extensive social welfare protections than the United States,[88] which suggests that its corporate governance should be more shareholder-oriented. On its face, however, Canadian corporate law seems to protect stakeholders as well as shareholders.[89] This suggests that Canada may combine social welfare protections and stakeholder governance, a combination that should not be sustainable if Bruner's theory is correct. Bruner solves the problem in two ways. First, based on a lengthy analysis

---

81. *Id.* at 159.

82. *See id.* at 160.

83. *Id.* (quoting John Clarke et al., *Remaking Welfare: The British Welfare Regime in the 1980s and 1990s*, *in* COMPARING WELFARE STATES 71, 76 (Allan Cochrane et al. eds., 2d ed. 2001)).

84. *Id.* at 192–93.

85. *Id.*

86. *Id.*

87. It seems a little puzzling that Australia toyed with the U.S. approach at all rather than simply adopting the U.K. approach. Jennifer Hill has suggested that the apparent shift came as a result of the cases first arising in the courts, before the Takeover Panel was set up. *See, e.g.*, Jennifer G. Hill, *Takeovers, Poison Pills and Protectionism in Comparative Corporate Governance* 5–6 (European Corporate Governance Inst., Working Paper No. 168/2010, 2010) *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1704745. For an argument that the mode of regulation—self-regulation in the United Kingdom, courts in the United States—has shaped the diverging U.S. and U.K. approaches to takeovers, see Armour & Skeel, *supra* note 77, at 1767–84.

88. BRUNER, *supra* note 16, at 200.

89. *See id.* at 83, 213 (discussing aspects of Canadian corporate law that make it appear more stakeholder friendly than it actually is).

of the *BCE, Inc.* case, Bruner contends that Canadian corporate law is far less stakeholder oriented in practice than it appears to be.[90] Although the analysis veers perilously close to the domain of special pleading, Bruner's conclusions seem more or less plausible. And it is of course essential to consider the law as it actually functions, not simply the law on the books. (Indeed, I will raise precisely this kind of concern about Bruner's characterization of U.S. law below.)

Second, Bruner characterizes Canada (as well as Australia) as occupying a middle ground between the United States and the United Kingdom, with governance that is more shareholder oriented than the United States but not quite so shareholder-centric as the United Kingdom.[91] Although this characterization seems accurate, it also is more worrisome for the explanatory power of Bruner's theory. If the shareholder and social welfare relationship is a continuum rather than a clear distinction, the theory becomes very difficult to falsify. Country A, which is somewhat shareholder oriented and has somewhat robust social welfare protections, and thus fits the theory, becomes hard to distinguish from Country B, which is somewhat shareholder oriented but has rather weak social welfare protections, or from Country C, which is not especially shareholder oriented but has fairly strong social welfare protection. In Bruner's defense, he does not put Canada in this category. He argues that both Canada and Australia are considerably closer to the United Kingdom than to the United States.[92] But a theory that allows for endless gradations and lacks clear causal relationships may be hard to sustain, given the inevitable messiness of history and the theory's reliance on only four countries as its data points.

The absence of crisp causal connections also raises the question whether omitted factors might further complicate the apparent relationship between shareholder-centrism and social welfare. One obvious candidate for consideration might be antitrust or competition law. In a country that permits concentrated industries, both shareholders and employees might favor a shareholder-centric approach. The United States has traditionally been more aggressive in enforcing competition law than the other three countries,[93] although regulators' easing up on antitrust enforcement in the late 1970s and early 1980s was one of the contributing factors to the U.S. takeover wave of the 1980s.[94]

---

90. Bruner spends six full pages developing his explanation of *BCE, Inc.* *Id.* at 89–95.

91. *Id.* at 77, 200.

92. *Id.* at 176.

93. *See id.* at 154 (describing the prevailing view that "U.K. competition regulation . . . was 'modest' in comparison with U.S. antitrust regulation").

94. George Bittlingmayer, *The Antitrust Emperor's Clothes*, REGULATION, Fall 2002, at 46, 48–49.

III. How Weak Are U.S. Shareholders?

In the next Part, I will consider the possibility that the United States is simply an outlier, while the United Kingdom, Australia, and Canada coherently combine social welfare and shareholder orientation. But let me first raise the question whether the analysis accurately describes the United States and the relationship between corporate law and social welfare in this country.

One question, it seems to me, is whether shareholders really are as weak in the United States as the Bruner thesis suggests. Although U.S. shareholders lack many of the powers seen in the United Kingdom, they have one lever that U.K. shareholders do not: a robust right to sue.[95] Due primarily to the generous rules for compensating plaintiffs' attorneys, shareholders can much more easily sue in the United States than in the United Kingdom.[96] From this perspective, shareholder litigation may supply an *ex post* substitute for the *ex ante* powers U.K. shareholders have. Bruner recognizes this possibility but rejects it.[97] Although shareholder litigation may be a partial substitute, it falls far short of closing the gap. "If greater capacity to sue were truly intended to substitute for strong shareholder governance powers," he argues, "then we might expect to find similarly strong expressions of commitment to shareholders in the articulation of directors' duties."[98] But to the contrary, Bruner concludes "the divergence between the express shareholder-centrism of the U.K. Companies Act and the ambivalent formulation of directors' fiduciary duties in Delaware is every bit as stark as the divergence between the shareholders' governance powers in the two jurisdictions."[99]

I think Bruner is right about this. Even effective *ex post* remedies are unlikely to fully substitute for *ex ante* governance powers, and the efficacy of shareholder litigation is subject to particular doubt.[100] But I am

---

95. *See, e.g.*, John Armour et al., *Private Enforcement of Corporate Law: An Empirical Comparison of the United Kingdom and the United States*, 6 J. EMPIRICAL LEGAL STUD. 687, 688–90, 721–22 (2009) (testing the hypothesis that the United States is more "litigation-friendly" with respect to private enforcement of corporate law and concluding that the United States has a higher "intensity of formal private enforcement[, which] may compensate for [its] modest substantive protections, producing a 'shareholder-friendly' end result").

96. *See, e.g.*, *id.* at 692 ("For instance, various features of civil procedure . . . are more favorable to plaintiffs in the United States. In particular, the facilitation of class actions and the use of contingency fees stimulate entrepreneurial attorneys, whereas the United Kingdom's 'loser pays' fees rule will discourage representative litigation.").

97. BRUNER, *supra* note 16, at 104–05.

98. *Id.* at 105.

99. *Id.*

100. Bruner notes one reason for this doubt: the difficulty of determining litigation's effectiveness in deterring director misbehavior. *Id.* at 101–02. The fraud-on-the-market doctrine—which presumes reliance if the stock in question is actively traded and which is a key feature of most securities litigation—is under serious attack, both in the courts and among

somewhat skeptical about his characterization of U.S. corporate governance for a different reason. Although Delaware's takeover jurisprudence is far less shareholder oriented than the United Kingdom's, as Bruner points out at length,[101] the directors of target corporations are now far less likely to resist a shareholder-benefitting takeover than they were in the early years of the 1980s takeover wave.[102] In the 1990s, directors and managers were increasingly paid in stock and stock options, rather than cash, which made them much less likely to thumb their noses at a lucrative takeover offer.[103] During this same period, shareholders have made much more active use of the governance levers they do have at their disposal. As a result, some commentators have argued that publicly held companies in the United States are run in a highly shareholder-oriented fashion.[104] Delaware doctrine may not be shareholder oriented, the reasoning goes, but the behavior of Delaware corporations is.

If this characterization is accurate, as I believe it is, one possible response might be to predict that this recent shareholder orientation will be accompanied by increasingly robust social welfare protections. The obvious evidence to marshal in support of this thesis is the enactment of healthcare legislation in 2010,[105] which Bruner mentions at several points[106] but does not really explore. One problem with such an account is that it is very hard to see the causal relationship between the new shareholder-centric reality and the recent healthcare reform. Perhaps this is, in part, because the legislation, like the Dodd-Frank reforms enacted shortly thereafter,[107] and which Bruner does discuss,[108] is still too new to put in historical perspective. There is a second, very different problem as well: if the United

---

academics. For an important recent academic critique, see William W. Bratton & Michael L. Wachter, *The Political Economy of Fraud on the Market*, 160 U. PA. L. REV. 69 (2011).

101. *See* BRUNER, *supra* note 16, at 36–42.

102. Marcel Kahan and Ed Rock were the first to draw attention to the implications of this development. Marcel Kahan & Edward B. Rock, *How I Learned to Stop Worrying and Love the Pill: Adaptive Responses to Takeover Law*, 69 U. CHI. L. REV. 871 (2002).

103. *See id.* at 896.

104. *See, e.g.*, *id.* at 899 (stating that managers have now adopted shareholder value maximization as their mantra). Ed Rock has recently suggested that shareholder–manager agency costs may no longer be the most important issue in U.S. corporate law. Edward B. Rock, *Adapting to the New Shareholder-Centric Reality*, 161 U. PA. L. REV. 1907, 1910 (2013). U.S. corporations are run in so shareholder oriented a fashion, at least at present, that shareholder opportunism vis-à-vis creditors is a more relevant risk. *Id.* at 1910–11.

105. Patient Protection and Affordable Care Act, Pub. L. No. 111-148, 124 Stat. 119 (2010) (codified as amended in scattered sections of 26 and 42 U.S.C.).

106. BRUNER, *supra* note 16, at 284–85, 290.

107. Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124 Stat. 1376 (2010) (codified as amended in scattered sections of 2, 5, 7, 12, 15, 18, 22, 26, 28, 31, 42, and 44 U.S.C.).

108. BRUNER, *supra* note 16, at 267–68, 270–72, 280–84. As noted earlier, Bruner focuses primarily on several provisions in the Dodd-Frank Act that give new authority to shareholders. *See supra* note 62 and accompanying text.

States is shareholder-centric too, all four countries line up on the same side of the spectrum. This would suggest that the diversity that motivated the book in the first instance is more illusory than real, although it also would raise the intriguing question of whether countries whose stock ownership is relatively dispersed will inevitably gravitate toward the shareholder-centric, robust social welfare side of the spectrum.

Let me speculate for a moment about this last possibility. The conventional explanation for the United Kingdom's robust social welfare system is that it was spurred in important part by the devastation England suffered from bombing in World War II, as well as the costs of the war in general.[109] Although Australia and Canada did not endure the same hardships, they were influenced by developments in the United Kingdom.[110] The United States, by contrast, charted a different course, adopting more limited social welfare protections and looking to employer-provided protections during a period when U.S. industry was dominant.[111] Perhaps this was unsustainable, and the United States also is headed toward the same shareholder-centrism and robust social welfare system as Bruner identifies in the other three common law nations.

IV. Is the United States Simply an Outlier?

I suggested in the last Part that U.S. corporate law currently may be more shareholder-centric than shareholders' limited formal powers suggest. In this Part, I will consider another possibility. Perhaps the United States is simply peculiar, an odd duck. I pursue this possibility by considering the impact of American federalism and the puzzling weakness of the American social safety net.

A. Federalism and the Limits on Shareholder-Centrism

The leading contemporary account of the political economy of American corporate law identifies federalism and American populism as two of the (mutually reinforcing) political reasons that managers have traditionally been strong and shareholders comparatively weak.[112] Thanks to populism, financial institutions have long been prevented from actively controlling American corporations, both by legal prohibitions on their stock ownership and by strong norms against their exerting control.[113] Managers

109. *See* John Clarke et al., *The Construction of the British Welfare State, 1945–1975*, *in* COMPARING WELFARE STATES, *supra* note 83, at 29, 34–44.

110. *See* BRUNER, *supra* note 16, at 176 (stating that the social welfare models of Australia and Canada resemble, "for broadly similar reasons," that of the United Kingdom, but that the resemblance is due to a unique set of factors).

111. *See id.* at 166–76.

112. *See* ROE, *supra* note 4, at x.

113. *Id.* at 48–49.

have been an important beneficiary of this straitjacket.[114] States' efforts to attract corporate business has further strengthened managers' hands, since managers tend to decide where companies incorporate and where they open a new plant.[115] Both sides in the endless (and now thankfully waning) debate over whether Delaware's dominant share of incorporations reflects a race to the top or a race to the bottom agree that Delaware is acutely sensitive to managers' interests, disagreeing primarily about whether managers' interests are aligned with those of shareholders'.[116]

One thing Delaware does not have is any particular reason to show concern for employees and the robustness of the social welfare system. Very few corporations have a significant number of employees in Delaware, since their headquarters and significant assets are elsewhere.[117] By contrast, nearly 20% of Delaware's annual income depends on the state continuing to keep the managers and/or shareholders of its corporations happy.[118] Delaware's resolution of most corporate governance issues—including takeovers, in stark contrast to both the United Kingdom and Australia—through common law judicial decision making also tends to favor the interests of managers.[119]

This does not refute the Brunerian thesis, of course. Even if Delaware does not have any particular interest in employees, it may nevertheless face pressure to take their interests into account. When the Delaware Supreme Court shifted from a relatively proshareholder approach to hostile takeovers (though still much less proshareholder than the United Kingdom) to a much more manager-oriented standard in 1989,[120] some commentators suggest that it is possible to attribute the shift to concerns that Congress might enact

---

114. *See id.* at 5 (describing the rise of "professional managers").

115. This feature has been discussed more by sociologists than by corporate law scholars. *See, e.g.*, Jacob S. Hacker & Paul Pierson, *Business Power and Social Policy: Employers and the Formation of the American Welfare State*, 30 POL. & SOC'Y 277, 290 (2002) (discussing concerns amongst state social policy reformers that reform efforts would discourage business development).

116. For the founding articles in this debate, see William L. Cary, *Federalism and Corporate Law: Reflections Upon Delaware*, 83 YALE L.J. 663 (1974), and Ralph K. Winter, Jr., *State Law, Shareholder Protection, and the Theory of the Corporation*, 6 J. LEGAL STUD. 251 (1977). *See also* Armour & Skeel, *supra* note 77, at 1765.

117. Kent Greenfield, *Democracy and the Dominance of Delaware in Corporate Law*, 67 LAW & CONTEMP. PROBS. 135, 136 (2004).

118. *See* ROBERTA ROMANO, THE GENIUS OF AMERICAN CORPORATE LAW 6–8 (1993) (noting that up to 17.7% of Delaware's total tax revenue comes from franchise taxes and noting that this number is very high compared to other states).

119. For a much more detailed analysis of this point in the takeover context, see Armour & Skeel, *supra* note 77, at 1780–84.

120. *See* Paramount Commc'ns, Inc. v. Time Inc., 571 A.2d 1140, 1142 (Del. 1989) (deferring to Paramount's managers' decision not to accept Time's offer because the offer was reasonably perceived by Paramount's board as a threat).

legislation discouraging takeovers, thus intruding on Delaware's turf.[121] Employees and their representatives were one of the groups urging congressional action.[122] If Delaware's shift was an effort to blunt the campaign for federal intervention by rendering it unnecessary, employees may thus have contributed indirectly to the Delaware ruling. Roe has offered a similar analytic on Delaware–Washington interaction: Delaware has limited reason to promote employee and other social welfare interests in the corporation, as it is boards and shareholders who decide whether to incorporate in Delaware.[123] If American corporate law were fully made in Washington, stakeholder interests would be more prominently involved.[124] Shareholders could put up with a Delaware tilt to managers (and an occasional venting of stakeholder interests), as the results for shareholders if corporate law were made in Washington might not be to their liking.[125]

Although Delaware's comparative disinterest in employees does not undermine Bruner's thesis, it does suggest that managers, whom Delaware clearly does attend to,[126] should be a central part of any story about American corporate law. Managers with authority may do a little extra for employees. The most obvious explanation for the manager-centrism of American corporate law, as compared to the United Kingdom's shareholder orientation, is American federalism. More nuanced historical factors, such as the contractual and self-regulatory traditions of U.K. corporate law and the more direct governmental role in the life of corporations in the United States may also have played a part.[127] Whatever the mix of factors, managers lie at the heart of corporate law in the United States but not in the United Kingdom. The general weakness of shareholder rights seems more closely related to managerial influence than to the limitations of the U.S. social welfare system. That is, Bruner's thesis needs to explain why American managerialism leads to both weaker shareholder power (than the United Kingdom) and some managerial noblesse oblige, in that the managers at times do things that are in employees' interest. Or, more subtly, perhaps managers, to maintain their authority in the firm over the long run, need to have political and social allies, such as employees and stakeholders.

---

121. *See* Jeffrey N. Gordon, *Corporations, Markets, and Courts*, 91 COLUM. L. REV. 1931, 1964–66 (1991) (explaining the argument that Delaware courts might have taken into consideration possible congressional action in making their ruling).

122. *See id.* at 1965.

123. Mark J. Roe, *Delaware's Politics*, 118 HARV. L. REV. 2491, 2500–02 (2005).

124. *Id.* at 2502–04.

125. *See id.* at 2515–16.

126. *See supra* notes 114–16 and accompanying text.

127. *See, e.g.*, Armour & Skeel, *supra* note 77, at 1767–84 (describing the importance of self-regulation to differences in U.S. and U.K. takeover law).

As astute an observer as he is, Bruner is well aware of this. In the final major chapter of the book, he increasingly relies on coalition theory—which posits that corporate governance is shaped by shifting coalitions of employees, managers, and shareholders[128]—to assess recent developments such as the U.K. government's retrenchment on social welfare protections and the United States' augmenting of social welfare through healthcare reform and inclusion of shareholder-oriented provisions in the Dodd-Frank Act.[129] This has the benefit of making managers a much more central factor in the story, but it complicates Bruner's own core story about social welfare and shareholder power as the key features of corporate governance, with the structure of the former being the primary determinant of the latter.

B. Why Is the U.S. Safety Net So Weak?

Perhaps the real puzzle is not why or whether the United States is so much less shareholder oriented than the other three common law countries. Perhaps the real puzzle is America's social safety net. Developed non-common law countries tend to have extremely robust social safety nets (in some cases probably too well developed, but that is another story).[130] Although the social safety nets have been somewhat more contested in the United Kingdom, Canada, and Australia, each has much more protection than the United States.[131] From this perspective, the most puzzling feature of the social welfare–shareholder centrism equation is the weakness of the U.S. social welfare system. Why is the U.S. safety net so much weaker than everyone else's?

Although social welfare experts would offer a more nuanced account, four distinctively American factors seem to me to figure in the contrast between the United States and other common law countries. The first is the federalism concerns I noted in the previous subpart. The ability for businesses to move out of states that require generous provisions for employees, or impose other costs, acts as a constraint on states' abilities to provide generous benefits. Given states' interests in local control, this factor may also translate to some extent to limits on federal programs.

---

128. BRUNER, *supra* note 16, at 130–31.

129. *See id.* at 265–67 (discussing coalitions and their effect on the post-crisis reform efforts and concluding that some reforms that would shift power to shareholders did make it into the Dodd-Frank Act); *id.* at 280–84 (noting that shifting dynamics have tended to promote greater social welfare protection (such as the health care law) in the United States); *id.* at 286 (noting the diverging postcrisis response in the United Kingdom, which has weakened stakeholder's interests).

130. *See, e.g.*, Liz Alderman, *Why Denmark Is Shrinking Its Social Safety Net*, ECONOMIX, N.Y. TIMES (Aug. 16, 2010, 12:12 PM), http://economix.blogs.nytimes.com/2010/08/16/why-denmark-is-shrinking-its-social-safety-net/ (discussing Denmark's "expensive, generous welfare state" and why it is shrinking).

131. *See* BRUNER, *supra* note 16, at 143.

The second is race.[132] The American experience with slavery and Jim Crow segregation complicates the politics of American social welfare in ways that sharply distinguish the United States from the other three countries under consideration. The most obvious beneficiaries of many forms of welfare legislation in the twentieth century would have been poor whites and poor blacks. Yet the racial divisions of the Jim Crow era and after made it nearly impossible to create a coalition consisting of both poor whites and poor blacks. This has created a very different politics of welfare legislation than is the norm in the United Kingdom, Canada, or Australia.

The third factor is ideological. The ideology of self-reliance that was once associated with the American Frontier has a stronger pull in the United States than elsewhere—especially as compared to European countries. This is reflected in the insistence that Americans who receive welfare benefits work for those benefits.[133] Although legal scholars periodically insist that every American should be entitled to a minimum level of income,[134] this contention has never seemed compelling to most Americans.[135]

The final factor is related to the third. America's bankruptcy discharge is considerably more generous than the discharge in most other countries, including the other common law countries. In the United States, a financially troubled consumer debtor nearly always has access to an immediate discharge of her debts based on the premise that both the debtor and her creditors are better off if an "honest but unfortunate"[136] debtor sheds her debts and is given a second chance.[137] In the United Kingdom, by contrast, debtors often receive a less generous discharge, and sometimes no discharge at all.[138] The American approach does less to discourage excessive borrowing, but it also appears to facilitate entrepreneurship, since an entrepreneur is not saddled by the obligations of a failed initial venture.[139] More importantly, it also serves as a partial proxy for the social

132. Thanks go to Mark Roe for encouraging me to consider the implications of race.

133. *See, e.g.*, THEDA SKOCPOL, THE MISSING MIDDLE: WORKING FAMILIES AND THE FUTURE OF AMERICAN SOCIAL POLICY 24–27 (2000).

134. *See* BRUCE ACKERMAN & ANNE ALSTOTT, THE STAKEHOLDER SOCIETY 21–44 (1999) (proposing that every citizen be given a sum of money when they reach maturity).

