Texas Law Review

FEBRUARY 2014 VOL. 92 NO. 3 PAGES 517 TO 772

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 3 February 2014

* * *

Texas Law Review

Volume 92, Number 3, February 2014

ARTICLES

BOOK REVIEWS

NOTES

* * *

Articles

Purging Patent Law of "Private Law" Remedies

Ted Sichelman*

This Article rejects the fundamental "private law" premise of patent law remedies that courts should always attempt to make the patentee "whole" in the event of infringement because the overarching aim of patent law is to promote innovation, not to remedy private wrongs. Specifically, make-whole damages may thwart optimal innovation incentives when they concern small components of complex products involving high-switching costs, generate large consumer deadweight losses, result in substantial duplicated costs during the pre-invention R&D process, or create transaction costs far in excess of the value of the invention. In other situations, a patentee should be made more than whole. For example, inducing socially valuable innovations that do not command large profits in the private market—such as drugs for rare diseases and technologies for the disabled—may require more than make-whole compensation.

More generally, the statutory remedies provisions of the Patent Act rest on a flawed foundation. Instead of correcting for private wrongs inflicted on private parties, patent law remedies should be tailored simply to promote the types and levels of innovation that most benefit society, taking into account administrative and error costs. As such, the patent system and its associated remedies should be viewed as part of a public regulatory regime designed to further societal goals rather than a private law system that protects individual interests.

---

* Professor, University of San Diego School of Law. I thank Michael Abramowicz, Olufunmilayo Arewa, Kenneth Ayotte, Scott Baker, Avi Bell, Bob Brauneis, Dan Burk, T.J. Chiang, Kevin Collins, Chris Cotropia, Tom Cotter, Shari Diamond, Peter DiCola, Gerrit De Geest, John Golden, Richard Gruner, Christi Guerrini, Jonathan Handel, Paul Heald, Scott Kieff, Josh Kleinfeld, Amy Landers, Peter Lee, Mark Lemley, Lydia Loren, Ed Manzo, Mark McKenna, John McGinnis, Neil Netanel, Michael Risch, Betsy Rosenblatt, Josh Sarnoff, Henry Smith, James Speta, Sam Vermont, John Whealan, and participants at the 2011 Samsung-Stanford Conference on Patent Remedies, the 2011 IP Scholars Conference, the "IP in the Trees" workshop at Lewis & Clark School of Law, the 2012 UCLA Media, Entertainment, and IP Workshop, the 2012 University of San Diego Conference on Patent Law Remedies, and faculty colloquia and workshops at the DePaul University College of Law, George Washington University School of Law, Northwestern University School of Law, and Washington University in St. Louis School of Law for their helpful comments and discussions. I also thank Gunjan Agarwal, Alan Chang, April Sun, and Dan Tasakalos for their invaluable research assistance.

Introduction

Private law—such as tort, property, and contract law—generally provides remedies for the infliction of wrongs on private individuals and entities.[1] For example, if a rambunctious bar patron punches you on the nose while you quietly sip a martini, you can sue the patron in tort for battery, collecting damages via a "liability" rule for at least the harm inflicted and—in many cases—an injunction via a "property" rule going forward to prevent future harm.[2] Tort law would provide you these sorts of remedies because we as a society believe the optimal state of the world is for you to go on quietly sipping your drink, free from the interference and physical damage caused by the rowdy patron. In other words, tort law—at least, traditionally—tends to assume that a world free from interference into

1. *See* ERNEST J. WEINRIB, THE IDEA OF PRIVATE LAW 143 (1995) ("When the remedy takes the form of an award of damages, a single amount undoes the injustice both of what the defendant has done and of what the plaintiff has suffered."); Ezra Ripley Thayer, *Public Wrong and Private Action*, 27 HARV. L. REV. 317, 326–27 (1914) (discussing private civil rights of action in tort).

2. *See* DAN B. DOBBS, LAW OF REMEDIES § 8.10, at 692 (2d ed. 1993) (observing how injunctions against ongoing risky practices may accompany tort damages in personal injury cases); Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092 (1972) (proposing the property–liability rule distinction); *see also* 5 WILLIAM BLACKSTONE, COMMENTARIES 7 (arguing that private law remedies redress violations of private rights "by either restoring to [the victim] his right" or providing remuneration sufficient to compensate for the violation).

an individual's or private entity's "sphere of autonomy" is the ideal state of the world.[3] As such, it generally seeks to return the private actor that has been harmed to the *status quo ante*, through damages, an injunction, or both.[4]

Patent law, on the other hand, is not designed to remedy private wrongs. Rather, its major aim is to promote innovation.[5] Nonetheless, patent remedies mirror traditional tort law remedies by attempting to restore the patentee to the *status quo ante*—namely, the state of the world in which there is no infringement of the patent.[6] No infringement occurs when third parties, including competitors, do not practice the patent or pay license fees to permissibly perform otherwise infringing acts.[7] Historically, patent law

---

3. *See* Donald P. Judges, *Of Rocks and Hard Places: The Value of Risk Choice*, 42 EMORY L.J. 1, 63 (1993) ("Tort law generally seeks to protect the autonomy of plaintiffs by compensating for and deterring the forced intrusions of defendants' tortious conduct."). Of course, there are those exceptional cases for which intentional intrusions into a person's sphere of autonomy are considered excused or justified, such as self-defense, duress, and the like. *See, e.g.*, RESTATEMENT (SECOND) OF TORTS § 65 (1965) (setting forth factors for establishing self-defense).