135. Dylan Matthews, *Obama Doesn't Want to Just Write Welfare Recipients Checks. But What If We Did?*, WONKBLOG, WASH. POST (Aug. 8, 2012, 3:54 PM), http://www.washingtonpost.com/blogs/wonkblog/wp/2012/08/08/obama-doesnt-want-to-just-write-welfare-recipients-checks-but-what-if-we-did/ (describing the idea of a universal basic income as unpopular).

136. Local Loan Co. v. Hunt, 292 U.S. 234, 244 (1934).

137. *See, e.g.*, DAVID A. SKEEL, JR., DEBT'S DOMINION: A HISTORY OF BANKRUPTCY LAW IN AMERICA 100 (2001).

138. *See id.* at 2, 238.

139. *See, e.g.*, John Armour & Douglas Cumming, *Bankruptcy Law and Entrepreneurship*, 10 AM. L. & ECON. REV. 303, 337 (2008) (stating that a generous bankruptcy discharge is associated with greater levels of self-employment in a sample of developed countries); Wei Fan & Michelle J. White, *Personal Bankruptcy and the Level of Entrepreneurial Activity*, 46 J.L. &

welfare benefits that are not available in the United States and accords with the ideological commitment to self-reliance rather than governmental support for those who are struggling.

The key question with the American preference for bankruptcy rather than robust social welfare protections is whether it is likely to endure, or whether the United States will eventually fall more closely in line with other developed nations. Predicting is always hazardous, but my guess is that the distinctions will remain. To be sure, a partial convergence may be underway, with Western Europeans retrenching somewhat on issues like the length of the workweek and age of retirement,[140] while the United States has added universal healthcare. But the structural and political factors that distinguish the United States from other common law countries, and from continental Europe, have not disappeared.

The Brunerian thesis easily accommodates these observations. If the American social welfare system remains weaker than that of other countries, the thesis would predict that shareholder powers will remain more limited in the United States. But the fit seems imperfect in at least two respects. First, as discussed earlier, U.S. governance appears to be much more shareholder oriented in practice than seems the case if we only consider Delaware doctrine. This shareholder-centrism seems real, and it does not seem connected in any discernible way with the recent, still greatly contested expansion of the U.S. social welfare system.

Second, Delaware corporate law has long oscillated between an emphasis on shareholders' interests, on the one hand, and a less shareholder-oriented emphasis on the corporate entity as a whole, on the other.[141] If shareholder-centrism is tightly linked to the strength of a country's social welfare system, we might expect to see some connection between these oscillations and changes in the social welfare system. But there do not seem to be any evident connections between the two.

Conclusion

Over the past several months, I have described the Brunerian thesis to a wide range of corporate law scholars—most from the United States, but some from the other common law countries as well—and asked them whether they find the thesis persuasive. Nearly every one has given one of

---

ECON. 543, 563 (2003) (claiming that there is more entrepreneurial activity in U.S. states that permit greater exemptions in bankruptcy).

140. *See, e.g.*, Ian Traynor, *Eurozone Demands Six-Day Week for Greece*, GUARDIAN, Sept. 4, 2012, http://www.theguardian.com/business/2012/sep/04/eurozone-six-day-week-greece; Robert Winnett, *Queen's Speech: Retirement Age Delayed to 67 to Fund Flat-Rate Pension Scheme*, TELEGRAPH, May 8, 2013, http://www.telegraph.co.uk/news/politics/queens-speech/10043364/Queens-Speech-retirement-age-delayed-to-67-to-fund-flat-rate-pension-scheme.html.

141. The classic account is William T. Allen, *Our Schizophrenic Conception of the Business Corporation*, 14 CARDOZO L. REV. 261, 264–75 (1992).

two responses, at least initially. Some explain why the thesis cannot be correct, and the others conclude that its insights are obvious. The reactions sound suspiciously similar to a definition I once heard of a successful scholarly presentation. The best papers and presentations, I was told, are the ones in which the audience is initially convinced that the scholar's thesis is completely wrong, and eventually concludes that it is obvious.

Bruner's claim that strongly shareholder-oriented governance—which sniffs of Wall Street rather than Main Street—is associated with robust social welfare protections—which sounds much more like Main Street—is both counterintuitive and plausible. Even if Bruner had not marshaled extensive supporting evidence, it would be a thesis that corporate law scholars, and perhaps social welfare experts as well, would need to grapple with. The elaborately detailed case that Bruner presents adds to its importance.

As my quibbles suggest, I am not sure whether Bruner is right. The mechanism that links the two halves of the thesis together is somewhat unclear, with connections that are more indirect than direct. But there is an undeniable logic to his thesis, and I do not believe that the connections he identifies are imaginary. Any future scholar who purports to provide an explanation of comparative corporate governance will need to consider how social welfare legislation may be shaping what he or she sees in the corporate governance of a particular country. It is hard to imagine a more compelling demonstration that U.S. corporate law scholars need not only to continue looking outside the United States but also to venture beyond the narrow confines of corporate law.

Notes

Improving Statutory Deadlines on Agency Action: Learning from the SEC's Missed Deadlines Under the JOBS Act*

The Securities and Exchange Commission (SEC) has received a great deal of negative attention, both from Congress and from the public, for missing deadlines imposed by various provisions of the JOBS Act. This Note argues that these missed deadlines were at least partially due to Congress's problematic use of statutory deadlines.

This Note begins by discussing the Administrative Procedure Act (APA), which sets forth the rulemaking procedures that an agency must follow. It also explains the legal implications of statutory deadlines, describing how they can increase the likelihood of a successful claim against an agency for unreasonable delay, affect a court's evaluation of agency compliance with the APA, and affect "arbitrary and capricious" review.

This Note then describes why statutory deadlines are used—to reduce regulatory delay, to align agency decision making with legislative intent, and to give Congress an easy way to narrow agency discretion in areas where it does not have expertise. It also addresses the problems created by their use, explaining how deadlines reduce agency flexibility, exacerbate resource constraints, can result in decreased quality of rulemaking, and can be used by Congress as a political tool.

Using these purposes and problems as a framework, this Note analyzes the use of statutory deadlines in the context of the SEC. It concludes that while deadlines can accelerate regulatory action by the SEC and align the SEC's rulemaking action with legislative intent, the SEC's failure to meet its deadlines is at least partially due to Congress's problematic use of deadlines. It argues that, in the context of the SEC, statutory deadlines can have a positive effect on agency action, as long as they are not used in a problematic way.

Lastly, this Note makes three recommendations for improving Congress's use of deadlines. First, Congress should expressly indicate the priority of deadlines, both within a statutory scheme and between statutory schemes

---

* I would like to thank William Stutts for his guidance and feedback on this Note. In addition, I would like to thank my wonderful family and friends—especially Mom, Dad, and Christina—for all of their love and support. I am also grateful to the editors of the *Texas Law Review* for their hard work and excellent editing.

imposing deadlines. Second, Congress should hold itself accountable for its legislative decisions—even when it delegates rulemaking to agencies—by setting more realistic deadlines and expressly waiving the notice-and-comment requirements when it feels that the need for quick agency action outweighs the benefits of APA procedure. Lastly, Congress should increase resource appropriations in line with increases in responsibility imposed by statutory deadlines. To do this, Congress could draft deadlines with specific checkpoints at which the Appropriations and Oversight Committees would assess the agency's progress and appropriate additional resources as required. In the context of the SEC, Congress could also consider allowing the agency to obtain resources through self-funding.

I. Introduction

On April 5, 2012, President Obama signed into law the Jumpstart Our Business Startups (JOBS) Act.[1] The Act—passed with wide bipartisan support—attempts to spur the economic growth of small-to-midsize businesses by making funding more easily available to entrepreneurs.[2] To do so, the Act amended several federal securities laws, including adding an exemption to the Securities Act of 1933 for "crowdfunding."[3]

The JOBS Act required the Securities and Exchange Commission (SEC) to implement rules before certain provisions became effective.[4] The SEC failed to comply with the deadlines for issuing these rules. For instance, § 201(a) required the SEC to promulgate rules to eliminate the prohibition against general solicitation in certain securities options within ninety days of the Act's enactment[5]—making the deadline July 4, 2012. The SEC issued a proposed rule, rather than an interim-final rule, on August 29, 2012, seeking public comment.[6] Although the public comment period ended in October 2012,[7] the SEC did not issue a final rule until July 10, 2013.[8]

The SEC's failure to comply with these deadlines has received a negative reaction from some members of Congress. For instance, Representative Patrick McHenry described the SEC's failure to meet the deadline imposed by § 201(a) as a "reflection of [SEC Chairman Mary Schapiro's] ideological opposition" to the Act and as a failure to follow the law.[9] Moreover, the House Oversight and Government Reform Committee posted a video entitled *SEC & the JOBS Act: Just Do Your Job*, which

---

1. Jumpstart Our Business Startups Act, Pub. L. No. 112-106, 126 Stat. 306 (2012) (codified in scattered sections of 15 U.S.C.).

2. President Barack Obama, Remarks by the President at JOBS Act Bill Signing (Apr. 5, 2012), *available at* http://www.whitehouse.gov/the-press-office/2012/04/05/remarks-president-jobs-act-bill-signing.

3. § 302, 126 Stat. at 315–21.

4. *Information Regarding the Use of the Crowdfunding Exemption in the JOBS Act*, U.S. SEC. & EXCHANGE COMMISSION (Apr. 23, 2012), http://www.sec.gov/spotlight/jobsact/crowdfundingexemption.htm.

5. § 201(a), 126 Stat. at 313–14.

6. Press Release, U.S. Sec. & Exch. Comm'n, SEC Proposes Rules to Implement JOBS Act Provision About General Solicitation and Advertising in Securities Offerings (Aug. 29, 2012), *available at* http://www.sec.gov/news/press/2012/2012-170.htm.

7. *See id.* (announcing, on August 29, 2012, that the SEC could seek public comment for thirty days).

8. Press Release, U.S. Sec. & Exch. Comm'n, SEC Approves JOBS Act Requirement to Lift General Solicitation Ban (July 10, 2013), *available at* http://www.sec.gov/News/PressRelease/Detail/PressRelease/1370539707782#.UidbRhaJL9Q.

9. Letter from Representative Patrick McHenry, Chairman, House Subcomm. on TARP, Fin. Servs. & Bailouts of Pub. & Private Programs, to Mary Schapiro, Chairman, U.S. Sec. & Exch. Comm'n (Aug. 16, 2012), *available at* http://oversight.house.gov/wp-content/uploads/2012/08/2012-08-16-PMC-to-Schapiro-SEC-general-solicitation-due-8-30.pdf.

asserted that Chairman Schapiro delayed promulgating the rules to protect her own legacy.[10]

This delay stands out sharply in the current environment, in which the public has anxiously awaited the promulgation of the rules regulating crowdfunding.[11] As one commentator noted, "The buzz surrounding equity-based crowdfunding has died down since the JOBS Act entered the purgatory of SEC rulemaking."[12] There is also a concern that online scammers may be "preying upon those entrepreneurs less familiar with the Act's requirements," who may believe that crowdfunding is currently legal.[13]

Despite the negative attention that the SEC has received, the agency's inability to meet the deadlines imposed by the JOBS Act is not entirely its fault. The thesis explored below is that the SEC's missed deadlines reflect a deeper problem with the use of statutory deadlines by Congress. Part II briefly reviews the Administrative Procedure Act (APA) and provides a description of the legal implications of the use of deadlines. Part III then introduces the concept of statutory deadlines by explaining their purpose and describing problems inherent in their use. Part IV analyzes the use of deadlines in the context of the SEC. Part V discusses ways in which Congress can improve its use of deadlines with regard to the SEC and more broadly.

II. The APA and the Legal Implications of Missed Deadlines for Regulatory Actions

A. A Brief Overview of the APA

Although administrative agencies have been a part of our government structure since its founding, their role increased dramatically as part of the governmental and legislative changes adopted in the context of the Great

---

10. *SEC & the JOBS Act: Just Do Your Job*, COMMITTEE ON OVERSIGHT & GOV'T REFORM (Dec. 14, 2012), http://oversight.house.gov/release/video-release-sec-the-jobs-act-just-do-your-job/.

11. *See* J.D. Harrison, *Crowdfunding Delays, SEC Silence Spark Hostility on Capitol Hill*, WASH. POST, Apr. 8, 2013, http://articles.washingtonpost.com/2013-04-08/business/38369133_1_non-accredited-investors-crowdfunding-entrepreneurs (describing how the delay in issuing rules related to crowdfunding has frustrated start-up "entrepreneurs and potential investors, many of whom are eager to link up through the new online portals" and entrepreneurs who are building crowdfunding sites "who jumped at the opportunity when lawmakers authorized crowdfunding but are still waiting to set their new ventures in motion").

12. J.J. Colao, *Steve Case: Crowdfunding Will Augment—Not Replace—Venture Capital*, FORBES (Mar. 22, 2013, 11:37 AM), http://www.forbes.com/sites/jjcolao/2013/03/22/steve-case-crowdfunding-will-augment-not-replace-venture-capital.

13. N. AM. SEC. ADMINS. ASS'N, SMALL BUSINESS ADVISORY: CROWDFUNDING (2013), http://www.nasaa.org/wp-content/uploads/2012/06/NASAA_Advisory_Crowdfunding_issuers.pdf.

Depression.[14] As part of the New Deal, "Congress created a myriad of new regulatory and benefit programs and created new administrative agencies to implement them."[15] Before these changes, the laws regulating agencies were restricted to common law causes of action and the statutes establishing each agency's existence.[16] In the later years of the New Deal, it became clear that these laws were inadequate for regulating the growing administrative state and that a new, more cohesive framework was necessary.[17] Congress established such a framework in the APA, passed in 1946.[18]

The APA regulates agencies by dictating the procedures agencies must follow[19] and by establishing a cause of action for review of agency action.[20] Section 553 sets out the procedural requirements for agency rulemaking.[21] First, it requires the agency to publish notice of a proposed rule in the Federal Register, including the "substance of the proposed rule or a description of the subjects and issues involved."[22] The agency must then give the public an opportunity to respond to the proposed rule for at least thirty days.[23] The purpose of this rule is to "give interested persons an opportunity to participate in the rule making."[24] After evaluating the comments, an agency can then issue a final rule and must do so at least thirty days before the rule is to become effective.[25]

The section also establishes a "good cause" exception, stating that the notice-and-comment procedure is not required "when the agency for good cause finds . . . that notice and public procedure . . . are impracticable, unnecessary, or contrary to the public interest."[26] If an agency satisfies the good cause standard, it can promulgate rules without engaging in the notice-and-comment procedure required by the APA.[27] These rules—referred to as "interim rules"—are effective immediately.[28] Additionally,

14. Richard E. Levy & Robert L. Glicksman, *Agency-Specific Precedents*, 89 TEXAS L. REV. 499, 501–03 (2011).

15. *Id.* at 503.

16. *Id.* at 503–04.

17. Paul A. Dame, Note, *Stare Decisis,* Chevron*, and* Skidmore*: Do Administrative Agencies Have the Power to Overrule Courts?*, 44 WM. & MARY L. REV. 405, 409 (2002).

18. Administrative Procedure Act, Pub. L. No. 79-404, 60 Stat. 237 (1946) (codified as amended in scattered sections of 5 U.S.C.).

19. *See* Administrative Procedure Act, 5 U.S.C. §§ 551–59 (2012).

20. *Id.* §§ 701–706.

21. *Id.* § 553.

22. *Id.* § 553(b).

23. *Id.* § 553(c)–(d).

24. *Id.* § 553(c).

25. Mark D. Shepard, *The Need for an Additional Notice and Comment Period When Final Rules Differ Substantially from Interim Rules*, 1981 DUKE L.J. 377, 379–80.

26. 5 U.S.C. § 553(b)(B).

27. Shepard, *supra* note 25, at 380.

28. *Id.* at 380–81.

an agency may adopt an "interim-final rule," which "become[s] effective without prior notice and public comment and that invite[s] post-effective public comment."[29] The agency will then modify the rule in light of any post-effective comments and issue a final rule.[30]

In addition, the APA creates causes of action under which individuals can seek review of agency actions.[31] First, § 706 states that a court can "compel agency action unlawfully withheld or unreasonably delayed."[32] This allows an individual to bring a claim if an agency fails to act or delays action. Courts, however, are often reluctant to compel agency action for unreasonable delay.[33] Second, the APA allows the court to hold unlawful any agency action that is "without observance of procedure required by law."[34] In addition, a court can also review the substance of the agency action and hold it unlawful if it finds the action is "arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law."[35] In determining whether an agency's action was arbitrary and capricious, "a court must consider whether the decision was based on a consideration of the relevant factors and whether there was a clear error of judgment."[36]

B. Legal Implications of Deadlines

The inclusion of deadlines in legislation has various impacts on the judicial review of an agency's rulemaking action. First, it may increase the likelihood that a court will compel agency action as a result of unreasonable delay. Statutory deadlines may also affect a court's evaluation of an agency's procedure in issuing rules and may affect a court's determination of whether an action is arbitrary and capricious. In addition, Congress may respond to an agency's failure to meet deadlines by enacting legislation that reduces the agency's budget or imposes additional deadlines on agency action.

---

29. Michael Asimow, *Interim-Final Rules: Making Haste Slowly*, 51 ADMIN. L. REV. 703, 704 (1999) (emphasis omitted).

30. *Id.*

31. 5 U.S.C. § 702 ("A person suffering legal wrong because of agency action, or adversely affected or aggrieved by agency action . . . is entitled to judicial review thereof.").

32. *Id.* § 706(1).

33. *See* Norton v. S. Utah Wilderness Alliance, 542 U.S. 55, 64 (2004) ("[A] claim under § 706(1) can proceed only where a plaintiff asserts that an agency failed to take a *discrete* agency action that it is *required to take*."); Jacob E. Gersen & Anne Joseph O'Connell, *Deadlines in Administrative Law*, 156 U. PA. L. REV. 923, 951 (2008) (noting that courts have "extremely limited jurisdiction" to compel action that is unreasonably delayed under § 706(1)).

34. 5 U.S.C. § 706(2)(D).

35. *Id.* § 706(2)(A).

36. Motor Vehicle Mfrs. Ass'n of U.S., Inc. v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 30–31 (1983).

1. Compelling Agency Action Unlawfully Withheld.—First, a missed deadline increases the likelihood that a third party can succeed in a claim against an agency for withholding action or unreasonable delay.[37] In *Norton v. Southern Utah Wilderness Alliance*,[38] the Supreme Court indicated that a statutory deadline requiring an agency to promulgate rules by a certain date could qualify as a discrete action the agency was required to take.[39] Several courts have compelled agency action under § 706(1) when a statutory deadline was in place.

For instance, in *Forest Guardians v. Babbitt*,[40] the Tenth Circuit stated that when Congress sets forth a deadline by which the agency must act, "[t]he agency must act by the deadline," and if it fails to do so, "a reviewing court must compel the action unlawfully withheld."[41] The court then held that the Secretary of the Interior must be ordered to comply with a deadline imposed by the Endangered Species Act of 1973 for designating the critical habitat of the Rio Grande silvery minnow, which it had failed to meet.[42] The court distinguished this statutorily required action, which involved a statutory deadline, from statutes that did not set out a specific deadline.[43] In those cases, the court noted, "§ 706 leaves in the courts the discretion to decide whether agency delay is unreasonable."[44] Other courts have made similar distinctions based on the presence of statutory deadlines.[45]

2. Observance of Procedural Requirements.—In addition, deadlines can affect the procedure used by an agency in promulgating rules and the way courts evaluate the agency's procedure under the APA.[46] As discussed above, § 553 of the APA sets out the procedure required for agency

37. *See* Gersen & O'Connell, *supra* note 33, at 952 ("Deadlines stand out as one of the few areas where courts will compel agencies to act despite multiple demands on their resources.").

38. 542 U.S. 55 (2004).

39. *See id.* at 71 (holding that a land use plan, "[q]uite unlike a specific statutory command requiring an agency to promulgate regulations by a certain date," cannot be the basis for suit under § 706(1)).

40. 174 F.3d 1178 (10th Cir. 1999).

41. *Id.* at 1190.

42. *Id.* at 1181, 1182 & n.5, 1193.

43. *Id.* at 1190.

44. *Id.*

45. *See, e.g.*, Home Builders Ass'n of Greater Chi. v. U.S. Army Corps of Eng'rs, 335 F.3d 607, 616 (7th Cir. 2003) ("Among the circuits that have considered the question, the consensus is that agency delay in face of a clear statutory duty (but in the absence of a statutory deadline) must be 'egregious' before it can convert agency inaction into a final action reviewable under the APA or warrant mandamus."); Cobell v. Norton, 240 F.3d 1081, 1096 (D.C. Cir. 2001) ("An agency's own timetable for performing its duties in the absence of a statutory deadline is due 'considerable deference.'").

46. *See* Gersen & O'Connell, *supra* note 33, at 956 (arguing that "agencies often forego notice and comment rulemaking . . . for deadline-driven actions").

rulemaking.[47] Moreover, under § 706 a court must set aside agency actions that are not in "observance of procedure required by law."[48] Thus, a court can set aside rules promulgated without the notice and comment required by § 553.[49] Deadlines, however, can make it difficult or impossible for an agency to comply with the required notice-and-comment procedure.[50] For this reason, agencies may choose to issue an interim-final rule on the grounds that it has good cause to forgo the notice-and-comment procedure.[51]

However, an agency takes a "considerable risk [by relying] on the good cause exception to adopt interim-final rules under a statute that imposes tight implementation deadlines."[52] While a statutory deadline may be an exigent circumstance that establishes a good cause exception,[53] some courts have held that "strict congressionally imposed deadlines, without more, by no means warrant invocation of the good cause exception."[54] This reluctance stems from a fear that "agencies might wait until the eleventh hour to issue rules, rather than organize their procedures to allow notice and comment within the time allotted."[55] Thus, courts typically use a multifactor analysis to determine whether a deadline constitutes good cause.[56] For instance, the D.C. Circuit has exempted an agency from the requirements of § 553 when "congressional deadlines are very tight and where the statute is particularly complicated."[57] In contrast, courts have been reluctant to exempt agency action where the agency had adequate time

47. *See supra* notes 21–23 and accompanying text.

48. Administrative Procedure Act, 5 U.S.C. § 706(2)(D) (2012).

49. *See id.* § 553(b)(B) (providing a good cause exception to the notice-and-comment requirement).

50. *See* Asimow, *supra* note 29, at 742 ("Often, agencies simply cannot complete work on the rules within the deadline date.").

51. *See id.* (noting that "[i]nterim-final rules are often employed where Congress has adopted a new statute that requires an agency to adopt implementing rules under time constraints").

52. *Id.*

53. *See id.* at 719–20 (explaining that to demonstrate good cause, "the agency must demonstrate the presence of exigent circumstances" and that "an imminent implementation deadline imposed by a statute . . . may qualify" (footnote omitted)).

54. Petry v. Block, 737 F.2d 1193, 1203 (D.C. Cir. 1984).

55. Ellen R. Jordan, *The Administrative Procedure Act's "Good Cause" Exemption*, 36 ADMIN. L. REV. 113, 136 (1984).

56. Gersen & O'Connell, *supra* note 33, at 958.

57. Methodist Hosp. of Sacramento v. Shalala, 38 F.3d 1225, 1226, 1236–37 (D.C. Cir. 1994) (holding that use of the good cause exception was justified when the Department of Health and Human Services was required to prepare "regulations to implement a complete and radical overhaul of the Medicare reimbursement system" within five months of the statute's enactment); *accord* Phila. Citizens in Action v. Schweiker, 669 F.2d 877, 880, 885 (3d Cir. 1982) (finding that a deadline requiring the promulgation of rules within forty-nine days justified the application of the good cause exception).

for a notice-and-comment period[58] or when the agency failed to comply with the deadline despite forgoing a notice-and-comment period.[59]

Thus, while courts may find that a statutory deadline constitutes good cause for forgoing the procedural requirements of the APA, whether it will do so depends on the specific facts of the case. This may require an agency to choose between complying with the requirements of the APA and risking that a court will find that its statutory deadline was not good cause to forego the notice-and-comment period. Congress, however, can reduce this risk by "authoriz[ing] or even requir[ing] an agency to adopt interim-final rules in order to set a new regulatory scheme in motion quickly."[60]

3. Substantive Review of Agency Rulemaking.—Lastly, the imposition of statutory deadlines can change a court's substantive review by affecting whether an action is given *Chevron*-style deference and whether an action is considered arbitrary and capricious.[61] In *Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc.*,[62] the Supreme Court set out a framework for determining whether a court should give deference to an agency's interpretation of a statute.[63] Under the *Chevron* framework, a court reviewing an agency's interpretation of a statute begins with "Step One," where the court determines "whether Congress has directly spoken to the precise question at issue."[64] If the intent of Congress is clear, the court and the agency "must give effect to the unambiguously expressed intent of Congress."[65] If the statute does not express Congress's intent as to a specific issue, the court proceeds to "Step Two," where it determines "whether the agency's answer is based on a permissible construction of the statute."[66] More recently, the Supreme Court has recognized another step in

58. *See, e.g.*, Kollett v. Harris, 619 F.2d 134, 145–46 (1st Cir. 1980) (finding that compliance with a deadline was not good cause for failing to meet the notice-and-comment requirement where there were fourteen months between the passage of the legislation and the expiration date and the agency provided no explanation of why it could not meet the requirements); U.S. Steel Corp. v. EPA, 595 F.2d 207, 213 & n.13 (5th Cir. 1979) (rejecting a good cause argument when the EPA had six months before the statutory deadline).

59. *See* U.S. Steel Corp., 595 F.2d at 213 ("[I]t is clear that the EPA did not regard the statutory deadline as sacrosanct, since the nonattainment list was not published until a full month after the deadline."); Sharon Steel Corp. v. EPA, 597 F.2d 377, 378, 380–81 (3d Cir. 1979) (holding that the Administrator of the Environmental Protection Agency did not have good cause to forego a notice-and-comment period in issuing a final rule determining the status of air quality in certain counties based on the national ambient air quality standards when he still failed to meet the statutory deadline).

60. *See* Asimow, *supra* note 29, at 712.

61. Gersen & O'Connell, *supra* note 33, at 960–63.

62. 467 U.S. 837 (1984).

63. *Id.* at 842–44.

64. *Id.* at 842.

65. *Id.* at 842–43.

66. *Id.*

the *Chevron* framework—christened "Step Zero"—that addresses whether a reviewing Court should apply the *Chevron* framework.[67] Developed over several cases,[68] Step Zero recognizes that "administrative implementation of a particular statutory provision qualifies for *Chevron* deference when it appears that Congress delegated authority to the agency generally to make rules carrying the force of law, and that the agency interpretation claiming deference was promulgated in the exercise of that authority."[69]

A deadline can affect whether an agency action is given *Chevron* deference in two ways. First, because a deadline may cause an agency to forego notice-and-comment procedure, it could be cause for the court to find, at Step Zero, that the action is not entitled to *Chevron* deference.[70] However, the lack of notice-and-comment procedure is not dispositive.[71] Moreover, if the court finds that the agency properly avoided the notice-and-comment procedure,[72] the agency action should satisfy Step Zero.[73] Second, statutory deadlines may make it more likely that a court will find that Congress's intent is unambiguous under *Chevron* Step One with respect to timing.[74] Thus, statutory deadlines can "make it easier for the reviewing court to find related language unambiguous and to strike down agency attempts to modify" the timing of the agency action.[75]

A deadline can also affect arbitrary and capricious review, although what that effect will be is not clear.[76] As discussed above, the court must determine whether the agency's decision was based on "relevant factors"

---

67. Cass R. Sunstein, Chevron *Step Zero*, 92 VA. L. REV. 187, 191 (2006).

68. *See* Barnhart v. Walton, 535 U.S. 212, 214–15, 222 (2002) (holding that the Social Security Administration's interpretation of disability was subject to the *Chevron* framework because of "the interstitial nature of the legal question, the related expertise of the Agency, the importance of the question to administration of the statute, the complexity of that administration, and the careful consideration the Agency has given the question over a long period of time"); United States v. Mead Corp., 533 U.S. 218, 221, 226–27 (2001) (holding that tariff-classification rulings by the United States Customs Service are due no deference); Christensen v. Harris Cnty., 529 U.S. 576, 586–87 (2000) (holding that an opinion letter issued by the Department of Labor did "not warrant *Chevron*-style deference" because it "lack[ed] the force of law").

69. *Mead*, 533 U.S. at 226–27.

70. Gersen & O'Connell, *supra* note 33, at 960; *see also Barnhart*, 535 U.S. at 222 ("[W]hether a court should give such deference depends in significant part upon the interpretive method used and the nature of the question at issue.").

71. *See Barnhart*, 535 U.S. at 221 ("[T]he fact that the Agency previously reached its interpretation through means less formal than 'notice and comment' rulemaking does not automatically deprive that interpretation of the judicial deference otherwise its due." (citation omitted)); *Mead*, 533 U.S. at 230–31 (noting that "as significant as notice-and-comment is in pointing to *Chevron* authority, the want of that procedure . . . does not decide the case").

72. *See supra* text accompanying notes 26–30.

73. Gersen & O'Connell, *supra* note 33, at 961.

74. *Id.* at 961 & nn.135–36.

75. *Id.* at 959.

76. *Id.* at 962–64.

and if there was a "clear error in judgment."[77] Deadlines have the capacity to increase the likelihood that agency actions will be found arbitrary and capricious because deadlines may force the agency to promulgate rules of lower quality.[78] Some scholars argue that courts should "reduc[e] the intensity of arbitrary and capricious review because of statutory deadlines."[79] This view does not appear to have been widely adopted by courts, but at least one court has expressed a similar view of deadlines.[80]

4. Congress's Response to Missed Deadlines.—Congress may also take action if an agency fails to meet its deadlines. Congress may respond to a failure to meet a deadline by imposing additional deadlines[81] or may require that the agency explain the reasons it was unable to meet the deadline.[82] In addition, Congress may decide to punish the agency for failing to meet the deadlines by cutting the agency's budget.[83]

Thus, deadlines can affect the ability of third parties to successfully bring suit to compel agency action and can affect a court's procedural and substantive review of agency rulemaking. Moreover, Congress may respond to an agency's failure to meet its deadlines by reducing the agency's budget or by attempting to impose additional deadlines.

---

77. *Supra* note 36 and accompanying text.

78. *See* Gersen & O'Connell, *supra* note 33, at 963 (arguing that "[w]hen agencies sacrifice deliberative process to meet deadlines, decisions seem more likely to fail the arbitrary and capricious inquiry"); Thomas O. McGarity, *Some Thoughts on "Deossifying" the Rulemaking Process*, 41 DUKE L.J. 1385, 1456 (1992) (describing how deadlines "may force agencies to give short shrift to public comments" and may lead to "[h]asty agency action" that "may result in flawed analyses and poor explanations, the hallmarks of arbitrary and capricious action"); *see also* Ethyl Corp. v. EPA, 541 F.2d 1, 69–70 (D.C. Cir. 1976) (MacKinnon, J., dissenting) (arguing that a thirty-day deadline imposed by another panel resulted in "flawed regulations" because it forced the agency to make "a hasty decision . . . on a highly complicated matter that was still under study").

79. Gersen & O'Connell, *supra* note 33, at 962–63.

80. *See* Cal. Human Dev. Corp. v. Brock, 762 F.2d 1044, 1051 (D.C. Cir. 1985) (holding that a Department of Labor action was not arbitrary and capricious when the Department had to make "[c]omplex decisions" in a "short time span").

81. Sidney A. Shapiro, *Political Oversight and the Deterioration of Regulatory Policy*, 46 ADMIN. L. REV. 1, 25–26 (1994); *see also* Alden F. Abbott, *The Case Against Federal Statutory and Judicial Deadlines: A Cost–Benefit Appraisal*, 39 ADMIN. L. REV. 171, 181 (1987) [hereinafter Abbott, *Case Against Deadlines*] (describing how Congress imposed more deadlines on the agency actions of the EPA after it failed to meet statutory deadlines).

82. Sidney A. Shapiro & Thomas O. McGarity, *Reorienting OSHA: Regulatory Alternatives and Legislative Reform*, 6 YALE J. ON REG. 1, 54 (1989).

83. *See* Gregory L. Ogden, *Reducing Administrative Delay: Timeliness Standards, Judicial Review of Agency Procedures, Procedural Reform, and Legislative Oversight*, 4 U. DAYTON L. REV. 71, 85, 88 (1979) ("Legislative imposition of standards for timeliness . . . provides a clear, articulable substantive standard easily usable by oversight committees at agency and budget review time.").

III. The Use of Statutory Deadlines

Statutory deadlines are a mechanism used by Congress to "control . . . the timing of administrative action."[84] The use of statutory deadlines has the potential to reduce administrative delay, align agency action with legislative intent, and provide Congress with an easy mechanism for limiting agency discretion. However, the use of such deadlines can also be problematic because deadlines can reduce an agency's flexibility, decrease the quality of rulemaking, and be used by Congress as a political tool. Moreover, deadlines do not adequately address resource constraints. This Part describes the purposes of statutory deadlines and the problems associated with their use.

A. The Purposes of Statutory Deadlines

1. Reducing Administrative Delay.—One of the main purposes of statutory deadlines is to prevent or reduce administrative delay.[85] Although rulemaking by agencies was once considered a mechanism to increase the speed of government decision making "in areas where Congress was thought to be inexpert or inefficient,"[86] administrative delay is now a well-recognized problem in administrative law.[87] The logic of deadlines as a mechanism to reduce administrative delay is fairly straightforward; deadlines are believed to spur agency action by setting a definite date by which the rules must be promulgated.[88] In addition, deadlines can decrease administrative delay by ensuring that a required action is made a priority or at least can bring the action to the agency's attention.[89] In the same vein, deadlines may also require the agency to rethink its allocation of personnel and monetary resources and justify shifting those resources to that particular policy area in order to comply with the deadlines.[90] Administrative delay may also be curbed because deadlines make it more

84. Gersen & O'Connell, *supra* note 33, at 925 (emphasis omitted).

85. Abbott, *Case Against Deadlines*, *supra* note 81, at 179; *see also* George A. Bermann, *Administrative Delay and Its Control*, 30 AM. J. COMP. L. (SUPPLEMENT) 473, 474 (1982) (defining administrative delay as "an agency's failure to initiate action on a timely basis or to bring such action, once initiated, to a timely conclusion").

86. Bermann, *supra* note 85, at 473.

87. *See* 4 S. COMM. ON GOVERNMENTAL AFFAIRS, 95TH CONG., STUDY ON FEDERAL REGULATION: DELAY IN THE REGULATORY PROCESS 1 (Comm. Print 1977) ("There is no question that undue delay is pervasive in the Federal regulatory process."); Gersen & O'Connell, *supra* note 33, at 927 (describing administrative delay as "an increasingly prominent fixture in administrative law").

88. Abbott, *Case Against Deadlines*, *supra* note 81, at 179.

89. *Id.* at 179–80; *see also* Ogden, *supra* note 83, at 85 (arguing that statutory deadlines demonstrate "the legislature's commitment to timely agency decisionmaking").

90. Abbott, *Case Against Deadlines*, *supra* note 81, at 179–80.

likely that intended beneficiaries of the statute can compel agency action in court.[91]

Although deadlines are widely used to reduce administrative delay in agency rulemaking, there is limited empirical evidence of their effectiveness in doing so.[92] One study, using data from the *Unified Agenda of Federal Regulatory and Deregulatory Actions*, evaluated the use and effectiveness of deadlines on rulemaking across thirty agencies.[93] This study indicates that, while the effects in some cases were "relatively modest," deadlines "do quicken the pace of agency decisions."[94]

2. Aligning Agency Decision Making with Legislative Intent.— Another purpose of statutory deadlines is to align agency decision making with Congress's intent. First, as discussed briefly above, statutory deadlines can be used to communicate that Congress views rulemaking as a priority.[95] Congress can also use deadlines to indicate the relative importance of different rulemaking requirements in the same legislation. For instance, if Congress passes a lengthy piece of legislation that requires a substantial amount of rulemaking by the agency on a variety of topics, it can use deadlines to indicate what the agency should tackle first.[96] However, the ability of a deadline to convey priority can be reduced in two ways. First, if Congress imposes too many deadlines on an agency, it "dilutes the import of any single deadline."[97] Second, if Congress does not clearly address the priority of newly issued deadlines as compared to deadlines imposed by previous legislation, it can create confusion as to their relative importance.[98]

A second way in which deadlines can align agency decision making with legislative intent is by reducing bureaucratic drift. Because agency rulemaking involves delegation of Congress's ability to make laws, it necessarily involves the "risk that the bureaucracy will alter policy."[99] This risk, referred to as bureaucratic drift, reflects that "changes in administrative agency policies" can "lead to outcomes inconsistent with the original

---

91. *See supra* section II(B)(1). *But see* R. Shep Melnick, *The Political Roots of the Judicial Dilemma*, 49 ADMIN. L. REV. 585, 596 (1997) (arguing that this might not be a desirable way to set public policy because "[u]nlike public administrators, private attorneys general have no responsibility for taking into account opportunity costs").

92. Gersen & O'Connell, *supra* note 33, at 937.

93. *Id.* at 938 n.57, 981 tbl.2.

94. *Id.* at 945, 948.

95. *See supra* note 89 and accompanying text.

96. *See infra* text accompanying notes 177–80.

97. Abbott, *Case Against Deadlines*, *supra* note 81 (internal quotation marks omitted).

98. *See id.* at 182 (explaining that the legislative process can often be "insensitive . . . to difficult trade-offs and priority conflicts among programs").

99. Jacob E. Gersen, *Temporary Legislation*, 74 U. CHI. L. REV. 247, 282 (2007).

expectations of the legislation's intended beneficiaries."[100] Although deadlines cannot require agencies to act in the way that Congress intended,[101] the imposition of deadlines that require rules to be promulgated during the current legislative session can reduce bureaucratic drift by "ensuring that the enacting Congress gets to see (and possibly object to) the final regulation."[102] Thus, Congress is given an opportunity to voice its objections to rules promulgated by the agency that are not in line with Congress's intent.

Similarly, deadlines can help combat legislative drift, which is the concern that a future Congress will "have different policy preferences than those of the current Congress, and therefore attempt to undo previous legislative outcomes."[103] Deadlines increase the likelihood that the regulations will be promulgated while the current Congress is in session.[104] Although a future Congress could still repeal the statute, it will have a harder time doing so because once regulations have been promulgated there is "some form of status quo bias."[105]

Lastly, deadlines align agency decision making with Congress's legislative intent by preventing or reducing agency capture. In severe cases, agency capture "denotes an agency that is effectively controlled by an industry that it regulates."[106] Agency capture can move agency decision making away from congressional intent by allowing third parties to manipulate policies "to serve the interests of the regulated industry rather than the intended beneficiaries of congressional action."[107] By limiting the time that agencies have to promulgate regulations, a deadline can also limit the time in which industry lobbyists and other outsiders can attempt to influence agency decision makers or delay action.[108] Thus, deadlines can "mitigat[e] outside pressures to avoid reaching a decision" and can give

100. Jonathan R. Macey, *Separated Powers and Positive Political Theory: The Tug of War over Administrative Agencies*, 80 GEO. L.J. 671, 672 (1992).

101. Sidney A. Shapiro & Robert L. Glicksman, *Congress, the Supreme Court, and the Quiet Revolution in Administrative Law*, 1988 DUKE L.J. 819, 836 (noting that a deadline "requires only that the agency act, not that it act in a particular way").

102. Gersen & O'Connell, *supra* note 33, at 936.

103. Gersen, *supra* note 99.

104. Gersen & O'Connell, *supra* note 33, at 936.

105. *Id.* at 936–37 (noting that status quo bias "may make it marginally harder to eliminate" a program).

106. Michael D. Sant'Ambrogio, *Agency Delays: How a Principal-Agent Approach Can Inform Judicial and Executive Branch Review of Agency Foot-Dragging*, 79 GEO. WASH. L. REV. 1381, 1393 (2011).

107. *Id.* at 1393–94.

108. *See* Abbott, *Case Against Deadlines*, *supra* note 81, at 176 (noting that "outsiders" may delay agency action when it serves their interests); S. COMM. ON GOVERNMENTAL AFFAIRS, *supra* note 87, at 12 (reporting that "interests or institutions outside the agency" also cause delay when it is in their best interest, thereby contributing to "regulatory lag").

"the agency a reason to end its analysis and make a difficult, but necessary decision."[109]

3. Ease of Application.—A third reason that Congress uses deadlines extensively is that they are easy to implement and monitor. Deadlines provide "clear, articulable substantive standard[s] easily usable by oversight committees at agency and budget review time."[110] Additionally, deadlines may allow Congress to "narrow an agency's discretion" in areas where Congress may not have the expertise to prescribe substantive standards to rein in the agency or to evaluate the product of agency decision making.[111]

Thus, Congress's use of deadlines can help reduce many of the unintentional effects of congressional delegation of rulemaking authority to agencies. By setting a definite end date, deadlines may encourage quicker agency rulemaking. In addition, deadlines can help ensure priority rulemaking is higher up on the agency's agenda and can cause shifting of resources and personnel to the priority rulemaking area. Deadlines can also align agency decisions with legislative purpose by clarifying regulatory priority and by reducing the effects of bureaucratic drift, legislative drift, and agency capture. Moreover, deadlines achieve this purpose in an easy-to-implement fashion.

B. Problems Created by Statutory Deadlines

1. Reduction in Agency Flexibility.—As discussed above, one way that statutory deadlines increase the speed of agency decision making is by justifying the reallocation of resources and personnel from a nondeadline action to an action with a deadline.[112] While this may help reduce administrative delay, it also reduces agency flexibility.[113] The reduction of agency flexibility caused by deadlines can impair one of the main purposes of delegation of rulemaking to agencies, which is that agencies should have wide discretion in their policy making decisions because they have "better information and greater expertise than Congress."[114] By imposing a deadline on one agency action or policy area, Congress reduces an agency's

---

109. Shapiro & Glicksman, *supra* note 101, at 830.

110. Ogden, *supra* note 83; *see also* Shapiro & Glicksman, *supra* note 101, at 831 ("Deadlines improve legislative oversight because Congress can easily determine whether a statutory deadline has been met.").

111. Shapiro & Glicksman, *supra* note 101, at 831.

112. *See supra* note 90 and accompanying text.