4. *See* Aaron Xavier Fellmeth, *Civil and Criminal Sanctions in the Constitution and Courts*, 94 GEO. L.J. 1, 60 (2005) ("In tort law, injunctions are granted . . . when the remedy of compensatory damages will not suffice to restore the status quo ante."); Karen E. Sandrik, *Reframing Patent Remedies*, 67 U. MIAMI L. REV. 95, 102 (2012) ("Further, patent and tort remedies largely mirror one another in that the goal of both remedial structures is to restore the aggrieved party to the status quo ante." (citing an earlier version of this Article)); *see also* RESTATEMENT (SECOND) OF TORTS § 936 (1979) (setting forth factors for when injunctions are appropriate).

5. *See* U.S. CONST. art. I, § 8, cl. 8 ("To *promote the Progress* of Science and useful Arts . . . ." (emphasis added)); Sears, Roebuck & Co. v. Stiffel Co., 376 U.S. 225, 229–31 (1964) ("Patents are not given as favors, as was the case of monopolies given by the Tudor monarchs, . . . but are meant to encourage invention by rewarding the inventor with the right . . . to exclude others from the use of his invention."); Peter S. Menell, *Intellectual Property: General Theories*, *in* 2 ENCYCLOPEDIA OF LAW AND ECONOMICS 129, 130–33 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000) ("The United States Constitution expressly conditions the grant of power to Congress to create patent and copyright laws upon a utilitarian foundation . . . ."). By using the term *innovation*, I intend to include not only invention but the commercialization and dissemination of the invention. *See, e.g.*, Jan Fagerberg, *Innovation: A Guide to the Literature*, *in* THE OXFORD HANDBOOK OF INNOVATION 1, 4 (Jan Fagerberg et al. eds., 2005) ("Invention is the first occurrence of an idea for a new product or process, while innovation is the first attempt to carry it out into practice.").

6. Mark A. Lemley, *Distinguishing Lost Profits From Reasonable Royalties*, 51 WM. & MARY L. REV. 655, 674 (2009) ("Patent damages are supposed to compensate patent owners for their losses, putting them back in the world they would have inhabited but for infringement."); *see also* Thomas F. Cotter, *Four Principles for Calculating Reasonable Royalties in Patent Infringement Litigation*, 27 SANTA CLARA COMPUTER & HIGH TECH. L.J. 725, 727 (2011) ("[T]he baseline damages recovery for prevailing patent owners should be the amount that restores them to the position they would have enjoyed but for the infringement.").

7. *See* 35 U.S.C. § 271(a) (2006) ("[W]hoever without authority makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention during the term of the patent therefor, infringes the patent."). There is also a fairly limited "experimental use" exception to patent infringement, which excuses otherwise

typically provided for damages for past infringement and injunctions on a going-forward basis to restore the *status quo ante*.[8] Backward-looking damages compensate the patentholder either for lost profits caused by the third party's infringing activity or for forgone royalties in the event the patentholder did not practice its patent (or, in some situations, where royalties would result in greater damages than lost profits).[9] Forward-looking injunctions restore the patentholder to the equitable *status quo ante* at the time of patent issuance, providing the holder an absolute right—backed by contempt sanctions—to prevent third parties from infringing the patent.[10]

In *eBay v. MercExchange*,[11] the Supreme Court shifted the calculus by holding that injunctions are not mandatory and should instead be awarded on the basis of a set of equitable factors.[12] In an influential concurrence,[13]

---

infringing acts. *See* 5 DONALD S. CHISUM, CHISUM ON PATENTS § 16.03[1] (2012) (collecting cases addressing the experimental use doctrine).

8. *See, e.g.*, Richardson v. Suzuki Motor Co., 868 F.2d 1226, 1247 (Fed. Cir. 1989) (reaffirming the long-standing rule of issuing injunctions to successful patentees); Connell v. Sears, Roebuck & Co., 722 F.2d 1542, 1548 (Fed. Cir. 1983) ("[A] patent is a form of property right, and the right to exclude recognized in a patent is but the essence of the concept of property."); *see also* Smith Int'l, Inc. v. Hughes Tool Co., 718 F.2d 1573, 1577–78 (Fed. Cir. 1983) (reasoning that the incentive to engage in research would be diminished without the right to injunctive relief); Ben Depoorter, *Property Rules, Liability Rules and Patent Market Failure*, 4 ERASMUS L. REV. 59, 61 (2008) ("[T]he equitable remedy of injunction has dominated the law of intellectual property.").

9. *See* 35 U.S.C. § 284 (2006 & Supp. V 2012) (providing for "no . . . less than a reasonable royalty for the use made of the invention by the infringer"); Lemley, *supra* note 6, at 655 ("Patent damages are designed to compensate patentees for their losses, not punish accused infringers or require them to disgorge their profits."); *id.* at 657 ("Giving patentees the profits they would have made absent the infringement effectively puts them in the same position as if they had had an injunction in place all along."); *see also* ROGER D. BLAIR & THOMAS F. COTTER, INTELLECTUAL PROPERTY: ECONOMIC AND LEGAL DIMENSIONS OF RIGHTS AND REMEDIES 12 (2005) (noting that damages for an infringement "may include an award of the plaintiff's lost profits attributable to the infringement; the amount of an established royalty; or a reasonable royalty").

10. *See* Osborn v. Bank of U.S., 22 U.S. (9 Wheat.) 738, 749 (1824) (stating that patent owners may obtain injunctions to prevent others from using the patent based on "the principle[] that the injury was consequential, not direct, and that it would be difficult, if not impossible, to estimate the damages"); Rite Hite Corp. v. Kelley Co., 56 F.3d 1538, 1562 (Fed. Cir. 1995) ("An injunction preserves the patentee's exclusive right to market embodiments of the patented invention.").