113. *See* Ogden, *supra* note 89, at 79 (arguing that imposing specific time limits on decision making "can be rigid and too inflexible").

114. Gersen & O'Connell, *supra* note 33, at 925–26. Courts have acknowledged that agencies are better equipped than judges to make decisions about how their resources should be used. *See* Heckler v. Chaney, 470 U.S. 821, 831–32 (1985) ("The agency is far better equipped than the courts to deal with the many variables involved in the proper ordering of its priorities.").

ability to "allocate resources according to need and importance across programs over time."[115] Furthermore, deadlines can increase the problem of new-risk bias.[116] New-risk bias occurs because of the "tendency to favor new, high-profile risks for regulation over older, more familiar risks that may be more serious."[117] This problem is made worse when deadlines are imposed on agency action to address newer risks.[118]

In addition, deadlines can reduce agency flexibility in deciding what procedure to use in promulgating rules.[119] Because deadlines may reduce the likelihood of a notice-and-comment period, deadlines may reduce public participation.[120] The inability of the public to participate in the rulemaking process can have negative impacts on the quality of the rulemaking because the agency may not have all of the information it could use to make a decision.[121] Moreover, if a deadline does not allow for a notice-and-comment period, it can take away from the democratic legitimacy that public participation provides the rulemaking process.[122]

Thus, the benefits of using statutory deadlines to decrease administrative delay should be weighed against the reduction in an agency's flexibility in allocating its resources and its ability to utilize the notice-and-comment procedure to encourage public participation, ensure it has complete information, and instill democratic legitimacy into its rulemaking.

2. Resource Constraints.—An agency may also suffer from an inability to meet deadlines with its current resources.[123] This problem is based on the reality that the imposition of a deadline does not guarantee that the agency will receive the funds it needs to meet the deadline.[124] This is

---

115. Gersen & O'Connell, *supra* note 33, at 974.

116. *Id.*

117. *Id.*

118. *Id.*

119. Abbott, *Case Against Deadlines*, *supra* note 81, at 185 (explaining that deadlines may preclude certain procedural techniques that would improve the decision-making process).

120. *See* Gersen & O'Connell, *supra* note 33, at 956 ("Deadlines impose significant constraints on agency resources, and, therefore, agencies often forego notice and comment rulemaking . . . for deadline-driven actions.")

121. *See infra* notes 142–43 and accompanying text.

122. *See* Batterton v. Marshall, 648 F.2d 694, 703 (D.C. Cir. 1980) ("The essential purpose of according § 553 notice and comment opportunities is to reintroduce public participation and fairness to affected parties after governmental authority has been delegated to unrepresentative agencies."); Guardian Fed. Sav. & Loan Ass'n v. Fed. Sav. & Loan Ins. Corp., 589 F.2d 658, 662 (D.C. Cir. 1978) ("[T]he procedure for public participation tends to promote acquiescence in the result even when objections remain as to substance."); Gersen & O'Connell, *supra* note 33, at 972 (noting that deadlines can worsen an agency's rulemaking process because the promulgated rule will not have the "democratic legitimacy produced by public participation").

123. *See* McGarity, *supra* note 78, at 1437 ("A very important reason for the agencies' failure to meet statutory and administrative deadlines is Congress's failure to appropriate sufficient resources to the agencies to undertake the ambitious rulemaking tasks Congress assigns them.").

124. Abbott, *Case Against Deadlines*, *supra* note 81, at 182.

because the budget is determined through a separate process—appropriations.[125] In practice, the amount the agency needs to meet deadlines often differs from the amount actually given to the agency in appropriations.[126] This is because in creating statutory regimes "Congress tends to set absolute goals, and to tie these goals to absolute deadlines," while the appropriations process must consider all agencies' needs and make "difficult trade-offs and priority decisions."[127]

Constrained resources can also be counterproductive to achievement of the goals of reducing agency delay and reducing agency capture. An agency will be less able to act quickly if it does not have the people and monetary resources to do so. In addition, when an agency does not have the necessary resources, it may need to go to outside sources to gain the information it needs to promulgate rules.[128] This information comes primarily from regulated industries, thus "encouraging the agency to develop a productive working relationship with the industry."[129] This can cause the agency to develop "an industry bias" because individuals do not have the resources to generate the needed information and their interests are "normally too small to justify" representation before the agency.[130]

The problems created by constrained resources can be exacerbated by the legal effects of the agency's failure to meet deadlines. As discussed above, deadlines can increase the likelihood that a third party can successfully bring a claim to compel agency action.[131] And a party also may bring suit to set aside the agency's promulgated rules based on the procedure[132] and substance of the rules.[133] If an agency is unable to meet a deadline because it lacks the resources to do so, the additional costs imposed by the litigation can add to its resource constraint.[134] Thus,

---

125. *See id.* (arguing that one reason that agencies fail to meet deadlines is the "tension between an ambitious authorizing statute and a more realistic appropriation"); Richard J. Pierce, Jr., *Judicial Review of Agency Actions in a Period of Diminishing Agency Resources*, 49 ADMIN. L. REV. 61, 84 (1997) ("Congress regularly orders agencies to take specific actions by specific times and just as regularly refuses to provide agencies with the resources required to take the statutorily mandated actions.").

126. *See* Pierce, *supra* note 125, at 77 (arguing that "Congress rarely coordinates its decisions to assign tasks to agencies with its decisions to appropriate").

127. Abbott, *Case Against Deadlines*, *supra* note 81, at 182.

128. Richard B. Stewart, *The Reformation of American Administrative Law*, 88 HARV. L. REV. 1669, 1686 (1975).

129. Sant'Ambrogio, *supra* note 106, at 1394; *see also* Stewart, *supra* note 128 (explaining that the outside information usually comes from "organized interests, such as regulated firms, that have a substantial stake in the substance of agency policy").

130. Stewart, *supra* note 128.

131. *See supra* section II(B)(1).

132. *See supra* section II(B)(2).

133. *See supra* section II(B)(3).

134. *See* Abbott, *Case Against Deadlines*, *supra* note 81, at 187 (arguing that litigation resulting from a failure to meet deadlines can produce wasteful costs when the deadlines were

"limited agency resources may be expended in litigation over deadlines rather than in writing regulations."[135] In addition, while some courts have considered resource constraints, others refuse to do so.[136]

3. Decreased Quality of Agency Rulemaking.—Some scholars have argued that the use of deadlines may also decrease the quality of agency rulemaking.[137] One scholar argues that the use of deadlines "may signify a determination by Congress about how the agency should balance the timeliness and quality of its decisions."[138] This decrease in quality may be the result of the limited time frame. The limited time frame can have even more of an effect on the quality of the rulemaking when the deadline is unreasonable.[139] In addition, quality of decision making may be affected if the agency is unable to properly utilize the notice-and-comment procedure—either because the time frame does not allow or because the agency does not have the resources.[140] As the D.C. Circuit recognized in *Guardian Federal Savings & Loan Ass'n v. Federal Savings & Loan Insurance Corp.*,[141] one purpose of the notice-and-comment procedure required by APA § 553 is to "assure[] that the agency will have before it the facts and information relevant to a particular administrative problem, as well as suggestions for alternative solutions."[142] Thus, the quality of

---

"impossible to meet" and "when the litigation itself does not speed the pace of rulemaking"); Alden F. Abbott, *Case Studies on the Costs of Federal Statutory and Judicial Deadlines*, 39 ADMIN. L. REV. 467, 470–71 (1987) [hereinafter Abbott, *Costs of Deadlines*] (discussing the EPA's inability to meet "unrealistic" statutory deadlines under the Federal Water Pollution Control Act, which resulted in litigation that "consumed scarce resources that might otherwise have been directed to the promulgation of regulations").

135. McGarity, *supra* note 78.

136. Pierce, *supra* note 125, at 63–64; *see also* Envtl. Def. Ctr. v. Babbitt, 73 F.3d 867, 868–69 (9th Cir. 1995) (holding that the Secretary of the Interior's failure to meet a deadline imposed by the Endangered Species Act was caused by the unavailability of funds and that the judgment should provide that his compliance is excused "until a reasonable time after appropriated funds are made available").

137. *See* Gersen & O'Connell, *supra* note 33, at 933 ("A straightforward potential result [of deadlines] is to decrease the quality of agency deliberations and decisions.").

138. Sant'Ambrogio, *supra* note 106, at 1417.

139. *See* Shapiro & Glicksman, *supra* note 101, at 835 (explaining that because deadlines do not actually "simplify the agency's substantive task," they can lead to "unrealistic time pressures").

140. *See* Gersen & O'Connell, *supra* note 33, at 944–45 (observing that "deadlines produce fewer chances for public input and less agency process, two variables typically associated with higher-quality and more legitimate decisions"); Richard J. Pierce, Jr., *Seven Ways to Deossify Agency Rulemaking*, 47 ADMIN. L. REV. 59, 85–86 (1995) (noting that the social benefits of notice-and-comment procedures in rulemaking include "enhanced quality of agency rules attributable to broad participation of all potentially affected groups in the policymaking process").

141. 589 F.2d 658 (D.C. Cir. 1978).

142. *Id.* at 662; *see also* Stephen M. Johnson, *Good Guidance, Good Grief!*, 72 MO. L. REV. 695, 702–03 (2007) (explaining that public participation "provides the agency with important information about the impacts of proposed decisions that enable the agency to administer the law in a rational, defensible manner").

rulemaking may be lowered if the agency does not utilize the notice-and-comment procedure because the rule might not reflect all of the relevant information. Deadlines make it less likely that the agency will be able or willing to "take advantage of the information and expertise produced by notice and comment."[143]

Moreover, imposition of a deadline in one policy area can have negative effects on the quality of an agency's actions in other policy areas.[144] A deadline causes an agency to divert resources and personnel from non-deadline activities to those with deadlines imposed.[145] The lack of resources and time spent on a particular policy area can reduce the quality of the rulemaking produced for that policy area.[146]

4. Deadlines as a Political Tool.—One aspect of deadlines that is particularly problematic is that Congress can use them as political tools. First, deadlines may allow Congress to bypass the procedural requirements of the APA without explicitly exempting the agency action from those requirements.[147] As previously discussed, courts may excuse an agency from the notice-and-comment procedure required by § 553.[148] By setting short deadlines, Congress can bypass the notice-and-comment requirement in order "to avoid the lengthy process of informal rulemaking."[149] This can be problematic for an agency, because whether or not a court will find that a deadline is enough to establish good cause for avoiding the procedure is fact-specific.[150] In addition, an agency may wish to take advantage of the information that can be made available through public comment and the democratic legitimacy that it can provide to the agency's rulemaking.[151]

Moreover, Congress may utilize deadlines, not to increase the speed of regulation, but as a symbolic "statement that a problem will be solved."[152] To indicate that it is taking a strong stance, Congress may include a strict deadline.[153] However, at the same time, Congress can strategically refuse to provide an agency the funds it requires to meet this deadline because it "may not actually want the agency to engage in much regulatory or

143. Gersen & O'Connell, *supra* note 33, at 972.

144. *Id.* at 933–34.

145. *See supra* note 90 and accompanying text.

146. Gersen & O'Connell, *supra* note 33, at 933–34.

147. *Id.* at 934.

148. *See supra* note 27 and accompanying text.

149. Gersen & O'Connell, *supra* note 33, at 934.

150. *See supra* section II(B)(2).

151. *See supra* text accompanying notes 121–122.

152. Abbott, *Case Against Deadlines*, *supra* note 81, at 180; *see also* Abbott, *Costs of Deadlines*, *supra* note 134, at 487 ("The process for establishing statutory time limits understandably is driven by political pressures to 'do something' in order to alleviate highly publicized problems.").

153. *See supra* section II(B)(2).

administrative activity, or it may prefer that the agency move slowly."[154] Similarly, if a previous Congress passed legislation that the current Congress does not support, it may choose to cut funding for the agency rather than repeal the statute.[155] Thus, Congress can please the public by showing that it is acting to correct a problem by passing legislation while preventing an agency from doing what it is required to do under that legislation.[156] In addition, individual legislators can "exert pressure on agencies to defer or postpone regulatory actions" by utilizing their position on a committee with oversight responsibility or power over appropriations for the agency.[157]

In the same vein, Congress can use delegation to push responsibility for acting and making difficult decisions onto an agency. Congress may use broad delegations so that it can "avoid taking responsibility for the consequences of legislation."[158] Instead of making the choices itself—"for fear of incurring the ire of a regulated constituency"—Congress may use broad statutory language and leave the tough decisions to the agency.[159] Delegation may also be used when Congress is unable to agree on specific agency choices.[160] When deadlines are used in conjunction with vague statutory mandates, deadlines can be difficult or impossible for agencies to meet.[161] This can not only result in the additional costs of litigation but can also have a negative effect on the reputation of the agency.[162] In addition, if Congress responds by imposing additional deadlines that also cannot be met, it can create "public cynicism about the ability of government to regulate."[163]

---

154. Sant'Ambrogio, *supra* note 106, at 1398 (describing how Congress may pass legislation for "purely symbolic reasons" and then use appropriations to delay agency action).

155. *Id.*

156. *See* DAVID SCHOENBROD, POWER WITHOUT RESPONSIBILITY: HOW CONGRESS ABUSES THE PEOPLE THROUGH DELEGATION 11–12 (1993) (using statutory deadlines imposed on the EPA to illustrate how legislatures pass legislation "although they [know] that the agency could never come close to discharging these duties with the time, resources, and political power given to it").

157. Sant'Ambrogio, *supra* note 106, at 1397.

158. DANIEL A. FARBER & PHILIP P. FRICKEY, LAW AND PUBLIC CHOICE: A CRITICAL INTRODUCTION 81 (1991).

159. Sant'Ambrogio, *supra* note 106, at 1391–92.

160. *Id.* at 1392.

161. *See* Shapiro & Glicksman, *supra* note 101, at 844 (arguing that the coercive model of delegation, which utilizes deadlines, is open to abuse because the imposition of deadlines "may enable legislators to strike a posture in favor of regulation while avoiding the difficult policy choices that regulation entails").

162. *See id.* at 836 n.81 (noting that a failure to meet deadlines can result in "congressional and public disrespect and distrust of the agency").

163. Shapiro, *supra* note 81.

In sum, while there are various policy objectives that may be advanced by the use of deadlines, they must be weighed against the various problems implicit in the use of deadlines.

IV. Assessing the SEC's Missed Deadlines Under the JOBS Act

The SEC has received critical attention due to missed statutory deadlines under the JOBS Act.[164] In order to assess the validity of these criticisms, this Part analyzes the purposes and problems of statutory deadlines as used to regulate the actions of the SEC. While deadlines can accelerate regulatory action by the SEC and align the SEC's actions with legislative intent, the SEC's failure to meet its deadlines under the JOBS Act is due, at least in part, to the problematic use of deadlines by Congress.

A. Purposes of Deadlines in the Context of the SEC

First, the SEC has at least some symptoms of administrative delay. This is illustrated, at least anecdotally, by the controversy surrounding Mary Schapiro, former chairman of the SEC, and her actions in implementing (or failing to implement) rules under the JOBS Act. Before the JOBS Act was passed by the Senate, Schapiro wrote a letter to the Senate Committee on Banking, Housing, and Urban Affairs explaining her concerns that the Act lacked necessary investor protections and that some of the changes could result in "real and significant damage to investors."[165] After the JOBS Act was passed, the SEC failed to meet its first rulemaking deadline, which required the agency to promulgate rules to eliminate the prohibition against general solicitation in certain securities.[166] Testifying before the House Subcommittee on TARP, Financial Services and Bailouts of Public and Private Programs, Schapiro indicated that she did not believe that the deadline would be met, but that she expected that it would "be done this summer."[167] However, in August 2012, the SEC issued a proposed rule

164. *See supra* notes 8–9 and accompanying text.

165. Letter from Mary L. Schapiro, Chairman, Sec. & Exch. Comm'n, to Tim Johnson, Chairman, Comm. on Banking, Hous. & Urban Affairs, U.S. Senate and Richard C. Shelby, Ranking Member, Comm. on Banking, Hous. & Urban Affairs, U.S. Senate 2 (Mar. 13, 2012), *available at* http://www.thevaluealliance.com/Schapiro_letter_Jobs_Act_031312.pdf.

166. *See* Jumpstart Our Business Startups Act, Pub. L. No. 112-106, § 201(a), 126 Stat. 306, 313–14 (2012) (codified in scattered sections of 15 U.S.C.) (requiring the rules to be issued within 90 days of the statute's enactment).

167. *The JOBS Act in Action Part II: Overseeing Effective Implementation That Can Grow American Jobs: Hearing Before the Subcomm. on TARP, Fin. Servs. & Bailouts of Pub. & Private Programs of the Comm. on Oversight & Gov't Reform*, 112th Cong. 23 (2012) [hereinafter *The JOBS Act in Action Part II*] (statement of Mary Schapiro, Chairman, United States Securities and Exchange Commission).

for eliminating the ban on general solicitation and general advertising.[168] By choosing to issue a proposed rule, which involved a comment period,[169] the SEC was not able to finish promulgating in the timeline that Schapiro had suggested.[170] Moreover, it does not appear that this was the intent of the staff when drafting the proposed rule. Originally, the SEC staff had planned to issue an interim-final rule, which would have allowed it to "bypass the lengthier public comment process" in order to lift the ban immediately and "potentially tweak the rule down the road."[171]

It is clear that some individuals at the SEC did not agree with Schapiro's decision. When the Commissioners voted on the proposal, Commissioner Daniel M. Gallagher stated, "I am not happy to be sitting here today, almost two months after the JOBS Act deadline for a final rule, voting on a proposal," especially when, "[f]or months, the Commission had been told that the Staff was recommending that we vote on an interim final rule."[172] Furthermore, there is evidence that Schapiro might have intentionally delayed rulemaking in order to protect her legacy.[173] Although this is anecdotal and not conclusive evidence that the SEC purposefully delayed promulgation of rules under the JOBS Act,[174] it does provide some indication that deadlines can serve a useful purpose by helping speed up the rulemaking process in the SEC.

In addition, deadlines under the JOBS Act also appear to serve the purpose of aligning the SEC's actions with Congress's intent. In the past, Congress has not imposed an overwhelming number of deadlines on the

---

168. Press Release, U.S. Sec. & Exch. Comm'n, SEC Proposes Rules to Implement JOBS Act Provision About General Solicitation and Advertising in Securities Offerings (Aug. 29, 2012), *available at* http://www.sec.gov/news/press/2012/2012-170.htm.

169. *See* Eliminating the Prohibition Against General Solicitation, 77 Fed. Reg. 54,464, 54,473–74 (proposed Sept. 5, 2012) (to be codified at 17 C.F.R. pts. 230, 239) (requesting comments on the proposed rule).

170. *See Eliminating the Prohibition Against General Solicitation and General Advertising in Rule 506 and Rule 144A Offerings*, U.S. SEC. & EXCHANGE COMMISSION, http://www.sec.gov/info/smallbus/secg/general-solicitation-small-entity-compliance-guide.htm (last updated Sept. 20, 2013) (noting that the regulation was not adopted until July 10, 2013, long after Schapiro's timeline).

171. Sarah N. Lynch, *Emails Suggest SEC's Schapiro Delayed JOBS Act Rule amid Concerns About Legacy*, REUTERS (Dec. 1, 2012, 9:59 PM), http://www.reuters.com/article/2012/12/02/sec-schapiro-idUSL1E8N201X20121202.

172. Daniel M. Gallagher, Comm'r, U.S. Sec. & Exch. Comm'n, Statement at SEC Open Meeting: Proposed Rules to Eliminate the Prohibition Against General Solicitation and General Advertising in Rule 506 and Rule 144A Offerings (Aug. 29, 2012) (emphasis omitted), *available at* http://www.sec.gov/News/Speech/Detail/Speech/1365171491084#.UmLrJ2Q6Ud5.

173. Lynch, *supra* note 171 (quoting an e-mail sent from Schapiro to Corporation Finance Director Meredith Cross, in which Schapiro stated, "I don't want to be tagged with an anti-investor legacy").

174. There are other potential reasons why the SEC would propose a proposed rule. For instance, the SEC has said that it requested public comments first because of the "'very real threat of a legal challenge' if the agency did not go through a more rigorous rule-making process." Lynch, *supra* note 171.

SEC as compared to other agencies.[175] This could indicate that Congress is able to use deadlines to set priorities more effectively with the SEC than other agencies because the importance of each individual deadline is not as diluted.[176] This, however, has changed, as Dodd-Frank and the JOBS Act have imposed many more deadlines on the SEC.[177] The increased number of deadlines could make it more difficult for the SEC to prioritize amongst the activities with deadlines.

Standing alone, the JOBS Act does a fairly good job of prioritizing the individual rulemaking requirements imposed on the SEC by staggering the deadlines. The statute required the SEC to promulgate rules regarding the general ban on solicitation within 90 days of enactment,[178] issue a report on how to streamline the registration process within 180 days of enactment,[179] and promulgate rules governing crowdfunding within 270 days of enactment.[180] However, Congress did not establish priority between the deadlines in the JOBS Act and other statutes, such as the Dodd-Frank Act.[181] When the JOBS Act was passed, the SEC had still not promulgated many of the key rules required by the Dodd-Frank Act, which were already a year overdue.[182] Without knowing which deadlines should be given priority, the SEC will be unable to shift the necessary resources to that area to make sure that those rulemaking requirements will be given top priority. This may have the effect of reducing the effectiveness of deadlines as a way of expressing congressional priorities.

It also appears that Congress itself is divided on which legislation should be given priority. At a hearing before the House Subcommittee on TARP, Financial Services and Bailouts of Public and Private Programs,

---

175. For an empirical study of the use of deadlines between 1988 and 2003, see Gersen & O'Connell, *supra* note 33, at 981 tbl.2. They found that the SEC had 25 deadlines imposed on the agency's actions during that period, while the EPA had 611 and the Department of Commerce had 940. *Id.*

176. *See supra* notes 97–98 and accompanying text.

177. *See Implementing the Dodd-Frank Wall Street Reform and Consumer Protection Act*, U.S. SEC. & EXCHANGE COMMISSION (last updated Jan. 16, 2014), http://www.sec.gov/spotlight/dodd-frank.shtml (indicating that the Dodd-Frank Act imposed more than ninety provisions mandating SEC rulemaking and many other deadlines that gave the SEC the discretion to regulate); *infra* notes 178–80 and accompanying text.

178. Jumpstart Our Business Startups Act, Pub. L. No. 112-106, § 201(a), 126 Stat. 306, 313–14 (2012) (codified in scattered sections of 15 U.S.C.).

179. § 108(b), 126 Stat. at 313.

180. §§ 302–304, 126 Stat. at 315–22.

181. *See* Daniel M. Gallagher, Comm'r, U.S. Sec. & Exch. Comm'n, Address at the IMFA Regional Conference: SEC Priorities in Perspective (Sept. 24, 2012), *available at* http://www.sec.gov/News/Speech/Detail/Speech/1365171491262#.UmP5h2Q6Ud4 (acknowledging that disparate statutory mandates leave the SEC on its own to prioritize rulemaking in accordance with different statutes).

182. Dan Froomkin, *SEC Stall Leaves Key Dodd-Frank Rules More than a Year Overdue*, HUFFINGTON POST (Apr. 17, 2012, 7:50 PM), http://www.huffingtonpost.com/2012/04/17/sec-dodd-frank-rules-year-overdue_n_1432839.html.

Representative Mike Quigley expressed that there was "no reason the JOBS Act should be prioritized in front of pending Dodd-Frank rulemakings."[183] In contrast, Representative Darrell Issa indicated that the JOBS Act should be given priority because "the creation of private sector jobs should come before the creation of new bureaucratic organizations, no matter how well meaning, in Washington."[184] Thus, because it is difficult for members of Congress to come to a conclusion as to which rulemaking responsibilities warrant immediate attention, it seems unlikely that the SEC will be able to determine which of the deadlines it should be focusing its resources and energy on.

Deadlines can also be helpful in the context of SEC rulemaking because there is evidence that the SEC is prone to agency capture. One scholar suggests that this is due to the revolving-door phenomenon, which may cause many in the agency to "identif[y] with the market participants they [are] ostensibly regulating" and may lead to the SEC becoming more lax in its regulatory and enforcement functions.[185] In addition, the SEC must obtain information from the organizations it regulates in order to do its job, which can lead to agency capture because it gives the SEC "a strong incentive to cooperate with entities directly subject to [its] regulatory decisions."[186] The likelihood of this occurrence increases when resources are scarce, as the SEC must rely on information from outside sources.[187] Deadlines can reduce or limit this agency capture by limiting the time during which the SEC is exposed to regulated firms and interest groups.[188]

Although a deadline may reduce some of the effect of agency cooperation with regulated firms by shortening the amount of time that the SEC is exposed to these firms, it does not appear to be a solution for the revolving-door phenomenon. However, as discussed above, deadlines can also have the positive effect of reducing bureaucratic drift because they give the members of the enacting Congress the ability to voice their

---

183. *The JOBS Act in Action: Overseeing Effective Implementation That Can Grow American Jobs: Hearing Before the Subcomm. on TARP, Fin. Servs. & Bailouts of Pub. & Private Programs of the Comm. on Oversight & Gov't Reform*, 112th Cong. 33 (2012) [hereinafter *The JOBS Act in Action*] (statement of Rep. Mike Quigley, Member, Comm. on Oversight & Gov't Reform).

184. *The JOBS Act: Importance of Prompt Implementation for Entrepreneurs, Capital Formation, and Job Creation: Joint Hearing Before the Subcomm. on TARP, Fin. Servs. & Bailouts of Pub. & Private Programs of the Comm. on Oversight & Gov't Reform and the Subcomm. on Capital Mkts. & Gov't Sponsored Enters. of the Comm. on Fin. Servs.*, 112th Cong. 33 (2012) (statement of Rep. Darrell E. Issa, Chairman, Comm. on Oversight & Gov't Reform).

185. Rachel E. Barkow, *Insulating Agencies: Avoiding Capture Through Institutional Design*, 89 TEXAS L. REV. 15, 47 (2010) (quoting Jonathan R. Macey, *Wall Street in Turmoil: State-Federal Relations Post-Eliot Spitzer*, 70 BROOK. L. REV. 117, 128 (2004)).

186. Mark Seidenfeld, *Bending the Rules: Flexible Regulation and Constraints on Agency Discretion*, 51 ADMIN. L. REV. 429, 464 (1999).

187. *See supra* notes 128–30 and accompanying text.

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

objections.[189] This also has the potential to reduce bureaucratic drift caused by industry bias. This would have been the case with the deadlines imposed by the JOBS Act, had they been met, because the Act required all rules to be promulgated within 270 days of the statute's enactment,[190] leaving a few days after the deadline for members of the enacting Congress to express their opinions on the rules as promulgated before the legislative session finished. And, had the deadlines been met, the effects of legislative drift could have been minimized by status quo biases.[191]

B. Problems with the Use of Deadlines in the Context of the SEC

Although there are several benefits that can accrue as a result of the use of deadlines by Congress, there are also problems that have arisen from their use under the JOBS Act. These problems likely contributed to the SEC's failure to meet its deadlines under the Act.

First, the deadlines reduced the SEC's flexibility because they required the agency to choose between using a full notice-and-comment procedure and meeting its deadlines under the JOBS Act. For instance, although the SEC staff originally recommended forgoing the traditional notice-and-comment period in order to meet the ninety-day deadline,[192] Schapiro changed her mind and decided that the quality of the rules would benefit from the notice-and-comment procedure because of the "high level of investor interest."[193] The SEC was also concerned about the risk of litigation if it did forego the notice-and-comment procedure.[194]

On the other hand, it could be argued that Congress made the choice of speed over quality when it set such a tight deadline.[195] However, if that were the case, Congress would have better achieved this purpose by also waiving the APA requirements as part of the JOBS Act.[196] Choosing not to do so put the burden on the SEC to decide whether to forego what it felt were necessary notice-and-comment procedures—exposing itself to

---

189. *See supra* text accompanying note 102.

190. Jumpstart Our Business Startups Act, Pub. L. No. 112-106, § 302(c), 126 Stat. 306, 320 (2012) (codified in scattered sections of 15 U.S.C.) (requiring the rules to be issued within 270 days of the statute's enactment).

191. *See supra* text accompanying notes 104–105.

192. Lynch, *supra* note 171.

193. *Id.*

194. *Id.*; *see also The JOBS Act in Action*, *supra* note 183, at 44–45 (statement of John C. Coffee, Jr., Professor of Law, Columbia University Law School) (asserting that "the SEC is today caught between the rock and the hard place" because it was "asked to expedite rules and it is trying to do so," but it faces the "risk that its rules could be found to be arbitrary [and] capricious").

195. Sant'Ambrogio, *supra* note 106, at 1417 (characterizing deadlines as a "determination by Congress about how the agency should balance the timeliness and quality of its decisions").

196. *See* Gersen & O'Connell, *supra* note 33, at 959 n.128 (noting that Congress can waive and has waived APA requirements as part of a statutory scheme).

litigation—or to commit to a public-comment period and miss the statutory deadline.

Secondly—and probably the greatest problem for the SEC in its implementation of the JOBS Act—the deadlines did not take into account the resource constraints imposed on the SEC. As discussed above, there is often a discrepancy between the funds needed to implement a statutory scheme and the funds actually appropriated to the agency.[197] Even before it was faced with the heavy rulemaking load required by the Dodd-Frank Act and the JOBS Act, the SEC had faced resource constraints—making it difficult to maintain sufficient staff to carry out its duties.[198] The passage of Dodd-Frank exacerbated these problems. In March 2012, a month before the JOBS Act was passed, Mary Schapiro testified before a Subcommittee of the House Committee on Appropriations that the SEC needed more funds in order to promulgate rules as required by the Dodd-Frank Act, requesting $1.566 billion.[199] The appropriations committee proposed increasing the budget, but to a lesser degree than what Schapiro and the SEC felt they needed.[200] Congress, however, was unable to agree on this proposal and instead signed a law continuing the appropriations from the previous year.[201]

At the signing of the JOBS Act, President Obama also expressed the importance of keeping the SEC well funded.[202] For 2014, President Obama proposed a 27% increase in the SEC's budget;[203] however, it is unclear if

197. *See supra* notes 125–27 and accompanying text.

198. *See Oversight of the U.S. Securities and Exchange Commission's Operations, Activities, Challenges, and FY 2012 Budget Request: Hearing Before the Subcomm. on Capital Mkts. & Gov't Sponsored Enters. of the Comm. on Fin. Servs.*, 112th Cong. 49 (2011) (prepared statement of Robert Khuzami, Director, Division of Enforcement, United States Securities and Exchange Commission, et al.) ("Over the past decade, the SEC has faced significant challenges in maintaining a staffing level and budget sufficient to carry out its core mission."); Joel Seligman, *Self-Funding for the Securities and Exchange Commission*, 28 NOVA L. REV. 233, 237–39 (2004) (describing the chairmanship of Arthur Levitt, during which "fees collected by the SEC far exceeded its annual appropriations" and it was not given "adequate budgetary or staff support").

199. Mary Schapiro, Chairman, U.S. Sec. & Exch. Comm'n, Testimony Before the Subcommittee on Financial Services and General Government Committee on Appropriations, United States House of Representatives (Mar. 6, 2012), *available at* http://www.sec.gov/news/testimony/2012/ts030612mls.htm.

200. The House Appropriations Committee proposed a bill setting the budget at $1.371 billion. H.R. 6020, 112th Cong. § 515 (2012).

201. Consolidated and Further Continuing Appropriations Act 2013, Pub. L. No. 113-6, §§ 1101–1113, 127 Stat. 198, 412–15 (2013).

202. *See* Obama, *supra* note 2 ("[I]t's going to be important that we continue to make sure that the SEC is properly funded, just like all our other regulatory agencies, so that they can do the job and make sure that our investors get adequate protections.").

203. OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, BUDGET OF THE U.S. GOVERNMENT, FISCAL YEAR 2014, at 142 (2013), *available at* http://www.whitehouse.gov/sites/default/files/omb/budget/fy2014/assets/budget.pdf.

Congress will agree to such an increase.[204] Thus, despite calls from the SEC itself, scholars, and the President, Congress has been unwilling to agree to increase the SEC's budget to what is required by the agency to fulfill its duties. What makes Congress's refusal to increase the budget more difficult to understand is that the SEC is not funded by tax revenue, but by the fees it gets from the companies that it regulates.[205] Thus, there is little justification for refusing to increase the agency's budget when necessary.

Moreover, the SEC may experience additional resource constraints if litigation is brought against it for failing to meet its deadlines.[206] While it does not appear that this type of litigation has been as problematic for the SEC as it has been for other agencies,[207] one case was brought against the agency in 2012 for failure to promulgate rules under the Dodd-Frank Act.[208] Although the case was dismissed on the stipulation of the parties,[209] it could indicate that similar suits may be forthcoming.

As a result of the very tight deadlines, the SEC may also experience a decrease in the quality of its rulemaking. The changes being implemented by the SEC under the JOBS Act involve "significant changes to the federal securities laws."[210] If the SEC were to issue rules on such novel and complex issues too quickly, the quality of the rules would suffer as a result because the agency would have less time to discuss and deliberate the issues. Moreover, it could be argued that these deadlines are unrealistic,[211] which exacerbates the effect of the short timeframe on the quality of the promulgated rules.[212] If the SEC begins promulgating rules of lower quality in order to satisfy short or even unreasonable deadlines, a court could find

---

204. The House Appropriations Committee maintained $1.371 billion as its proposal for 2014, while the Senate bill would increase the budget to $1.674 billion. H.R. 2786, 113th Cong. § 518 (2013); S. 1371, 113th Cong. § 526 (2013).

205. *See* James B. Stewart, *As a Watchdog Starves, Wall Street Is Tossed a Bone*, N.Y. TIMES, July 15, 2011, http://www.nytimes.com/2011/07/16/business/budget-cuts-to-sec-reduce-its-effectiveness.html?pagewanted=all&_r=0 (pointing out that reducing the SEC's budget does not save taxpayers' money and "could cost the Treasury millions in lost fees and penalties" because Dodd-Frank contains a mandate stating that fees collected by the SEC cannot exceed its budget).

206. *See supra* note 134 and accompanying text.

207. *See generally* Abbott, *Costs of Deadlines*, *supra* note 134 (explaining that litigation over the EPA's failure to meet statutory deadlines consumed a portion the agency's scarce resources).

208. Complaint at 1, Oxfam Am., Inc. v. U.S. Sec. & Exch. Comm'n, No. 12-CV-10878-DJC (D. Mass. May 16, 2012).

209. Stipulation of Dismissal, Oxfam Am., Inc. v. U.S. Sec. & Exch. Comm'n, No. 12-CV-10878-DJC (D. Mass. Dec. 3, 2012).

210. *The JOBS Act in Action Part II*, *supra* note 167, at 7.

211. *See id.* at 5 (arguing that the ninety-day deadline for removing the ban on general solicitation was unrealistic and "[did] not provide time for drafting a new rule with a rigorous economic analysis, considering public input, and reviewing a proposed final rule at the Commission level").

212. *See supra* note 140 and accompanying text.

that the agency's actions were arbitrary and capricious.[213] In addition, if the SEC makes the decision to forgo notice-and-comment procedure to meet the deadline, it may increase the chances of an arbitrary and capricious finding because, while the doctrine "does not demand procedural formality, some degree of formality is often required implicitly."[214]

Lastly, the use of deadlines under the JOBS Act is problematic because Congress, by requiring the SEC to implement rules but not giving the agency the resources or time that it needed to complete them, intentionally or unintentionally shifted the blame for not taking action onto the SEC.[215] In addition, because the JOBS Act was not well drafted, it left many of the harder questions to the SEC to figure out.[216] The poor drafting may have resulted from the speed at which it passed through Congress.[217] The lack of necessary resources and the ambiguous drafting of the provisions of the JOBS Act are likely contributing factors to the SEC's inability to promulgate rules before the deadline.

In sum, although the SEC has shouldered the majority of the blame for its inability to comply with the deadlines created by the JOBS Act, at least part of the blame falls on Congress for its use of deadlines in the Act.

V. Recommendations for Improvement in the Use of Deadlines

While there are certainly problems created by the use of deadlines, they still serve the valuable purpose of reducing administrative delay and aligning agency action with legislative intent.[218] In order to improve the use of deadlines, in the context of the SEC and more broadly, I recommend improvements that fall into three categories. First, Congress should improve its use of deadlines to express priority. Second, Congress must take more responsibility for the legislation it passes, even legislation that requires rules to be promulgated by the SEC or another agency. And third, resources appropriated must be more in line with the funds necessary to meet the required deadlines.

---

213. *See supra* note 78 and accompanying text.

214. Gersen & O'Connell, *supra* note 33, at 973.

215. *See supra* notes 154–56 and accompanying text.

216. *See* C. Steven Bradford, *The New Federal Crowdfunding Exemption: Promise Unfulfilled*, 40 SEC. REG. L.J. 195, 198 (2012) (criticizing the crowdfunding exemption as "poorly drafted" because it left "many ambiguities and inconsistencies for the SEC or the courts to resolve").

217. *See* Dina ElBoghdady, *JOBS Act Falls Short of Grand Promises*, WASH. POST, Mar. 28, 2013, http://articles.washingtonpost.com/2013-03-28/business/38084496_1_jobs-act-president-obama-measure (remarking that the JOBS Act was passed "perhaps too fast" and "[e]ven supporters say they expected more time to work out the kinks" before it was signed into law).

218. *See supra* sections III(A)(1)–(2).

A. Improvement in the Use of Deadlines to Convey Priority

First, as discussed above, deadlines are used to align agency action with legislative intent by indicating that Congress views the particular action as a priority.[219] However, the effect of a deadline is diluted if Congress imposes too many deadlines[220] or if Congress does not address the priority of newly implemented deadlines relative to deadlines already in place.[221] As demonstrated by the SEC's experience under the Dodd-Frank and JOBS Acts,[222] multiple statutes imposing various deadlines can make it difficult for any agency to prioritize where its limited resources should be directed. In order to better align agency action with legislative intent, Congress should expressly indicate the priority of deadlines, not only within a statutory scheme, but also between different statutory programs. Additionally, by imposing fewer deadlines, Congress can increase the likelihood that the deadlines it does set can be complied with.

B. Congress Should Hold Itself More Accountable for the Choices It Makes in Legislation

Second, Congress should hold itself more accountable for the laws that it makes, even when it delegates rulemaking responsibility to agencies like the SEC. To do this, I recommend that Congress take the time to set deadlines that are more realistic. However, it is difficult, or impossible, to predict in advance how much time it will take an agency to promulgate rules, especially on novel or complex legislation like the new crowdfunding exemption under the JOBS Act.[223]

To remedy this problem, one scholar has suggested that agencies should set their own deadlines because they "can establish more accurate and more realistic estimates of how long a particular type of decision will take than can legislative bodies."[224] While this solution would lead to more realistic deadlines, it seems unlikely that Congress will be willing to give up this control. A more realistic solution would be for Congress to take into consideration the opinions of the agency as to what would be a realistic deadline based on its current workload, other deadlines, and resources. Moreover, just as allowing public participation in the rulemaking process has the effect of promoting public acceptance of the resulting rule, even if those opposing the rule continue to disagree,[225] allowing an agency to

---

219. *See supra* section III(A)(2).

220. *See supra* note 97 and accompanying text.

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

222. *See supra* subpart IV(A).

223. *See* Sant'Ambrogio, *supra* note 106, at 1415 ("It is difficult if not impossible to know in advance how much time many decisions will reasonably require.").

224. Abbott, *Case Against Deadlines*, *supra* note 81, at 185.

225. *See* sources cited *supra* note 122.

participate in setting the deadline could make the agency feel more responsible for achieving its objectives.[226]

Furthermore, Congress should not attempt to set short deadlines in order to indirectly bypass the notice-and-comment procedure required by APA § 553. This puts the agency in the position of choosing between missing the deadline and forgoing the notice-and-comment procedure, creating the risk that a court will set aside the agency action if it finds the agency did not have good cause to do so.[227] Thus, instead of exposing the agency to that risk, Congress should expressly waive the notice-and-comment requirements of the APA if it feels that the need for quick agency action outweighs the many benefits of utilizing the procedure.

C. *Resource Appropriation Should Be Increased in Line with Increases in Responsibility*

Lastly, one of the greatest challenges faced by agencies in meeting their statutorily imposed deadlines is the lack of funding needed to complete the task in a short timeframe.[228] In order to improve an agency's ability to comply with deadlines, Congress should better align the responsibilities assigned to agencies with the amount appropriated to each agency. One way that Congress could do this is by drafting deadlines that include specific checkpoints at which time an agency's progress would be assessed, perhaps by the appropriate oversight committee, and the agency could express any need for additional resources. This assessment process could make the agency better able to comply with deadlines by making sure it has the funds it needs. The process would also keep Congress informed on the agency's actions and how the appropriated resources are being utilized.

In addition, Congress has a unique opportunity to make sure the SEC has the funds it requires to meet deadlines. Currently, the SEC "submits its budget to the White House Office of Management and Budget."[229] For the SEC to actually receive the funds, Congress must authorize and appropriate the funds.[230] However, as noted before, the SEC is not funded through taxation but through its fee-collection mechanism, which often takes in more money than the SEC requests for its budget.[231] Thus, Congress could allow the SEC to obtain resources through self-funding, which could

---

226. Similarly, Abbott has suggested that requiring agencies to set nonbinding deadlines could make them "truly accountable for the failure" to meet those deadlines. *See* Abbott, *Case Against Deadlines*, *supra* note 81, at 201.

227. *See supra* section II(B)(2).

228. *See supra* section III(B)(2).

229. Seligman, *supra* note 198, at 253.

230. *See id.* (describing the process by which funds are authorized and appropriated and explaining that it often results in underfunding of the SEC).

231. Stewart, *supra* note 205.

eliminate the difference between the funds needed by the agency and the funds actually appropriated by Congress. Although it is not clear that Congress is willing to loosen the reins on the SEC's budget, it has recently considered it. The Senate version of what became the Dodd-Frank Act included a provision for self-funding for the SEC;[232] however, the final version of the Act did not.[233]

VI. Conclusion

Although the SEC has been the target of a great deal of criticism as a result of missing statutory deadlines imposed by the JOBS Act, the failure is at least partially due to Congress's problematic use of deadlines. This Note addresses this problem by setting forth several recommendations for improvement. First, in order to better align agency action with legislative intent, Congress should expressly state the priority of deadlines, both within a statutory scheme and with regard to other statutory schemes. Second, Congress should set deadlines that are more realistic by taking into consideration the opinions of each agency as to what time frame would be realistic based on its workload, other deadlines, and available resources. Lastly, Congress should better align the amount appropriated to each agency with the agency's responsibilities. By improving its use of deadlines, Congress can minimize the problems that are caused by deadlines and that impair each agency's ability to meet those deadlines.