11. 547 U.S. 388 (2006).

12. *See id.* at 391–92 (holding that the well-established principles of equitable relief "apply with equal force to disputes arising under the Patent Act"); Carl Shapiro, *Injunctions, Hold-Up, and Patent Royalties*, 12 AM. L. & ECON. REV. 280, 282 (2010) ("The Supreme Court ruled unanimously that the district court has discretion whether to grant or deny injunctive relief based on traditional principles of equity, using a four-factor test.").

13. *See eBay*, 547 U.S. at 395–97 (Kennedy, J., concurring). A large number of opinions have specifically cited Justice Kennedy's concurring opinion with approval. *See, e.g.*, Robert Bosch LLC v. Pylon Mfg. Corp., 659 F.3d 1142, 1150 (Fed. Cir. 2011); Salinger v. Colting, 607 F.3d 68, 82 (2d Cir. 2010); N. Am. Med. Corp. v. Axiom Worldwide, Inc., 522 F.3d 1211, 1228 (11th Cir. 2008); ATCS Int'l LLC v. Jefferson Contracting Corp., 807 F. Supp. 2d 516, 519 (E.D. Va. 2011); i4i Ltd. P'ship v. Microsoft Corp., 670 F. Supp. 2d 568, 600 (E.D. Tex. 2009); Hynix

Justice Kennedy argued that entities that do not practice their patents (so-called nonpracticing entities or "NPEs")—typically, by forgoing manufacturing and product sales in favor of licensing—generally should not be awarded an injunction because it would give them "undue leverage" over third parties, particularly when the patent covers a "small component" of a complex product.[14] Implicit in Justice Kennedy's reasoning is the tort law *status quo ante* rationale: because an NPE would by definition license its patents, a liability rule providing damages on a forward-looking basis—assuming it adequately reflected the market royalty rate—would return the NPE exactly to that state of the world that would have existed but for the infringement.[15] Conversely, Justice Kennedy assumed that a practicing entity would not generally license its patents, instead choosing to leverage its patents by earning supernormal profits from selling products and

---

Semiconductor Inc. v. Rambus Inc., 609 F. Supp. 2d 951, 966 (N.D. Cal. 2009); Amgen, Inc. v. F. Hoffman-La Roche Ltd., 581 F. Supp. 2d 160, 211 (D. Mass. 2008); MGM Studios, Inc. v. Grokster, Ltd., 518 F. Supp. 2d 1197, 1215–16 (C.D. Cal. 2007); Commonwealth Sci. and Indus. Research Org. v. Buffalo Tech. Inc., 492 F. Supp. 2d 600, 605 (E.D. Tex. 2007); MPT, Inc. v. Marathon Labels, Inc., 505 F. Supp. 2d 401, 419–20 (N.D. Ohio 2007); z4 Techs., Inc. v. Microsoft Corp., 434 F. Supp. 2d 437, 441 (E.D. Tex. 2006). Additionally, a well-known empirical study indicates that Justice Kennedy's opinion has been far more influential than Chief Justice Roberts's opinion, which suggested following the historical practice of typically awarding injunctive relief. *See eBay*, 547 U.S. at 394–95 (Roberts, C.J., concurring) ("From at least the early 19th century, courts have granted injunctive relief upon a finding of infringement in the vast majority of patent cases. . . . When it comes to discerning and applying those standards, in this area as others, a page of history is worth a volume of logic." (internal quotation marks omitted)); FTC, THE EVOLVING IP MARKETPLACE: ALIGNING PATENT NOTICE AND REMEDIES WITH COMPETITION app. b at 256–59 (2001), *available at* http://www.ftc.gov/sites/default/files/documents/reports/evolving-ip-marketplace-aligning-patent-notice-and-remedies-competition-report-federal-trade/110307patentreport.pdf (presenting the results of Steven Malin's survey of post-*eBay* cases, which found that district courts have denied injunctions to nonpracticing entities more than 50% of the time).

14. *See eBay*, 547 U.S. at 396–97 (Kennedy, J., concurring) ("When the patented invention is but a small component of the product . . . and the threat of an injunction is employed simply for undue leverage in negotiations, legal damages may well be sufficient to compensate for the infringement and an injunction may not serve the public interest.").

15. *See id.* at 396 ("For [NPEs], an injunction, and the potentially serious sanctions arising from its violation, can be employed as a bargaining tool to charge exorbitant fees to companies that seek to buy licenses to practice the patent."). Presumably, Justice Kennedy viewed the fees generated from an injunction as "exorbitant," *see id.*, for NPEs, but not for practicing entities, because—by widely licensing instead of practicing (or exclusively licensing) their patents—NPEs do not seek to exclude others from practicing their patents in the marketplace, nor do NPEs market products potentially subject to an accused infringer's patents. *See* FTC, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY, ch. 3, at 38–39 (2003), *available at* http://www.ftc.gov/os/2003/10/innovationrpt.pdf ("The potential for hold-up to result in mutually assured destruction means firms actively participating in the industry—patent practicing entities (PPEs)—are unlikely to employ this hold-up strategy against each other. . . . [H]owever, identified firms referred to as non-practicing entities (NPEs) . . . can successfully employ a hold-up strategy without fear of retaliation.") *cited in eBay*, 547 U.S. at 396 (Kennedy, J., concurring). As such, any leverage gained by NPEs from the threat of curtailing such activity via an injunction as litigation remedy is "undue" on this view. *eBay*, 547 U.S. at 396–97 (Kennedy, J., concurring).