—*Caitlin A. Bubar*

232. S. 3217, 111th Cong. § 991 (2010).

233. Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, § 991, 124 Stat. 1376, 1950–55 (2010) (codified in scattered sections of 15 U.S.C.).

* * *

A Status Quo Bias: Behavioral Economics and the Federal Preliminary Injunction Standard*

I. Introduction

A federal court's decision about whether to issue a preliminary injunction is one that is fraught with uncertainty and marred by a patchwork of differing standards and policy justifications.[1] The tests applied by the various federal circuit courts of appeals are diverse, with some requiring that all of the preliminary injunction factors meet a certain threshold[2] and others using a sliding scale approach that allows lesser showings on some factors when others are met more strongly.[3]

Some courts also consider the effect of a preliminary injunction on the status quo existing between the parties to the case.[4] Many courts note that the purpose of a preliminary injunction is to preserve the status quo existing between the parties.[5] Such statements of purpose, while notable, need not go beyond functioning as mere platitudes, throwaway lines in a judicial opinion that do not figure in the substantive test applied by the court.[6]

More interestingly, some federal circuit courts incorporate the status quo issue into their substantive preliminary injunction doctrine and require a greater showing by movants seeking preliminary relief that alters the

---

* I would like to thank Professor Sean Williams for his guidance during the drafting of this Note and Professor Mechele Dickerson, whose Remedies course inspired my interest in the law of preliminary injunctions. Thanks also go to the editors of the *Texas Law Review*, particularly Brent Rubin and Spencer Patton, for their very helpful efforts in preparing this Note for publication.

1. *See, e.g.*, Lea B. Vaughn, *A Need for Clarity: Toward a New Standard for Preliminary Injunctions*, 68 OR. L. REV. 839, 840–41 (1989) (noting that the "lack of uniformity" among the courts has led to "confusion" and "havoc in litigation"); Rachel A. Weisshaar, Note, *Hazy Shades of* Winter*: Resolving the Circuit Split over Preliminary Injunctions*, 65 VAND. L. REV. 1011, 1015, 1032–48 (2012) (discussing the continued disagreement among the circuit courts of appeals about whether a "sliding scale" approach or "sequential," elemental-style approach should be used and about whether a sliding scale is still permitted under current Supreme Court precedent).

2. *See, e.g.*, Bethany M. Bates, Note, *Reconciliation After* Winter*: The Standard for Preliminary Injunctions in Federal Courts*, 111 COLUM. L. REV. 1522, 1545–46 (2011) (outlining the Fourth Circuit doctrine, based upon that court's reading of the U.S. Supreme Court's decision in *Winter v. Natural Resources Defense Council, Inc.*, 555 U.S. 7 (2008), that preliminary injunction movants must meet all four traditional factors).

3. *See, e.g.*, *id.* at 1538 (concluding that the Second, Ninth, and Seventh Circuits continue to use a sliding scale approach when evaluating preliminary injunctions).

4. Thomas R. Lee, *Preliminary Injunctions and the Status Quo*, 58 WASH. & LEE L. REV. 109, 110 (2001).

5. Vaughn, *supra* note 1, at 849.

6. *Cf.* Lee, *supra* note 4, at 138 (discussing early American cases where the status quo was not offered as part of a test but merely as a statement of purpose or of what the status quo usually accomplishes).

status quo.[7] For example, one court requires that movants prove not only that the court's normal preliminary injunction factors are met but also that these factors "weigh heavily and compellingly" in the movant's favor.[8] Another court requires that a movant "must show not only a likelihood, but a clear or substantial likelihood, of success on the merits[] where the injunction sought is mandatory—i.e., it will alter rather than maintain, the status quo."[9]

These heightened requirements for preliminary injunctions that alter the status quo have been much criticized.[10] Judge Richard Posner gave a particularly incisive critique in a recent Seventh Circuit opinion: "Preliminary relief is properly sought only to avert irreparable harm to the moving party. Whether and in what sense the grant of relief would change or preserve some previous state of affairs is neither here nor there. To worry these questions is merely to fuzz up the legal standard."[11] If consideration of the status quo merely fuzzes up the legal standard, why are some courts worrying themselves over it?

Former Judge Michael McConnell of the Tenth Circuit, concurring in *O Centro Espirita Beneficiente Uniao do Vegetal v. Ashcroft*,[12] provided an interesting justification.[13] He wrote:

> Disrupting the status quo may provide a benefit to one party, but only by depriving the other party of some right he previously enjoyed. Although the harm and the benefit may be of equivalent magnitude on paper, in reality, deprivation of a thing already possessed is felt more acutely than lack of a benefit only hoped for.[14]

---

7. *Id.* at 115–16.

8. Fundamentalist Church of Jesus Christ of Latter-Day Saints v. Horne, 698 F.3d 1295, 1301 (10th Cir. 2012) (quoting Dominion Video Satellite, Inc. v. Echostar Satellite Corp., 269 F.3d 1149, 1154 (10th Cir. 2001)).

9. Louis Vuitton Malletier v. Dooney & Bourke, Inc., 454 F.3d 108, 114 (2d Cir. 2006) (internal quotation marks omitted).

10. *See, e.g.*, 11A CHARLES ALAN WRIGHT ET AL., FEDERAL PRACTICE AND PROCEDURE § 2948, at 127 (3d ed. 2013) (asserting that it is "regrettable" when consideration of the status quo leads to the denial of a preliminary injunction when the "important conditions" have been met); John Leubsdorf, *The Standard for Preliminary Injunctions*, 91 HARV. L. REV. 525, 546 (1978) ("Emphasis on preserving the status quo is a habit without a reason."); Vaughn, *supra* note 1, at 850 (arguing that the status quo "does not inform the deliberations for a preliminary injunction in any meaningful way").

11. Chi. United Indus., Ltd. v. City of Chi., 445 F.3d 940, 944 (7th Cir. 2006) (citations omitted).

12. 389 F.3d 973 (10th Cir. 2004) (en banc) (per curiam), *aff'd sub nom.* Gonzales v. O Centro Espirita Beneficiente Uniao do Vegetal, 546 U.S. 418 (2006).

13. *See* DOUGLAS LAYCOCK, MODERN AMERICAN REMEDIES: CASES AND MATERIALS 449 (4th ed. 2010) (discussing Judge McConnell's concurrence and his arguments in defense of the status quo-based heightened burden).

14. *O Centro*, 389 F.3d at 1015–16 (McConnell, J., concurring).

Judge McConnell went on to cite studies from the social sciences—in particular, the field of behavioral economics—to justify this assertion.[15] He discussed two well-established phenomena observed by researchers: "loss aversion" and the "endowment effect."[16] Each of these phenomena supports McConnell's basic point that losses loom larger than gains and that people value things to a greater degree when they already possess them.[17]

Setting aside Judge McConnell's focus on litigants, there are two other possible, more judge-centric explanations for the preoccupation with the status quo in some federal courts. Each explanation also invokes the research and findings of behavioral economics and psychology. Earlier in his concurring opinion, Judge McConnell unconsciously nodded towards these two explanations when he stated:

> Fundamentally, the reluctance to disturb the status quo prior to trial on the merits is an expression of judicial humility. As Judge Murphy points out, a court bears more direct moral responsibility for harms that result from its intervention than from its nonintervention, and more direct responsibility when it intervenes to change the status quo than when it intervenes to preserve it.[18]

Judge McConnell's concern about the "moral responsibility" of the court for status quo-altering interventions calls to mind the same sort of loss aversion he applied to litigants. But this passage, and the behavior of other courts in being suspicious of preliminary injunctions that alter the status quo, may reveal that loss aversion is making an impact on judges themselves. In particular, a species of loss aversion termed "status quo bias"—one that has been repeatedly confirmed experimentally and in the field[19]—provides one ready explanation for this judicial practice that has puzzled and bothered commentators and judges for years.[20] Furthermore, Judge McConnell's statements call to mind another psychological pattern: the heuristic that bad outcomes resulting from omissions are less morally blameworthy than bad outcomes resulting from commissions.[21] Such a judicial "omission bias" may provide another behaviorally based explanation for what underlies the hesitance to grant status quo-altering preliminary injunctions.

---

15. *Id.* at 1016.

16. *Id.*

17. *See id.* (explaining the meaning of loss aversion and the endowment effect and citing studies that demonstrate their existence).

18. *Id.* at 1015.

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

20. *See supra* notes 10–11 and accompanying text.

21. *See* Robert A. Prentice & Jonathan J. Koehler, *A Normality Bias in Legal Decision Making*, 88 CORNELL L. REV. 583, 587 (2003) (defining "omission bias" as "the tendency of people to find more blameworthy bad results that stem from actions than bad results that stem from otherwise equivalent omissions"); Cass R. Sunstein, *Moral Heuristics and Moral Framing*, 88 MINN. L. REV. 1556, 1575 (2004) (arguing that the "omission bias reflects a moral heuristic").

In the pages that follow, I will bring to bear the rich experimental literature of behavioral law and economics (BLE) in an area where it has never to this point been comprehensively applied: the role of the status quo in federal judicial decision making with respect to preliminary injunctions.[22] In so doing, I will argue that BLE can *explain* why federal courts concern themselves with the status quo at the preliminary injunction stage. More importantly, however, and contrary to Judge McConnell's view, I will demonstrate that BLE cannot *justify* invocation of the status quo in deciding whether to issue these injunctions. I will ultimately side with those courts and commentators that have explicitly rejected consideration of the status quo in deciding preliminary injunction motions.

This Note will proceed as follows. Part II will present a brief summary of the standards governing the adjudication of motions for preliminary injunction in the various federal circuit courts of appeals. I will pay special attention to how the status quo affects these standards. Part III will introduce the field of behavioral law and economics and summarize the relevant phenomena: loss aversion, the endowment effect, status quo bias, and omission bias. Furthermore, I will present past experimental research that suggests that judges are not immune to the sorts of irrationalities fundamental to BLE. In Part IV, I will argue that while some of these BLE-based phenomena can explain courts' invocation of the status quo in preliminary injunction decision making, none can justify it. I will go on to endorse the position held by many that the status quo has no place in this area of the law. I will briefly conclude in Part V.

II. The Law of Preliminary Injunctions

In its recent decision in *Winter v. Natural Resources Defense Council, Inc.*,[23] the U.S. Supreme Court held that a "plaintiff seeking a preliminary injunction must establish that he is likely to succeed on the merits, that he is likely to suffer irreparable harm in the absence of preliminary relief, that the balance of equities tips in his favor, and that an injunction is in the public interest."[24] This four-factor test generally aligns with the traditional one

22. One scholar has to this point connected BLE and the issue of the status quo in preliminary injunctions. *See* Eyal Zamir, *Loss Aversion and the Law*, 65 VAND. L. REV. 829, 869 (2012) (noting that "courts are far more willing to issue preliminary injunctions that preserve the status quo than preliminary injunctions that disrupt or alter it" and arguing that this "tendency is compatible with loss aversion"). Nonetheless, Zamir's article only notionally referred to the issue and did not explore the standards used by the relevant courts, in particular some courts' employment of heightened standards of review for status quo-altering injunctions. Furthermore, he did not appraise the normative claim that courts should take seriously litigant loss aversion or consider the impact of the omission bias, either as a descriptive or normative matter. This Note is the first legal scholarship to both fully explore the law in this area and tie it to BLE research.

23. 555 U.S. 7 (2008).

24. *Id.* at 20.

outlined by authorities in the field.[25] The apparently broad agreement glides above a murkier situation in which many commentators over the years have criticized the varied preliminary injunction standards applied by the federal courts.[26] Indeed, despite the Supreme Court's weighing in with a four-factor test of its own in *Winter*, the issue of how to apply these factors is still greatly unsettled, with substantial dispute as to whether the traditional four factors function as independent elements or if they are better construed as operating on a sliding scale.[27]

Meanwhile, the Supreme Court has not resolved what role the status quo should play in preliminary injunction doctrine, prolonging a fundamental disagreement among the federal circuit courts of appeals.[28] Some of these courts refer to the status quo when deciding whether to issue preliminary injunctions.[29] They often do so in two contexts: first, in stating that the purpose of a preliminary injunction is to preserve the status quo[30] and, second, in providing that a heightened showing is required of movants who seek to alter the status quo through such an injunction.[31] The following discussion will focus on the second context.

A. A Heightened Standard for Changes to the Status Quo

Wright and Miller's esteemed treatise *Federal Practice and Procedure* sets forth that the purpose of a preliminary injunction is "to protect [the] plaintiff from irreparable injury and to preserve the court's power to render a meaningful decision after a trial on the merits."[32] This comports with the statement of purpose handed down by the Supreme Court in *University of Texas v. Camenisch*[33]: "The purpose of a preliminary injunction is merely to preserve the relative positions of the parties until a trial on the merits can be

---

25. *See* Lee, *supra* note 4, at 111 & n.5 (summarizing the four factors as including, in prongs two and three, the irreparable harms to the plaintiff or defendant should the preliminary injunction be denied or granted respectively and collecting cases citing this framework); Vaughn, *supra* note 1, at 839 (providing four factors and apparently replacing the equities prong of the Supreme Court test with a "balance of harms" test with respect to the plaintiff and defendant); WRIGHT ET AL., *supra* note 10, at 122–24 (stating a similar formulation to Vaughn).

26. *See, e.g.*, Morton Denlow, *The Motion for a Preliminary Injunction: Time for a Uniform Federal Standard*, 22 REV. LITIG. 495, 539 (2003) (concluding that "in the absence of a definitive Supreme Court decision, the standards applied by the circuit courts of appeals are not consistent," leading to "confusion for lawyers and judges and an unpredictable application by the courts").

27. *See* Bates, *supra* note 2, at 1537–38 (summarizing the circuit split over how to interpret *Winter* between the Second, Seventh, and Ninth Circuits on the one hand and the Fourth Circuit on the other).

28. *Cf.* Lee, *supra* note 4 (stating that the federal circuit courts of appeals are in "substantial disarray" on how to treat the status quo issue).

29. *Id.*

30. *See infra* notes 34–35 and accompanying text.

31. *See infra* sections II(A)(1)–(3).

32. WRIGHT ET AL., *supra* note 10, § 2947, at 112.

33. 451 U.S. 390 (1981).

held."[34] The Court's reference to "preserv[ing] the relative positions of the parties" seems to touch upon the status quo issue. Indeed, preservation of the status quo is a commonly identified purpose of preliminary injunctions.[35]

Some courts go beyond mere statements of purpose, however, and respond to the status quo by requiring a heightened burden of proof from movants seeking preliminary relief that alters the status quo. The Tenth, Ninth, and Second Circuits all require this heightened burden in some form.[36] Language about the disfavored nature of status quo-altering preliminary injunctions also crops up in stray opinions of courts in other circuits, leaving an odd and unpredictable patchwork.[37]

1. Tenth Circuit and the Three Forms of Disfavored Relief.—The Tenth Circuit held in *SCFC ILC, Inc. v. Visa USA, Inc.*[38] that three types of preliminary injunctions are disfavored by that court: (1) those that "disturb[] the status quo," (2) those that are "mandatory as opposed to prohibitory," and (3) those that give the movant "substantially all the relief" it is seeking.[39] When one of these types of relief is requested, the court held, a movant must show that the traditional four factors weigh "heavily and compellingly" in its favor.[40] The court justified its holding by stating that "[a] preliminary injunction that alters the status quo goes beyond the traditional purpose for preliminary injunctions, which is only to preserve the status quo until a trial on the merits may be had."[41] A later Tenth Circuit panel stated that the purpose of the heightened burden is to "minimize any injury that would not have occurred but for the court's intervention."[42]

The court went on to retreat from the "heavily and compellingly" standard, though not the heightened burden, in *O Centro Espirita Beneficiente Uniao do Vegetal v. Ashcroft*.[43] The per curiam opinion, issued by the court sitting en banc, "jettison[ed]" the "heavily and

34. *Id.* at 395.

35. Vaughn, *supra* note 1, at 849.

36. *See* Lee, *supra* note 4, at 115–16 (indicating that these circuits have all adopted a "bifurcated preliminary injunction standard" that "imposes a heavier burden where the moving party seeks to upset the status quo"). Lee's piece provided an essential jumping-off point for my discussion in this subpart, which has been supplemented by research from more recent cases.

37. *See infra* section II(A)(3).

38. 936 F.2d 1096 (10th Cir. 1991), *overruled in part on other grounds by* O Centro Espirita Beneficiente Uniao do Vegetal v. Ashcroft, 389 F.3d 973 (10th Cir. 2004) (en banc) (per curiam).

39. *Id.* at 1098–99.

40. *Id.*

41. *Id.* at 1099.

42. RoDa Drilling Co. v. Siegal, 552 F.3d 1203, 1209 (10th Cir. 2009).

43. 389 F.3d 973, 975 (10th Cir. 2004) (en banc) (per curiam), *aff'd sub nom.* Gonzales v. O Centro Espirita Beneficiente Uniao do Vegetal, 546 U.S. 418 (2006).

compellingly" standard and instead required that requests for relief in the disfavored categories be "more closely scrutinized" in recognition of the "extraordinary" nature of the remedy.[44] To further tip the scales against the disfavored injunction types, the court stated that "a party seeking such an injunction must make a strong showing both with regard to the likelihood of success on the merits and with regard to the balance of harms, and may not rely on our modified likelihood-of-success-on-the-merits standard."[45] Under the modified standard, a movant who can demonstrate that the irreparable harm, balance of harms, and public interest factors are in its favor need only show that there are serious and doubtful questions ripe for litigation to satisfy the first factor (likelihood of success on the merits).[46]

Despite the en banc court's clear intention to "jettison" the "heavily and compellingly" standard, it appears to have been revived recently. In an opinion published in 2012, a Tenth Circuit panel stated that a movant seeking a disfavored preliminary injunction must meet the "heavily and compellingly" standard.[47] The case has since been cited by two other Tenth Circuit decisions for the same standard.[48]

2. Second and Ninth Circuits and Mandatory Injunctions.—The Second and Ninth Circuits adopt a slightly different—and less demanding—approach to the status quo. Both courts base their heightened burden on whether the injunction sought is mandatory as opposed to prohibitory, with mandatory injunctions being defined in part as those that will alter the status quo.[49] *Black's Law Dictionary* defines a mandatory injunction as one which "orders an affirmative act or mandates a specified course of conduct."[50] A prohibitory injunction, on the other hand, is defined as one "that forbids or restrains an act."[51] The Second Circuit's basic preliminary injunction standard requires that a movant show "(1) irreparable harm and (2) either (a) a likelihood of success on the merits,

44. *Id.*

45. *Id.* at 976.

46. *Id.* at 1002 (Seymour, J., concurring in part and dissenting in part).

47. Fundamentalist Church of Jesus Christ of Latter-Day Saints v. Horne, 698 F.3d 1295, 1301 (10th Cir. 2012).

48. Hobby Lobby Stores, Inc. v. Sebelius, No. 12-6294, 2012 WL 6930302, at *1 (10th Cir. Dec. 20, 2012); Lodgeworks, L.P. v. C.F. Jordan Constr., LLC, 506 F. App'x 747, 751 n.3 (10th Cir. 2012).

49. *See* Sunward Elecs., Inc. v. McDonald, 362 F.3d 17, 24 (2d Cir. 2004) (changing the standard of review when the "injunction sought is mandatory—i.e., it will alter, rather than maintain, the status quo"); Stanley v. Univ. of S. Cal., 13 F.3d 1313, 1320 (9th Cir. 1994) (stating that prohibitory injunctions simply preserve the status quo whereas mandatory injunctions go "well beyond simply maintaining the status quo *pendente lite*" and are thus "particularly disfavored" (quoting Anderson v. United States, 612 F.2d 1112, 1114 (9th Cir. 1980)) (internal quotation marks omitted)).

50. BLACK'S LAW DICTIONARY 855 (9th ed. 2009).

51. *Id.*

or (b) sufficiently serious questions going to the merits of its claims to make them fair ground for litigation, plus a balance of the hardships tipping decidedly in favor of the moving party."[52] When a mandatory injunction is sought, the movant must make a heightened showing under prong two that there is a "clear or substantial likelihood of success."[53]

The Ninth Circuit requires those seeking mandatory preliminary injunctions to establish that "the facts and law clearly favor the moving party."[54] District courts appear to give this standard serious consideration. For example, in *Korab v. McManaman*,[55] the court evaluated the plaintiffs' motion for a preliminary injunction of implementation of a health program for certain lawful immigrant residents.[56] The plaintiffs' motion turned in part on a claim that the program violated the Equal Protection Clause of the Fourteenth Amendment.[57] The court found that the plaintiffs' constitutional claim had "some" likelihood of ultimate success on the merits but that the law did not appear to be "clearly" in their favor.[58] The court denied the motion, citing the plaintiffs' failure to establish a "clear likelihood of success on the merits, irreparable harm to the class, or that the balance of the equities and/or public interest weigh in their favor."[59]

As an aside, it is worth noting that basing a standard on the distinction between the mandatory or preliminary character of an injunction has a couple of obvious flaws. First, any injunction stated in mandatory terms can rather easily be reformulated in prohibitory language.[60] For example, a mandatory injunction requiring a party to "perform your contractual obligations" can be changed to the prohibitory command "do not breach your contract." Second, Professor John Leubsdorf has persuasively argued that distinguishing between prohibitory and mandatory injunctions does little to determine whether one or the other will preserve the status quo.[61] Professor Thomas Lee found that the Second Circuit has been rather candid

52. Oneida Nation of N.Y. v. Cuomo, 645 F.3d 154, 164 (2d Cir. 2011) (quoting Monseratte v. N.Y. State Senate, 599 F.3d 148, 154 (2d Cir. 2010)).