services in the marketplace.[16] Importantly, both the historical doctrine and slightly modified approach of *eBay*[17] assume that the *status quo ante* endgame, to the extent it can be costlessly and accurately implemented by a court, is in fact the ideal remedy.[18] In other words, the courts—and essentially the entire body of academic literature—have assumed that treating the patentee like a private right holder entitled to traditional private law remedies optimally promotes innovation.[19]

I argue in this Article, however, that the traditional view is wrong in at least three contexts. First, along the lines of Justice Kennedy's suggestion, when a patent covers a minor component of a complex product that a third party has already implemented—and there are large switching costs in implementing a substitute for the patented component—then providing any

16. *See eBay*, 547 U.S. at 396 (Kennedy, J., concurring) ("An industry has developed in which firms use patents not as a basis for producing and selling goods but, instead, primarily for obtaining licensing fees.").

17. Justice Thomas's and Chief Justice Roberts's opinions in *eBay* make no explicit distinction between practicing entities and NPEs. *See id.* at 390–95 (majority opinion and Roberts, J., concurring). Following *eBay*, however, lower courts have followed Justice Kennedy's distinction—routinely granting injunctions to practicing entities but denying them to NPEs. *See* FTC, *supra* note 13 (presenting findings that, in cases in which practicing and nonpracticing entities were distinguished, district courts granted injunctions to practicing entities about 85% of the time but only about 45% of the time to NPEs following *eBay*).

18. *See* Aro Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 507 (1964) (holding that in determining damages, the question is "had the Infringer not infringed, what would Patent Holder-Licensee have made?" (internal quotation marks omitted)); Pall Corp. v. Micron Separations, Inc., 66 F.3d 1211, 1223 (Fed. Cir. 1995) ("[T]he purpose of compensatory damages is not to punish the infringer, but to make the patentee whole."); *see also* Amy L. Landers, *Patent Valuation Theory and the Economics of Improvement*, 88 TEXAS L. REV. SEE ALSO 163, 166 (2009) ("[P]atent damages are a make-whole remedy, intended to restore the patentee to the same position as before infringement.").

19. *See* General Motors Corp. v. Devex Corp., 461 U.S. 648, 653–55 (1983) ("When Congress wished to limit an element of recovery in a patent infringement action, it said so explicitly. . . . Congress' overriding purpose [is] affording patent owners complete compensation."); 6 JOHN GLADSTONE MILLS III ET AL., PATENT LAW FUNDAMENTALS § 20:65 (2d ed. 2013) ("[I]t is only by means of injunctive relief that a patentee can realize 'the right to exclude others'. . . ."); Roger D. Blair & Thomas F. Cotter, *Rethinking Patent Damages*, 10 TEX. INTELL. PROP. L.J. 1, 4 (2001) ("[W]e recommend the application of traditional tort-law doctrines of cause-in-fact and proximate cause to patent questions, and generally reject the idea that patent infringement is materially different . . . ."); Frank H. Easterbrook, *Intellectual Property Is Still Property*, 13 HARV. J.L. & PUB. POL'Y 108, 109 (1990) ("Patents give a right to exclude, just as the law of trespass does with real property."); Paul J. Heald, *Optimal Remedies for Patent Infringement: A Transactional Model*, 45 HOUS. L. REV. 1165, 1171 (2008) ("In establishing what constitutes infringement and what remedies apply, patent law's secondary function looks like tort law."); Amy L. Landers, *Liquid Patents*, 84 DENV. U. L. REV. 199, 252–53 (2006) (contending that patent remedies should differ as between practicing and nonpracticing entities because each incurs different kinds of "harm"); Mark A. Lemley & Carl Shapiro, *Patent Holdup and Royalty Stacking*, 85 TEXAS L. REV. 1991, 2036 (2007) (arguing that retaining injunctions for practicing entities "is justified in part for reasons of equity"); *see also* Megan M. La Belle, *Patent Law as Public Law*, 20 GEO. MASON L. REV. 41, 41 (2012) ("Patent litigation historically has been regarded as private law litigation, meaning disputes between private parties about private rights." (internal quotation marks omitted)).

patentee, *practicing or not*, with an injunction on a going-forward basis will generally yield market rewards (or settlement payments) far in excess of the value of the innovative component to society.[20] This windfall to the patentee results in the patent system providing too great an incentive for component and incremental innovations relative to discrete, whole product innovations.[21] Indeed, Mark Lemley and Carl Shapiro have convincingly argued as much in the context of nonpracticing entities.[22] Yet, because Lemley, Shapiro, and Justice Kennedy fall prey to the assumption that patent remedies should mirror traditional private law remedies, they mistakenly conclude that injunctions should still be generally available to practicing entities.[23] Rather, as I show below, their arguments against issuing injunctions to NPEs are often just as applicable to practicing entities.[24]

Second, in many contexts, the patent system provides excessive incentives to generate needed R&D and commercialization activity.[25] For example, the costs of invention and commercialization in the software industry are far below those in the pharmaceutical industry.[26] Yet, the

---

20. *See infra* subpart II(B).

21. For instance, such differential incentives may unduly increase incentives to innovate—or at least seek patent protection for plausible innovations—in the software industry relative to the pharmaceutical industry. *See* Julie E. Cohen & Mark A. Lemley, *Patent Scope and Innovation in the Software Industry*, 89 CALIF. L. REV. 1, 26 n.94 (2001) (noting that a patented invention typically covers a small part of a software product as compared with an entire pharmaceutical product). Relatedly, it may provide an inefficient advantage to large firms, which tend to incrementally innovate, relative to small firms, which tend to radically innovate. *See* Jonathan M. Barnett, *Is Intellectual Property Trivial?*, 157 U. PA. L. REV. 1691, 1736–37 (2009) ("[L]arge firms tend to undertake low-risk, incremental innovation projects that preserve market share while small firms tend to undertake high-risk, radical innovation projects that seek to capture market share.").

22. *See generally* Lemley & Shapiro, *supra* note 19 (discussing some of the problems associated with providing injunctive relief to patentees). As an important point of clarification, Lemley would effectively treat an NPE that exclusively licenses a patent as a practicing entity, since the NPE essentially stands in the shoes of the sole practicing entity from a market perspective. Lemley, *supra* note 6, at 673.

23. *See infra* subparts II(B)–(C).

24. *See infra* subparts II(B)–(C).

25. *See* Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEXAS L. REV. 1031, 1058–65 (2005) (positing that providing patents with the full social value of an invention would often result in overcompensation); *see also* Vincenzo Denicolò, *Do Patents Over-Compensate Innovators?*, 22 ECON. POL'Y 679, 713 (2007) (proposing a model to determine whether patent owners are overcompensated by comparing profit ratios to the elasticity of the supply of inventions); *cf.* Peter S. Menell, *The Challenges of Reforming Intellectual Property Protection for Computer Software*, 94 COLUM. L. REV. 2644, 2646 (1994) ("Excessive protection for first generation innovation can impede later stages, thereby undermining some of the salutary effects of strong intellectual property protection.").

26. *See* Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1581–82 (2003) (noting that R&D costs for software are considerably lower than those for pharmaceuticals); Gregory N. Mandel, *Will America Reinvent Itself? Patent Reform in 2011*, BUS. L. TODAY, Aug. 2011, at 1, 2 ("Developing a new drug or biologic routinely takes a decade or

duration of software and pharmaceutical patents are exactly the same (indeed, in practice, software patents last longer),[27] and—at least in rough conceptual terms—the scope of software patents often exceeds the scope of pharmaceutical patents.[28] If the broad scope and long duration afforded software patents is unnecessary to incentivize innovation in that industry, then the rewards provided by the patent system are excessive.[29] This result can create windfalls for innovators, which in turn can foster needless consumer deadweight losses, particularly when (1) the patentholder enjoys the ability to price its patented goods over the competitive price,[30] or

---

more, costs hundreds of millions or billions of dollars, and often requires testing hundreds of alternatives or compounds. . . . New software applications can be produced on much shorter time scales and for a much more limited investment . . . ."); *see also* PETER TOLLMAN ET AL., THE BOST. CONSULTING GRP., A REVOLUTION IN R&D: HOW GENOMICS AND GENETICS ARE TRANSFORMING THE BIOPHARMACEUTICAL INDUSTRY 12 (2001) (estimating that the cost to discover, develop, and commercialize each patented drug is about $880 million); Shanling Li et al., *Why Do Software Firms Fail? Capabilities, Competitive Actions, and Firm Survival in the Software Industry from 1995 to 2007*, 21 INFO. SYS. RES. 631, 642, 643 tbl.1 (2010) (surveying 870 publicly owned software companies and finding that average R&D expenditures per company were approximately $27 million annually).

27. *See* Emily Michiko Morris, *The Myth of Generic Pharmaceutical Competition Under the Hatch-Waxman Act*, 22 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 245, 266–67 (2012) (noting that the effective patent term for pharmaceuticals, even with patent term restoration under the Hatch-Waxman Act, is fourteen years while the effective term for nonpharmaceutical patents is eighteen-and-a-half years); *see also* Daniel R. Cahoy, *An Incrementalist Approach to Patent Reform Policy*, 9 N.Y.U. J. LEGIS. & PUB. POL'Y 587, 648 (2006) (arguing that the effective patent term for software patents is too long and for pharmaceutical patents is too short).

28. *See* Emily Michiko Morris, Res *or Rules? Patents and the (Uncertain) Rules of the Game*, 18 MICH. TELECOMM. & TECH. L. REV. 481, 498–99 (2012) (stating that the scope of software patents is defined more by concepts than by physical or functional structures, unlike chemical patents, whose scope corresponds only to a limited number of chemical structures); *see also* Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 843 (1990) (explaining that effective patent scope depends on the type of technology at issue).

29. *See* Burk & Lemley, *supra* note 26, at 1687–88 ("While most biotechnological and chemical inventions require broad patent protection because of their high cost and uncertain development process, the opposite is true in the case of software development. Software inventions tend to have a quick, cheap, and fairly straightforward post-invention development cycle."); Linda R. Cohen & Roger G. Noll, *Intellectual Property, Antitrust and the New Economy*, 62 U. PITT. L. REV. 453, 469 (2001) (arguing that software patents provide excessive protection); Richard R. Nelson, *Intellectual Property Protection for Cumulative Systems Technology*, 94 COLUM. L. REV. 2674, 2674 (1994) (arguing for a moderate protection scheme to meet the protective needs of the software industry).

30. WILLIAM D. NORDHAUS, INVENTION, GROWTH, AND WELFARE: A THEORETICAL TREATMENT OF TECHNOLOGICAL CHANGE 76 (1969) (describing a model for determining an optimal patent term by balancing increased incentives for innovation against greater deadweight losses); Jonathan M. Barnett, *Private Protection of Patentable Goods*, 25 CARDOZO L. REV. 1251, 1269 (2004) (noting that a patent's social costs include "supracompetitive pricing power exerted by the patent holder (or, more specifically, the deadweight loss resulting from the patent holder's output restrictions)").