53. *Sunward Elecs.*, 362 F.3d at 24 (quoting Tom Doherty Assocs., Inc. v. Saban Entm't, Inc., 60 F.3d 27, 34 (2d Cir. 1995)) (internal quotation marks omitted).

54. *Stanley*, 13 F.3d at 1320 (quoting *Anderson*, 612 F.2d at 1114).

55. 805 F. Supp. 2d 1027 (D. Haw. 2011).

56. *Id.* at 1029–30.

57. *Id.* at 1030.

58. *Id.* at 1040–41 (emphasis omitted).

59. *Id.* at 1042.

60. *See Developments in the Law—Injunctions*, 78 HARV. L. REV. 994, 1062 (1965) (arguing that any rule based on the distinction between mandatory and prohibitory injunctions is "ridiculously easy to circumvent" and any mandatory injunction can be restated in prohibitory terms).

61. *See* Leubsdorf, *supra* note 10, at 535 (asserting that "whether an injunction is mandatory has nothing to do with whether it preserves the status quo").

in noting the difficulty of making this distinction.[62] These admissions notwithstanding, the mandatory–prohibitory distinction continues to prevail in the Second and Ninth Circuits.

3. Incoherence in Other Circuits.—Some other circuits have made reference to the status quo in ways that are unpredictable and largely incoherent.[63] The Fourth Circuit provides an apt example. In the recent case of *Pashby v. Delia*,[64] the court stated, "[A] preliminary injunction's tendency to preserve the status quo determines whether it is prohibitory or mandatory."[65] If mandatory and, therefore, status quo altering, the *Pashby* court stated it would apply a heightened standard of review.[66] The opinion cited a previous Fourth Circuit case which held, "Mandatory preliminary injunctions do not preserve the status quo and normally should be granted only in those circumstances when the exigencies of the situation demand such relief."[67] In apparent contradiction to *Pashby*, however, a different Fourth Circuit panel has held that mandatory injunctions and "preservation of the *status quo*" do not require any different legal test than that normally applied when evaluating motions for preliminary relief.[68]

The Third Circuit evinces a similar split of authority. One of that court's opinions cited the Second Circuit for the rule that parties seeking mandatory, status quo-altering preliminary injunctions "must meet a higher standard of showing irreparable harm in the absence of an injunction."[69] Another opinion stated the opposite rule: "[W]e disagree that the preservation of the *status quo* operates as a separate test" and the legal standard is actually governed by the "traditional four-pronged test."[70]

---

62. *See* Lee, *supra* note 4, at 119 (observing that the Second Circuit has acknowledged the "definitional ambiguities" in the mandatory–prohibitory distinction and has granted that the "proposed dichotomy" between these two types of injunctions "is illusory").

63. *See, e.g.*, *id.* at 122 & n.60 (noting that "in some instances, the same circuits that have suggested some ambiguous relevance of the status quo elsewhere have questioned the viability of a variable standard" and citing examples).

64. 709 F.3d 307 (4th Cir. 2013).

65. *Id.* at 320. Another recent Fourth Circuit panel echoed these sorts of status quo considerations. *See* Perry v. Judd, 471 F. App'x 219, 223–24 (4th Cir. 2012) (stating that mandatory preliminary injunctions are "disfavored" and review is "even more searching" than normal when one is sought (quoting *In re* Microsoft Corp. Antitrust Litig., 333 F.3d 517, 525 (4th Cir. 2003))).

66. *Pashby*, 709 F.3d at 320.

67. Wetzel v. Edwards, 635 F.2d 283, 286 (4th Cir. 1980), *quoted in Pashby*, 709 F.3d at 320.

68. Rum Creek Coal Sales, Inc. v. Caperton, 926 F.2d 353, 360 (4th Cir. 1991) (quoting Ortho Pharm. Corp. v. Amgen, Inc., 882 F.2d 806, 814 (3d Cir. 1989)).

69. Bennington Foods LLC v. St. Croix Renaissance, Grp., LLP, 528 F.3d 176, 179 (3d Cir. 2008) (citing Tom Doherty Assocs., Inc. v. Saban Entm't, Inc., 60 F.3d 27, 33–34 (2d Cir. 1995)).

70. *Ortho Pharm. Corp.*, 882 F.2d at 813 (internal quotation marks omitted).

C. *Judicial Disregard of the Status Quo*

Some courts and judges have, however, been clear in their disregard of the status quo. Referring to *United Food & Commercial Workers Union, Local 1099 v. Southwest Ohio Regional Transit Authority*,[71] Professor Lee noted that the Sixth Circuit "most explicitly has confronted and rejected the heightened burden adopted" by some courts.[72] In that case, the court held that there is "little consequential importance to the concept of the status quo" and that the "distinction between mandatory and prohibitory injunctive relief is not meaningful."[73] The court explicitly rejected the Tenth Circuit "heavily and compellingly" test.[74] In the Seventh Circuit, Judge Richard Posner has also been clear in his rejection of the status quo test.[75] Judge Posner has persuasively stated the view that the true purpose of preliminary relief is to "avert irreparable harm to the moving party" and that consideration of the status quo does not assist the court in advancing this purpose.[76]

Against this backdrop of case law, one must ask: why are courts so divided on this point? Some courts hold tightly to status quo-based heightened standards. Other courts (and most commentators) discard such tests, sometimes deriding them as meritless and conceptually incoherent. Recalling Judge McConnell's discussion in *O Centro*, I will now introduce the field of behavioral law and economics to determine if it can provide an explanation or a justification for those federal courts that require heightened showings when reviewing motions for status quo-altering preliminary injunctions.

III. Behavioral Law and Economics

Law and economics has been defined as the "systematic application of theories of rational choice to legal problems."[77] The field seeks to "determine the implications of . . . rational maximizing behavior in and out of markets, and its legal implications for markets and other institutions."[78] Generally speaking, the field assumes that individuals have a "stable set of preferences" and will "maximize their utility" by their actions, thus

---

71. 163 F.3d 341 (6th Cir. 1998).

72. Lee, *supra* note 4, at 123.

73. *United Food*, 163 F.3d at 348.

74. *Id.*

75. *See* Roland Mach. Co. v. Dresser Indus., Inc., 749 F.2d 380, 383 (7th Cir. 1984) (discussing the ambiguity of the status quo concept and giving a dismissive treatment to the issue).

76. Chi. United Indus., Ltd. v. City of Chi., 445 F.3d 940, 944 (7th Cir. 2006).

77. Gary Lawson, *Efficiency and Individualism*, 42 DUKE L.J. 53, 53 n.1 (1992).

78. Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471, 1476 (1998).

functioning as so-called "rational actor[s]."[79] Behavioral law and economics, on the other hand, analyzes law from the perspective of observed human behavior patterns rather than a hypothesized rational actor.[80] The research underlying BLE has repeatedly demonstrated a number of ways in which people deviate from the utility-maximizing picture of the rational actor provided by classical law and economics scholars.[81] I will elaborate a few key deviations that will be helpful to understanding the issue of the status quo and preliminary injunctions.

A. Loss Aversion and the Endowment Effect

A core component of the classical economic model is expected utility theory, which rests on two key assumptions: (1) a choice's utility is based on the utility of its ultimate outcome and (2) the calculation of utility will be based on the final quantity of assets that result from the choice, regardless of whether a gain or loss occurred from the pre-choice state.[82] In short, the theory holds that preferences will not be affected by what a person possesses when they make a decision.[83] This tenet of choice theory—"reference independence"—was flatly contradicted by experiments conducted by Amos Tversky and Daniel Kahneman, who found that their results were better explained by a "reference-dependent" model.[84] Reference dependence "suggests that values are coded as gains and losses relative to a reference point."[85]

Reference dependence manifests itself in the concept of "loss aversion," which is defined by the simple axiom that "losses . . . loom larger than . . . gains."[86] In other words, a loss produces greater disutility than the utility produced by an equally sized gain. For an example of loss aversion, consider its application to consumer choices. Loss aversion implies that a consumer would experience a larger reduction in happiness from a price increase of a good (over a certain reference price) than the

---

79. *Id.* at 1473, 1476 (quoting GARY S. BECKER, THE ECONOMIC APPROACH TO HUMAN BEHAVIOR 14 (1976)).

80. *See id.* at 1476 (explaining that behavioral law and economics explores "the implications of *actual* (not hypothesized) human behavior for the law").

81. *See id.* at 1476–77 (noting aspects of BLE that "draw into question" these classical economic ideas and represent "significant way[s] in which most people depart from the standard economic model").

82. Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263, 263–64 (1979).

83. Amos Tversky & Daniel Kahneman, *Loss Aversion in Riskless Choice: A Reference-Dependent Model*, 106 Q.J. ECON. 1039, 1039 (1991).

84. *Id.* at 1046 (emphasis omitted).

85. Dilip Soman, *Framing, Loss Aversion, and Mental Accounting*, *in* BLACKWELL HANDBOOK OF JUDGMENT AND DECISION MAKING 379, 383 (Derek J. Koehler & Nigel Harvey eds., 2004).

86. Tversky & Kahneman, *supra* note 83, at 1047.

increase in happiness he or she would experience because of a price decrease of equal magnitude.[87] In response, the consumer would cut back on purchases to a greater degree in the price-increase condition than he or she would boost his or her purchases in the price-decrease condition.[88] Studies of consumer choice have demonstrated such an effect in the markets for eggs and orange juice.[89]

An important corollary of loss aversion is the "endowment effect." This term describes the phenomenon that "people tend to value goods more when they own them than when they do not."[90] This effect was demonstrated by an experiment involving Cornell undergraduate students conducted by Daniel Kahneman, Jack Knetsch, and Richard Thaler.[91] Half of the students were given mugs, and a market was created.[92] Each holder of a mug could give up his or her property for a set market price, and, alternatively, each mug-less student could pay the market price and receive a mug.[93] Because the mugs were distributed without regard to the preferences of the subjects, classical economic theory assumes that the buyers and sellers would roughly meet at the market price and that about half of the mugs would be traded (because about half of the students should have liked mugs more than the other half of students).[94] In reality, the mug holders demanded roughly two times more money than the prospective buyers were willing to pay.[95] As a consequence, the number of trades was less than half the expected amount.[96]

Evaluating experiments like this, one pair of commentators has noted that the power of the endowment effect is demonstrated by its appearance in situations ranging from the transfer of "banal" goods like mugs and pens to the valuation of public goods like parks and wildlife.[97] The effect has been called "undoubtedly the most significant single finding from behavioral economics for legal analysis to date."[98] As just one example of its impact,

---

87. Colin F. Camerer, *Prospect Theory in the Wild: Evidence from the Field*, *in* ADVANCES IN BEHAVIORAL ECONOMICS 148, 152 (Colin F. Camerer et al. eds., 2004).

88. *Id.*

89. *Id.*

90. Russell Korobkin, *The Endowment Effect and Legal Analysis*, 97 NW. U. L. REV. 1227, 1228 (2003).

91. Daniel Kahneman et al., *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, J. ECON. PERSP., Winter 1991, at 193, 194–95.

92. *Id.* at 195. A second experiment involving ballpoint pens was also conducted. *Id.* It had the same basic format and produced similar results. *Id.* at 195–96.

93. *Id.* at 195.

94. *Id.* at 195–96.

95. *Id.* at 196.

96. *Id.*

97. Jeffrey J. Rachlinski & Forest Jourden, *Remedies and the Psychology of Ownership*, 51 VAND. L. REV. 1541, 1552 (1998).

98. Korobkin, *supra* note 90, at 1229.

the hugely influential Coase theorem—which holds that the "assignment of a legal entitlement by the state will not affect the ultimate ownership of that entitlement" because people will bargain for that entitlement should they want it—is called into question by the endowment effect.[99] The observed distance between buying and selling prices caused by the endowment effect might make the law's placement of an entitlement quite sticky.[100]

B. Status Quo Bias

Status quo bias is a phenomenon showing that people's choices are affected by consideration of whether a course of action will involve a departure from the status quo.[101] The bias is heightened when a person is presented with more choices and reduced when a person's preferences are more clearly held.[102]

The effect was first named and demonstrated in a paper by William Samuelson and Richard Zeckhauser.[103] The two men conducted a series of experiments involving students at the Boston University School of Management and the Kennedy School of Government at Harvard University.[104] They presented the subjects with questionnaires asking them to make a series of decisions.[105] In the neutral condition, the decision options were presented as new alternatives.[106] In the experimental condition, the first option was designated the status quo.[107] Samuelson and Zeckhauser's results indicated a robust change in response rate between the neutral and experimental conditions. When an option was in the status quo position, it received the most selections; in the neutral condition, the same option received fewer; and in the position of being an alternative to the designated status quo, it received even fewer.[108] The data also showed that relatively unpopular options received the greatest boost from appearing in

---

99. *See id.* at 1231–32 (noting that if there is a divergence between the amount one is willing to pay for an entitlement and the amount one will accept to sell said entitlement, then "the Coase Theorem [is] incorrect, or at least incomplete").

100. *See id.* (expounding that an implication of the endowment effect is that "the broad range of legal prescriptions based on Coase's insight requires reevaluation").

101. *See* Kahneman et al., *supra* note 91, at 197–98 (explaining that the bias, as an implication of loss aversion, emerges "because the disadvantages of leaving [the status quo] loom larger than advantages").

102. William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7, 8 (1988).

103. *Id.*

104. *Id.* at 14.

105. *Id.* at 12, 14.

106. *Id.* at 12–13.

107. *Id.*

108. *Id.* at 14.

the status quo position,[109] suggesting that the status quo bias can do substantial work when preferences are less strongly held.

Status quo bias can have major effects in the realm of public policy. Take the case of Pennsylvania's and New Jersey's automobile insurance reforms. Each state gave drivers a choice about their car insurance: they could either keep the unfettered right to sue in tort for accidents—and pay more for their policy—or have their right to sue restricted and pay less in premiums.[110] The two states presented this choice differently in an important respect: New Jersey made the restricted-right-to-sue-but-cheaper-policy option the default—with the corresponding requirement that the driver essentially buy the expanded right to sue—while Pennsylvania made the full right to sue a default.[111] The empirical evidence appears to show the impact of the status quo bias: 75% of Pennsylvanians kept their right to sue while only 20% of New Jersey drivers were willing to leave the status quo and pick the exact same option.[112]

C. *Omission Bias*

Omission bias refers to "the tendency to judge harmful acts as worse than equally harmful omissions."[113] It has been experimentally demonstrated in numerous papers.[114] For example, one experiment involved a questionnaire completed by fifty-seven University of Pennsylvania students.[115] One of the questions presented a fictional scenario where two tennis competitors share a prematch meal and one of the players knows that the house salad dressing will upset his competitor's stomach.[116] The respondents were asked to rate the wrongfulness of a series of different courses of conduct, including situations where (1) the player allows his competitor to order the bad dressing by saying nothing about the potential for sickness and (2) the player actively seeks to influence the

---

109. *Id.* at 19.

110. Kahneman et al., *supra* note 91, at 199.

111. Eric J. Johnson et al., *Framing, Probability Distortions, and Insurance Decisions*, 7 J. RISK & UNCERTAINTY 35, 48 (1993).

112. *See id.* (asserting that this example "illustrates that framing can have sizable economic consequences").

113. Prentice & Koehler, *supra* note 21, at 593.

114. *E.g.*, Jonathan Baron & Ilana Ritov, *Omission Bias, Individual Differences, and Normality*, 94 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 74, 83–84 (2004); Johanna H. Kordes-de Vaal, *Intention and the Omission Bias: Omissions Perceived as Nondecisions*, 93 ACTA PSYCHOLOGICA 161, 169 (1996); Ilana Ritov & Jonathan Baron, *Status-Quo and Omission Biases*, 5 J. RISK & UNCERTAINTY 49, 59–60 (1992) [hereinafter Ritov & Baron, *Status-Quo and Omission Biases*].

115. Mark Spranca et al., *Omission and Commission in Judgment and Choice*, 27 J. EXPERIMENTAL SOC. PSYCHOL. 76, 82 (1991).

116. *Id.*

competitor's decision in favor of the house dressing.[117] Sixty-five percent of those surveyed rated the omissions in this scenario as less bad than corresponding commissions.[118]

There are a number of explanations for the presence of this bias. One is based on a "norm-theory account" proposed by Kahneman and Miller.[119] Under this theory, "omissions tend to be considered as the norm, and commissions tend to be compared to what would have happened if nothing had been done."[120] As a result, omissions are seen as neutral whether they result in positive or negative outcomes.[121] The response to commissions, on the other hand, correlates with the nature of the outcome.[122] The aversion to harmful commissions leads to peculiar interactions with the decision whether to act. Professors Prentice and Koehler, surveying research in this field, concluded, "People are so averse to injuring others actively, that they will remain passive even when they know that more people will probably be hurt by their passivity."[123]

Omission bias seems justifiable as a rule of thumb in some cases. For example, bad outcomes from omissions are less likely to demonstrate specific intent to do harm than commissions because the general intent to act at all is absent.[124] Taken a step further, one would assume that a person with intent to do harm will more often cause harm (and go to the effort needed to cause it) than those who have no intent one way or the other. In light of the value it sometimes provides, Professor Sunstein has characterized the omission bias as a "moral heuristic," which he defines as "moral shortcuts, or rules of thumb, that work well most of the time, but that also systematically misfire."[125] This systematic misfiring can be seen in experiments where subjects found harmful commissions more blameworthy than omissions despite the fact that intent and knowledge of the actors (or nonactors as the scenario dictated) were held constant.[126] With no distinctions in knowledge or intent, the omission–commission

117. *Id.* at 82–83.

118. *Id.* at 84.

119. Ritov & Baron, *Status-Quo and Omission Biases*, *supra* note 114, at 60.

120. *Id.*

121. *Id.*

122. *Id.*

123. Prentice & Koehler, *supra* note 21, at 592.

124. *See* Spranca et al., *supra* note 115, at 76.

125. Sunstein, *supra* note 21, at 1558, 1575; *see also* Frances Howard-Snyder, *Doing vs. Allowing Harm*, STAN. ENCYCLOPEDIA PHIL., http://plato.stanford.edu/archives/win2011/entries/doing-allowing/ (last modified Dec. 20, 2011) (concluding that there is a group of situations where the distinction between doing and allowing harm leads to a "conflict between theory and intuition[]").

126. Spranca et al., *supra* note 115, at 101–02.

distinction should lose all relevance in evaluating whether a party was blameworthy.[127]

D. *Experimental and Survey Demonstrations of Heuristics and Biases in Judges*

A series of law review articles hypothesize that judges are subject to the same biases and heuristics that appear in the average person's thinking. Using experimental research involving actual judges, these articles provide evidence that judges are afflicted by the same bounded rationality observed by behavioral psychologists in the general population.[128]

One article, for example, surveyed 167 federal magistrate judges with a questionnaire that sought to draw out the presence of five particular heuristics and biases.[129] The authors found that the surveyed judges were susceptible to all five of these deviations from perfect rationality.[130] Furthermore, they compared their research with experimental findings in broader populations and found that judges were equally as susceptible as laypeople to three of those deviations (with judges being slightly less susceptible than other decision makers to framing and the representativeness heuristic).[131] The authors were appropriately cautious in noting that the presence of these biases in judges' thinking in the experimental setting did not necessarily mean that the biases would affect their courtroom decision making.[132] Nonetheless, the authors found anecdotal support for the functional importance of these heuristics and biases in the content of certain legal doctrine either created or applied by judges.[133]

Heuristics and biases can also potentially explain judicial behaviors observable in the real world. Professors Guthrie and George have studied a phenomenon in the federal circuit courts of appeals that they term the "affirmance effect."[134] In short, their research has found that these courts

---

127. Howard-Snyder, *supra* note 125 (arguing "that there is no decisive reason to say that any of these distinctions is morally significant, as long, that is, as we remember that intention plays no part in the distinction between doing and allowing harm").

128. *See, e.g.*, Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 829 (2001) (concluding that their "study demonstrates that judges rely on the same cognitive decision-making process as laypersons and other experts, which leaves them vulnerable to cognitive illusions that can produce poor judgments").

129. *Id.* at 786–87.

130. *Id.* at 816.

131. *Id.*

132. *Id.* at 819.

133. *See id.* at 821 (arguing that heuristics and biases are present in courts' application of the doctrine of res ipsa loquitur in tort and the prudent-investor rule in trustee liability cases).