(2) when multiple parties needlessly duplicate R&D in "racing" for an excessive patent prize.[31]

Third, when reasonable minds differ over whether a given patent is infringed, valid, or enforceable, it may be economically efficient for third parties to forgo large transaction costs in negotiating a license, choosing instead to infringe.[32] Like the theory of efficient breach in contract law,[33] I argue below that efficient infringement can occur when the transaction costs of negotiation dwarf the value of the innovation at issue, which can result when there is large uncertainty in the underlying patent rights or simply when the economic value of the innovation is fairly minimal.[34] In these situations, it may be optimal to preclude injunctions in order to foster efficient infringement.[35]

On the other hand, I assert that while full compensatory damages may be excessive in the three situations described above, there are at least two other situations in which the patentee may need *more than* compensatory damages to generate optimal innovation incentives.[36] First, if detection of infringement is difficult, damages should be enhanced to compensate for undetected infringement—activity for which the accused infringer reaps profit but pays no reward to the patentee.[37] This condition should apply

---

31. *See* SUZANNE SCOTCHMER, INNOVATIONS AND INCENTIVES 112–13 (2004) (discussing how broad intellectual property rights can incite "races" for patents, resulting in duplicated costs from different inventors expending time and money to achieve the same goal inefficiently); Partha Dasgupta & Joseph Stiglitz, *Uncertainty, Industrial Structure, and the Speed of R&D*, 11 BELL J. ECON. 1, 11–14 (1980) (describing how parties engage in duplicative R&D in a monopolistic controlled market in a race to attain the monopoly pricing prize); Mark F. Grady & Jay I. Alexander, *Patent Law and Rent Dissipation*, 78 VA. L. REV. 305, 308 (1992) ("The defect in the system is that if multiple inventors expend resources in competition for the patent monopoly, the benefit to society of having the invention will be dissipated by the cost of numerous, redundant, development efforts."); *see also* Kenneth W. Dam, *The Economic Underpinnings of Patent Law*, 23 J. LEGAL STUD. 247, 252 n.14 (1994) ("A further aspect of rent seeking beyond unnecessary duplication of R&D expenditures is that the race for the patent will cause R&D expenditures to be made at a faster than optimal rate.").

32. *See infra* section II(C)(1).

33. *See* Ian R. Macneil, *Efficient Breach of Contract: Circles in the Sky*, 68 VA. L. REV. 947, 950–53 (1982) (positing that "efficient breach" is efficient when the transaction costs of renegotiating the contract outweigh the transaction costs from breach).

34. *See infra* section II(C)(1).

35. *Cf.* Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341, 345–47 (2010) (proposing a new "commercialization" patent that would grant an affirmative equitable right to make and sell a product that could infringe a traditional "invention" patent).

36. *See infra* section II(C)(2).

37. *See* Michael Abramowicz, *A Unified Economic Theory of Noninfringement Opinions*, 14 FED. CIR. B.J. 241, 254 (2004) ("If infringers pay full damages, their conduct will be optimized, so if patentees will enforce their rights only some of the time, enhanced damages are appropriate, with the enhanced damages multiplier equal to the inverse of the probability of detection."); Roger D. Blair & Thomas F. Cotter, *An Economic Analysis of Damages Rules in Intellectual Property Law*, 39 WM. & MARY L. REV. 1585, 1591 (1998) (remarking that damage multipliers may be necessary to compensate for low detection levels).

regardless of whether the infringer was "willful," unless the willfulness somehow increases the odds of nondetection. Second, if the private economic value of a patented innovation to a single patentee is substantially less than the social value of the innovation—so much so that the patentee has insufficient incentives to invest in R&D and commercialization to produce the innovation—then the patentee will need more reward than mere compensatory damages in the event of infringement.[38] However, this reward multiplier should not typically be borne by the infringer—who, like the patentee, generally enjoys none of the greater social benefits of the innovation—but rather by society as a whole (for example, the government).[39]

More generally, I argue that instead of focusing on the substantive rule at issue—here, identifying those actions that should count as "patent infringement"—policymakers and scholars should also examine ways to adjust the manner of enforcement, judicial procedure, and remedies to achieve effective substantive aims.[40] When the cost of particularized substantive rulemaking is high, as in the case of patent law,[41] particularized enforcement, procedure, and remedies may provide a better route for achieving optimal outcomes.[42] Unfortunately, modifying enforcement

---

38. As I argue below, most patentees will not require the full social value of their invention to appropriately incentivize them. *See infra* subpart II(B). However, there will surely be situations for which that is not so. *Cf.* Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, J. ECON. PERSP., Winter 1991, at 29, 31 (arguing that the only way to ensure that socially desirable innovations are researched is to allow research firms to "collect as revenue all the social value they create"); Steven Shavell & Tanguy van Ypersele, *Rewards Versus Intellectual Property Rights*, 44 J.L. & ECON. 525, 529 (2001) ("[I]ncentives to invest in research are inadequate because monopoly profits are less than the social surplus created by an innovation.").

39. *See infra* section II(C)(2).