134. *See* Chris Guthrie & Tracey E. George, *The Futility of Appeal: Disciplinary Insights into the "Affirmance Effect" on the United States Courts of Appeals*, 32 FLA. ST. U. L. REV. 357, 358

have affirmed decisions up for review at a high and steadily rising rate.[135] For example, the October 2002 term saw the circuit courts affirm 91% of rulings, a number much higher than that shown by the U.S. Supreme Court, which hears far fewer cases.[136] The professors present a series of possible explanations for this fact.[137] Among these, they note the possible influence of the status quo bias and the omission bias.[138] Given the experimental findings discussed above, it is not an unreasonable connection to make that appellate judges are being influenced by the same heuristics and biases that were shown to influence trial judges.

Another article surveyed a number of areas of the common law where the courts appear to be crafting doctrine in a way that reflects a differential treatment of gains and losses. Such areas of law include adverse possession, limitations on lost-profits damages, and the disparate treatment of completed gifts and unenforceable gratuitous promises.[139] The authors term this differing treatment of gains and losses the "valuation disparity."[140] The difference in terminology notwithstanding, the authors explicitly link their discussion to the research underlying behavioral economics and the concepts of loss aversion and the endowment effect.[141] Because the common law develops as an "intuitive, non-empirical interpretation of community mores and individual preferences," the authors argue that it is not surprising to find federal judges developing legal doctrine in reflection of the "valuation disparity."[142] As I will argue in the next Part, it is similarly unsurprising to find federal judges relying upon the heuristics and biases explained by behavioral economics while developing doctrine in the judge-controlled area of preliminary injunction law.

IV. Can BLE Explain or Justify the Judicial Preoccupation with the Status Quo in Federal Preliminary Injunction Doctrine?

With the law of preliminary injunctions and the research underlying BLE in mind, I will proceed to ask two fundamental questions. First, do heuristics and biases explain why some federal courts are preoccupied with the status quo when considering motions for preliminary injunctions?

(2005) (quoting Kevin M. Clermont & Theodore Eisenberg, *Litigation Realities*, 88 CORNELL L. REV. 119, 150 (2002)).

135. *See id.* at 360 (indicating an affirmance rate of 72% in 1945 and of 91% in 2003).

136. *Id.* at 359 & n.13.

137. *See id.* at 358 (providing explanations drawn from political science and the rational-actor model on the one hand and behavioral economics and the model of bounded rationality on the other).

138. *Id.* at 377–80.

139. David Cohen & Jack L. Knetsch, *Judicial Choice and Disparities Between Measures of Economic Values*, 30 OSGOODE HALL L.J. 737, 751–53, 757–59 (1992).

140. *Id.* at 737.

141. *Id.* at 743–45.

142. *Id.* at 749.

Second, if BLE explains this issue, are heightened burdens for status quo-altering preliminary injunctions justified by BLE theories? I will explore these questions by first looking at the possibility that judges are behaving sensitively in response to loss aversion of litigants. I will next discuss whether judges are instead exhibiting bounded rationality themselves and being influenced in their thinking by the status quo bias and the omission bias.

A. Litigant Loss Aversion

In the introduction to this Note, I mentioned Judge McConnell's explanation of the Tenth Circuit's doctrine with respect to status quo-altering injunctions.[143] He stated that the court's heightened burden was a reasonable response to the loss aversion of nonmovants, who would feel more acutely the loss of the status quo than the movants would feel the gain from its alteration.[144]

As a matter of the psychology of judges, there is no particular way to measure whether or not loss aversion is actually a factor working upon the subconscious of judges. It is worth noting that federal judges are not explicitly invoking this rationale to justify the issuance or denial of preliminary injunctions in *any* context. A search of all federal cases on Westlaw revealed that Judge McConnell's concurrence was the only case in the entire database which used the words "loss aversion" or "behavioral law and economics" (or similar phrases) in concurrence with the phrase "preliminary injunction."[145] Nonetheless, I cannot prove that judges are not considering this issue outside of the four corners of their written opinions, and so it may indeed provide some sort of explanation.

Judge McConnell's rationale, however, provides minimal justification for the heightened burden applied by the Tenth Circuit and other courts. Professor Laycock has convincingly argued that the status quo in a case is often very difficult to define, with both parties able to legitimately lay claim to their desired outcome being the status quo.[146] For example, as Professor Laycock points out, there was a muddled status quo determination in the very case where Judge McConnell made his litigant-loss-aversion

---

143. *See supra* text accompanying notes 12–17.

144. O Centro Espirita Beneficiente Uniao do Vegetal v. Ashcroft, 389 F.3d 973, 1015–16 (10th Cir. 2004) (McConnell, J., concurring), *aff'd sub nom.* Gonzales v. O Centro Espirita Beneficiente Uniao do Vegetal, 546 U.S. 418 (2006).

145. This assertion is based on a WestlawNext search conducted on February 12, 2014. I searched for "preliminary injunction" in all federal cases and then searched within the results for any of the following terms: "behavioral law and economics," "behavioral psychology," "behavioral economics," "loss aversion," "prospect theory," "omission bias," and "endowment effect."

146. Brief of Douglas Laycock as Amicus Curiae in Support of Respondents at 22–23, Gonzales v. O Centro Espirita Beneficiente Uniao do Vegetal, 546 U.S. 418 (2006) (No. 04-1084) [hereinafter Laycock Brief].

argument.[147] On the one hand, the appellants in *O Centro* could lay claim to the status quo being their long-time use of the banned substance *hoasca* during religious ceremonies (that being the "last . . . uncontested" state of affairs).[148] On the other, the government could say its subsequent banning of appellants' importation and use of *hoasca* under the Controlled Substances Act (CSA) was the actual status quo (that being the state of affairs in existence at the time of the court's decision).[149] Thus, as Judge Seymour wrote in his separate opinion, there are "two plausible status quos, each of them important."[150] Indeed, the appellants' use of *hoasca* was a longer-held entitlement and, as such, would reasonably be due greater weight when considering the problem of litigant loss aversion.[151] Ultimately, a majority of the en banc court concluded that the government's enforcement of the CSA was the status quo, causing the appellants to be subjected to a heightened burden.[152] Judge McConnell himself joined in this decision.[153]

This example illustrates that what one might define as the "status quo" in a litigation can at times be legitimately claimed by both sides in the same case. In such a situation, the loss-aversion-conscious court will have the basically impossible task of deciding which party is more subject to the endowment effect and loss aversion and, as a result, which decision will cause greater psychic harm (thus requiring greater judicial caution). In this scenario, Judge McConnell's BLE-based argument—which Professor Laycock aptly defines to be an inquiry into irreparable harm by proxy[154]—breaks down. Faced with a quandary like *O Centro*, a court is no more likely to get the balance of irreparable harms right by using the status quo as a proxy determinant of that balance than if it simply grappled with the question directly. Indeed, a majority of the *O Centro* court found the government's enforcement of the CSA to be the status quo when, as a matter of intuition, one could persuasively argue that the appellants' ceremonial use of *hoasca* was more potently endowed with value, with the loss of its use posing a greater risk of irreparable harm. Consideration of the status quo, on this basis, looks to be essentially valueless and unjustified.

147. *Id.*
148. *O Centro*, 389 F.3d at 1006–07 (Seymour, J., concurring in part and dissenting in part).
149. *Id.* at 1007.
150. *Id.*
151. *See* Laycock Brief, *supra* note 146.
152. *O Centro*, 389 F.3d at 980–81 (Murphy, J., concurring in part and dissenting in part).
153. *Id.* at 976.
154. Laycock Brief, *supra* note 146, at 22.

B. *Bounded Rationality and Judges*

As detailed above, experimental research and studies of judicial opinions have shown that judges may be subject to the same heuristics and biases as other decision makers.[155] Applying that research here, two biases, the status quo bias and the omission bias, seem to fit nicely with the circumstances surrounding a heightened burden for status quo-altering preliminary injunctions. These biases may be working to alter the thinking of federal judges who are deciding and reviewing motions for preliminary injunctions.

1. Status Quo Bias.—Status quo bias involves the alteration of decision making when one option is presented as the default, or status quo.[156] It has been demonstrated experimentally and by research into real life outcomes, as in the case of the Pennsylvania and New Jersey auto insurance reforms.[157]

A key factor leads me to conclude that status quo bias may have an operative effect on federal judges' consideration of motions for preliminary relief. Status quo bias has greater significance in situations where a decision maker's preferences on an issue are less strongly held.[158] For example, the average person surely has little reason to know whether it is better to have a cheaper auto insurance policy and the restricted right to sue or the opposite. The right to sue is abstract, and its necessity is uncertain and (presumably) temporally distant. Thus, one would expect that the status quo bias would have a substantial effect, and field research appears to bear out that it did in the auto insurance example.[159]

Now consider the trial judge. A trial judge hears many cases and presumably has no particular attachment to any one of them (indeed the judge must be impartial in his or her decision making).[160] In close cases where the facts and law do not allow the judge to have a strongly held preference as between plaintiff and defendant, it is reasonable to think that a judge might be affected by the status quo bias, just as happens to the average decision maker when faced with difficult decisions involving uncertain options and loosely held preferences. Indeed, Professors Guthrie and George's exploration of the "affirmance effect" in federal appellate courts provides anecdotal support that status quo bias plays a part in judicial thinking.[161]

---

155. *See supra* subpart III(D).
156. *See supra* note 101 and accompanying text.
157. *See supra* notes 103–12 and accompanying text.
158. *See supra* note 109 and accompanying text.
159. *See supra* notes 110–12 and accompanying text.
160. CODE OF CONDUCT FOR UNITED STATES JUDGES Canon 3 (2011).
161. *See supra* notes 134–38 and accompanying text.

If one is to grant that the status quo bias provides an explanation for why the status quo is part of preliminary injunction doctrine in some federal courts, the next question is whether that bias justifies its inclusion. The answer must surely be no. A substantial change in decision outcomes, if a decision maker were behaving rationally, should result from the preference for one option or another arising from those options' actual characteristics (outside of the decisional context). The fact that one option's designation as the status quo can cause an immediate and substantial increase in popularity with no change in its underlying quality is profoundly irrational. An example of this irrationality comes from experimental research, which has shown that the status quo bias has a particularly pronounced effect on those decision options that are most unpopular when otherwise not designated as the status quo.[162]

Bringing these threads together, I believe a heightened burden for status quo-altering preliminary injunctions arises from and ultimately exacerbates the status quo bias (by focusing the judge's attention on one decision option as the designated "status quo"). Thus, the bias dissociates the judge's decision from the underlying merits of the two parties' positions and gives one party an undue boost as a result of occupying the status quo position.

2. Omission Bias.—Omission bias manifests itself in the behavior of some decision makers when they evaluate commissions as more blameworthy than omissions, even when omissions and commissions produce the same outcome.[163] This bias can be justified in some instances as a useful rule of thumb for determining when actors possess bad intent and are therefore more likely to act wrongly again.[164]

There are many reasons to believe that omission bias is playing a role in the decision making of judges. Judge McConnell's opinion in *O Centro* argues that the reluctance to disturb the status quo is an expression of "judicial humility" and of a recognition that the "court bears more direct moral responsibility for harms that result from its intervention than from its nonintervention, and more direct responsibility when it intervenes to change the status quo than when it intervenes to preserve it."[165] This discussion of the increased moral culpability from intervening at the preliminary injunction stage is couched in the precise terms of the omission bias. Another Tenth Circuit opinion, this time by Judge Kelly, echoed the same sentiment, stating that a heightened burden had the purpose of

---

162. *See supra* note 109 and accompanying text.

163. *See supra* subpart III(C).

164. *See supra* notes 124–26 and accompanying text.

165. O Centro Espirita Beneficiente Uniao do Vegetal v. Ashcroft, 389 F.3d 973, 1015 (10th Cir. 2004) (McConnell, J., concurring), *aff'd sub nom.* Gonzales v. O Centro Espirita Beneficiente Uniao do Vegetal, 546 U.S. 418 (2006).

"minimiz[ing] any injury that would not have occurred but for the court's intervention."[166]

Aside from this anecdotal confirmation, one can look to the scholarship on the "affirmance effect" in the federal circuit courts of appeals for further evidence of omission bias at work.[167] While I discussed that material earlier in the context of the status quo bias, one could also see how actively overruling a lower court and being wrong in doing so might appear worse to an appellate judge than simply allowing a possibly flawed disposition to stand. Consider also the notion that affirmance is a "neutral" condition and that reversal is not. With affirmance being "neutral," only reversal would cause the judge's feelings about the disposition to correlate with the positive or negative nature of the outcome.[168]

Setting aside the issue of whether omission bias explains the phenomenon, it does not justify judges' caution about granting status quo-altering injunctions. My argument proceeds analogically from one advanced by Professors Vermeule and Sunstein. They argued in a recent article that the distinction between acts and omissions is not morally relevant with respect to governments.[169] The act–omission distinction breaks down in this context because "unlike individuals, governments always and necessarily face a choice between or among possible policies for regulating third parties."[170] Furthermore, "government is in the business of creating permissions and prohibitions. When it explicitly or implicitly authorizes private action, it is not omitting to do anything or refusing to act."[171]

This appraisal of the moral position of government necessarily applies to the individuals who carry out its mission as agents.[172] I now extend Sunstein and Vermeule's argument to judges faced with a motion for a preliminary injunction. A judge's decision to grant or deny such a motion will necessarily alter the balance of entitlements and hardships between the parties.[173] Furthermore, a judge's disposition of the motion will determine

166. RoDa Drilling Co. v. Siegal, 552 F.3d 1203, 1209 (10th Cir. 2009).

167. *See supra* notes 134–38 and accompanying text.

168. *See supra* notes 119–22 and accompanying text (summarizing the "norm" explanation of the omission bias).

169. Cass R. Sunstein & Adrian Vermeule, *Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs*, 58 STAN. L. REV. 703, 720–24 (2005).

170. *Id.* at 721.

171. *Id.*

172. *See id.* at 724 (arguing that natural persons form the staff of government but that, nonetheless, it is "irresponsible, indeed incoherent," for them to "invok[e] the natural person's liberty not to 'act'").

173. *Cf.* O Centro Espirita Beneficiente Uniao do Vegetal v. Ashcroft, 389 F.3d 973, 1015–16 (10th Cir. 2004) (McConnell, J., concurring), *aff'd sub nom.* Gonzales v. O Centro Espirita Beneficiente Uniao do Vegetal, 546 U.S. 418 (2006) ("Disrupting the status quo [by granting or

how weakly or strongly similar conduct by other individuals is deterred.[174] Thus, whatever moral salience the act–omission distinction has for the average person, it is absent in the judicial context. In light of this absence, the primary justification for the omission bias's existence evaporates, and one is left with irrational behavior based on a heuristic. Ultimately, omission bias provides little justification for the courts' heightened burden when evaluating status quo-altering injunctions.

C. Better to Reject Consideration of the Status Quo

The better course is to take the path endorsed by the Sixth Circuit and Judge Posner and simply ignore the status quo.

Wright and Miller have noted that invocation of the status quo is "unobjectionable when used simply to articulate the desire to prevent defendant from changing the existing situation to plaintiff's irreparable detriment."[175] Indeed, the historical roots of the status quo in preliminary injunction decisions go back to the English Court of Chancery in the nineteenth century, where the chancellor brought up the status quo not as part of any independent test, but simply as a means to "describe the usual effect of preliminary injunctive relief" (i.e., the avoidance of irreparable injury to the movant).[176] This primarily descriptive use of the status quo phrase appears to have been the prevailing practice among nineteenth century American courts as well.[177]

Wright and Miller have also argued that, in courts employing a status quo test, the test can serve as "a harmless makeweight" when sufficient independent grounds exist for denying the injunction according to the traditional irreparable harm analysis.[178] It is not always so harmless however. Professor Laycock has noted the mischief that the status quo can do when it is given independent doctrinal force. First, he points out that in situations where irreparable harm will be caused by the continuation of the status quo, the traditional irreparable injury analysis is at odds with the status quo test, and the status quo burden must be overcome in order to do justice.[179] In these cases, the "status quo test is clearly an obstacle to justice

---

denying a motion for an injunction] may provide a benefit to one party, but only by depriving the other party of some right he previously enjoyed.").

174. *Cf.* Freedman v. Maryland, 380 U.S. 51, 59 (1965) (limiting the duration of prior restraint in advance of final judicial review under an obscenity law to the "preservation of the status quo for the shortest fixed period compatible with sound judicial resolution" in part to "minimize the deterrent effect of an interim and possibly erroneous denial" of the right to screen the film in question).

175. WRIGHT ET AL., *supra* note 10, at 125.

176. *See* Lee, *supra* note 4, at 132–33.

177. *Id.* at 138.

178. WRIGHT ET AL., *supra* note 10, at 127.

179. Laycock Brief, *supra* note 146, at 20.

rather than an aid."[180] Laycock notes a second key problem: the status quo is "highly manipulable," and thus, giving it doctrinal force leads to substantial argument between parties attempting to define this slippery concept.[181] Arguing about the status quo therefore has major downsides from the perspective of judicial efficiency. Parties are required to bear the expense of arguing about the status quo in addition to the cost associated with making their case according to the traditional four factors.[182] This outcome shows that the status quo focus is having a perverse effect, as one of its justifications is that it serves as a rule of thumb that can reduce litigation costs by easing the decision process for parties and judges.[183]

V. Conclusion

The test for whether to grant preliminary relief in the federal courts need not, as a conceptual matter, incorporate the status quo. Numerous academics and judges have rejected its application as unnecessary at best and pernicious at worst. Behavioral law and economics provides a potential means to explain why some courts appear wedded to a heightened burden in cases of status quo-altering preliminary injunctions. In particular, the status quo bias and the omission bias help to explain why the doctrine in these courts has an undue preference for choosing an option framed as the status quo and for avoiding judicial commissions in situations of uncertainty.

The possible BLE-based justifications for this preference are unsuccessful. First, litigant loss aversion should not play a particularly large role in judges' thinking because the status quo as interpreted by the courts is often unclear and potentially involves a return to some state of affairs that has already since passed. Furthermore, the party that the court ultimately deems to be in the status quo position as a legal matter may not be the one with the more powerfully "endowed" entitlement, thus leading to a potential subversion of the very litigant-focused rationale that was supposed to justify the test in the first place.

Second, status quo bias provides no justification on its face, as that bias operates to change decision making purely based on an artifact of the decision process, not on anything about the choices themselves. Finally, to the extent that the omission bias has any basis as a useful rule of thumb, it is not appropriate to apply it to the actions of judges. When a judge elects not to intervene, he or she is making a decision that necessarily has an impact upon the parties before the court. Judicial "omissions" are not even omissions in the truest sense, and they certainly do not bear the same moral characteristics as omissions by ordinary people.

---

180. *Id.*

181. *Id.* at 20–22.

182. Lee, *supra* note 4, at 163–64.

183. *Id.* at 164 n.296.

Going forward, I believe two courses of action are called for. First, those federal courts that continue to promulgate a heightened burden for status quo-altering preliminary injunctions should cease doing so and move toward the status quo-agnostic view expounded by most academics and many courts. Second, researchers may find it fruitful to put the BLE theories I have applied to this context into experimental form, possibly by surveying federal district judges in such a way as to elicit their views about mandatory and status quo-altering injunctions. Only through such experimental research can it be more rigorously demonstrated that bounded rationality is infecting this particular area of the law. If such a demonstration were made, it would further strengthen the case that this particular facet of the law—a barnacle attached to the law of remedies—should be scrubbed away.

—James Powers

* * *

JAMAIL CENTER FOR LEGAL RESEARCH
TARLTON LAW LIBRARY
THE UNIVERSITY OF TEXAS SCHOOL OF LAW

The Tarlton Law Library Oral History Series features interviews with outstanding alumni and faculty of The University of Texas School of Law.

Oral History Series

No. 1 - *Joseph D. Jamail, Jr.* 2005. $20
No. 2 - *Harry M. Reasoner* 2005. $20
No. 3 - *Robert O. Dawson* 2006. $20
No. 4 - *J. Leon Lebowitz* 2006. $20
No. 5 - *Hans W. Baade* 2006. $20
No. 6 - *James DeAnda* 2006. $20
No. 7 - *Russell J. Weintraub* 2007. $20
No. 8 - *Oscar H. Mauzy* 2007. $20
No. 9 - *Roy M. Mersky* 2008. $25

Forthcoming:

Gloria Bradford, Patrick Hazel, James W. McCartney, Michael Sharlot, Ernest E. Smith, John F. Sutton, Jr.

Other Oral Histories Published by the Jamail Center for Legal Research

Robert W. Calvert (Texas Supreme Court Trilogy, Vol. 1). 1998. $20

Joe R. Greenhill, Sr. (Texas Supreme Court Trilogy, Vol. 2). 1998. $20

Gus M. Hodges (Tarlton Law Library Legal History Series, No. 3). 2002. $20

Corwin Johnson (Tarlton Law Library Legal History Series, No. 4). 2003. $20

W. Page Keeton (Tarlton Legal Bibliography Series, No. 36). 1992. $25

Jack Pope (Texas Supreme Court Trilogy, Vol. 3). 1998. $20

Order online at http://tarlton.law.utexas.edu/ click on Publications
or contact Publications Coordinator,
Tarlton Law Library, UT School of Law,
727 E. Dean Keeton St., Austin, TX 78705

phone (512) 471-6228; *fax* (512) 471-0243;
email tarltonbooks@law.utexas.edu

* * *

* * *

* * *


TEXAS
TLR
LAW REVIEW