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

41. Patent law is rife with industry variation in the economics of innovation incentives, but Congress and the courts have had difficulty in substantially differentiating the law for specific industries. Alan Devlin, *Patent Law's Parsimony Principle*, 25 BERKELEY TECH. L.J. 1693, 1707–08 (2010) ("[B]ecause it is so difficult to determine why particular inventors innovate, patent law has typically declined to incorporate inventor- or even industry-specific principles into its doctrine[, generally operating instead] on a 'one-size-fits-all' basis, attempting to spur optimal levels of innovation through the provision of largely uniform reward structures." (footnotes omitted)). Burk and Lemley argue that patent law is effectively differentiated in application. *See* Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155, 1156–57 (2002) (discussing the fact that general legal standards lead to diverse results in different technology areas such as software and pharmaceuticals); Burk & Lemley, *supra* note 26, at 1577 ("[D]espite the appearance of uniformity, patent law is actually as varied as the industries it seeks to foster."). However, such differentiation is ultimately constrained by patent law's textual uniformity and cannot achieve fully efficient differentiation. *See* Clarisa Long, *Our Uniform Patent System*, FED. LAW., Feb. 2008, at 44, 48 ("At present, the patent statute does provide for some technology-specific variation and exempts certain groups from liability, but these are the rare exceptions rather than the rule.").

42. *Cf.* Leandra Lederman & Ted Sichelman, *Enforcement as Substance in Tax Compliance*, 70 WASH. & LEE L. REV. 1679, 1747–49 (2013) (contending that tailoring enforcement rates by

approaches has generally been overlooked as a means for compensating for defects in the primary substantive law at issue.[43] Most analytical treatments of enforcement emphasize its costly nature and concern various mechanisms for increasing compliance while maintaining or reducing administrative costs.[44] In these models, if enforcement were costless, the ideal approach would be to punish legal actors for every violation of the law in order to ensure 100% compliance.[45]

However, Ian Ayres and Paul Klemperer have offered an alternative model,[46] extended by me in other work,[47] in which *the aim of enforcement is less than 100% compliance.*[48] Specifically, we show that imperfect, probabilistic enforcement of a patent may result in welfare outcomes superior to ironclad enforcement.[49] Such a counterintuitive goal is justified when the applicable substantive rule generates substantial unnecessary costs—for example, deadweight losses and duplicated development—and these costs cannot be easily remedied at the legislative level.[50] In these instances, modifying enforcement in the judicial or executive domains may significantly reduce the costs imposed by the substantive rule.[51] If these modification costs are relatively low, then "measuring" enforcement or

---

industry and product markets can sometimes reduce deadweight losses and other social costs resulting from taxation more efficiently than modifying the substantive tax law).

43. *See id.* at 1681 ("Scholars often assume that perfect enforcement of the laws, though unrealistic, is the ideal, and have focused on achieving the highest level of compliance possible at the lowest cost.").

44. *Id.*; *cf.* Gary S. Becker & George J. Stigler, *Law Enforcement, Malfeasance, and Compensation of Enforcers*, 3 J. LEG. STUD. 1, 1 (1974) ("Both the normative and positive approaches to legislation . . . generally have taken enforcement of laws for granted, and have not included systematic analyses of the cost of enforcing different kinds of laws.").

45. Lederman & Sichelman, *supra* note 42, at 1690–93; *see also* Blair & Cotter, *supra* note 37, at 1619 ("In order to deter infringement, we must have a set of rules that renders an infringement unprofitable."); Blair & Cotter, *supra* note 19, at 9 (assuming that infringement will necessarily reduce "incentive[s] to innovate" below the optimal level).

46. Ian Ayres & Paul Klemperer, *Limiting Patentees' Market Power Without Reducing Innovation Incentives: The Perverse Benefits of Uncertainty and Non-Injunctive Remedies*, 97 MICH. L. REV. 985, 988–89 (1999) ("[T]his Article shows that a regime with some uncertainty and delay can produce this reward more efficiently than a regime in which enforcement is instantaneous and certain.").

47. *See* Ted Sichelman, *Quantum Game Theory and Coordination in Intellectual Property* 5 (San Diego Legal Studies, Working Paper No. 10-035, 2010), *available at* http://ssrn.com/abstract=1656625 (extending the Ayres and Klemperer model to the context of patent races).

48. *See* Ayres & Klemperer, *supra* note 46, at 994–1000 (exploring the potential benefits of a patent regime in which remedies for infringement are awarded only a fraction of the time); Sichelman, *supra* note 47 (expanding and analyzing the Ayers and Klemperer model using a variant of quantum game theory).

49. *See* Ayers & Klemperer, *supra* note 46, at 994–1000 (demonstrating "how uncertainty and delay can produce higher welfare than an 'idealized' patent regime"); Sichelman, *supra* note 47, at 14–20 (same).

50. *See* Burk & Lemley, *supra* note 26, at 1635 ("[R]ewriting the patent law for each industry would involve substantial administrative costs and uncertainties."); *infra* section II(C)(1).

51. *See infra* subpart II(C)(1).

remedies to change the effect of a substantive rule may be a superior alternative to modifying the substantive rule directly via legislation or regulation.[52]

In broad terms, I argue that the blanket assumption of courts and scholars that traditional private law remedies are optimal for patent law is flawed. This Article proceeds to do so as follows. Part I describes the aims of patent law, showing that there is nearly universal agreement that the patent system's primary goal is to promote innovation, rather than to vindicate individual, private rights. In so doing, I draw upon scholarly work criticizing patent doctrine for overincorporating traditional contract and private property law concepts, but argue that this line of scholarship is incomplete for failing to critique the incorporation of traditional tort law principles into patent law.[53]

Part II explains in more detail why existing models of patent law remedies are deficient. In particular, contrary to the prevailing wisdom, I show that injunctions and full compensatory damages may be suboptimal not only for nonpracticing entities but also for practicing entities, in at least four situations: (1) the patent covers a minor component of a complex product and switching costs are high; (2) the costs and risks of innovation are relatively low; (3) large duplicated R&D costs are generated by inefficient rent-seeking; or (4) bargaining costs are far greater than the value of the innovation at issue. Conversely, I argue that awarding more than compensatory damages, even in the absence of willful infringement, may be optimal when infringement is difficult to detect or when the private returns from innovation fall far below the level needed to generate optimal social returns.

Part III addresses possible limitations and criticisms of my proposed model. First, I consider arguments that allowing judges to tailor remedies so as to effectively modify the substantive law threatens traditional separation of powers and related democratic concerns. I conclude that while Congress would need to provide the courts additional discretion to fully eliminate the make-whole approach, the courts have long retained a significant level of equitable discretion to fashion remedies that does not threaten democratic concerns. Second, I address potential concerns that a system unmoored from make-whole remedies would raise innovation costs by increasing overall uncertainty. Relatedly, I examine the concern that judges are insufficiently competent and knowledgeable to impose forward-looking damages in place of injunctions. Partially agreeing with this line of

52. *See* Lederman & Sichelman, *supra* note 42, at 1685 (proposing a regime of "measured enforcement" wherein the government intentionally engages in imperfect enforcement so as to reduce deadweight losses from taxation).

53. *See, e.g.*, Lemley, *supra* note 25, at 1032 (arguing that intellectual property should not be viewed through the lens of real property law, but overlooking similar problems in classifying intellectual property infringement as a species of tort).

critique, I advocate a near-term regime whereby traditional remedies are kept intact for most situations but are jettisoned in those cases that clearly lead to inefficient results. However, I propose a long-term regime in which remedies in all cases are fashioned with an eye towards optimizing innovation incentives.

Part IV discusses this long-term approach, contending that the remedies framework of the current Patent Act is fundamentally flawed. More generally, I conclude by arguing that the patent system should aim to develop sufficient competence and knowledge in its adjudicators in order to fully discard the notion that patentees should be returned to the *status quo ante*. Instead, patent law remedies should be tailored in each case in order to promote the types and levels of innovation that most benefit society. In this regard, drawing on the work of Shubha Ghosh,[54] I briefly propose a conceptual model that properly views patent law as a public regulatory mechanism, rather than a system of private law rights, duties, and remedies.

I. The Aim of Patent Law Is Optimally Promoting Innovation, Not Protecting Private Harms

In the United States, the overriding goal of patent law is to promote technological innovation.[55] In this regard, the Supreme Court has roundly rejected the natural rights notion that patents are a species of personal right designed to protect an inventor's interest in liberty, security, or personhood. As the Court explained at length in 1917 in *Motion Picture Patents*[56]:

> Since *Pennock v. Dialogue* was decided in 1829 this court has consistently held that the primary purpose of our patent laws is not the creation of private fortunes for the owners of patents but is to promote the progress of science and useful arts . . . .
>
> . . .
>
> Thirty years later this court, returning to the subject, in *Kendall v. Winsor*, again pointedly and significantly says:
>
> It is undeniably true, that the limited and temporary monopoly granted to inventors was never designed for their exclusive profit or advantage; the benefit to the public or community at large was . . . doubtless the primary object in granting and securing that monopoly.

---

54. Shubha Ghosh, *Patents and the Regulatory State: Rethinking the Patent Bargain Metaphor After* Eldred, 19 BERKELEY TECH. L.J. 1315 (2004) (proposing a public, regulatory model of patent law).

55. *See* U.S. CONST., art. I, § 8, cl. 8 ("To *promote the Progress* of Science and useful Arts . . . ." (emphasis added)); Bilski v. Kappos, 130 S. Ct. 3218, 3234 (2010) (Stevens, J., concurring) ("[T]he patent system is intended to protect and promote advances in science and technology . . . ." (alteration in original) (internal quotation marks omitted)).

56. Motion Picture Patents Co. v. Universal Film Mfg. Co., 243 U.S. 502 (1917).

> This court has never modified this statement of the relative importance of the public and private interests involved in every grant of a patent . . . .[57]

Nor has the Court wavered from this position since *Motion Picture Patents* was decided.[58]

Moreover, scholars generally agree—not just doctrinally but also as a policy matter—that the foundation of the patent laws should be utilitarian.[59] This prevailing view draws upon Thomas Jefferson's apt philosophy, which—in the Supreme Court's rendering in *Graham v. John Deere Co.*[60]—posited that "[t]he patent monopoly was not designed to secure to the inventor his natural right in his discoveries. Rather, it was a reward, an inducement, to bring forth new knowledge."[61] Because the patent system rests on a utilitarian basis, as the Court explained in *Kendall v. Winsor*,[62] any *private* benefit that an inventor receives from a patent is merely a means to an end of providing a benefit "to the public or community at large."[63] As such, any form of private law right afforded to the patentee is

---

57. *Id.* at 510–11 (citations omitted) (internal quotation marks omitted). Of course, some aspects of patent law promote individual interests at the expense of utilitarian aims. *See generally* Tom G. Palmer, *Are Patents and Copyrights Morally Justified? The Philosophy of Property Rights and Ideal Objects*, 13 HARV. J.L. & PUB. POL'Y 817 (1990) (proposing a Lockean labor theory approach to patent law). Yet, the primary goal—and the dominant explanation—of U.S. patent law is utilitarian in approach. *See* Menell, *supra* note 5, at 130 ("The United States Constitution expressly conditions the grant of power to Congress to create patent and copyright laws upon a utilitarian foundation."); *infra* note 59 and accompanying text.

58. *See* Quanta Computer, Inc. v. LG Elecs., Inc., 553 U.S. 617, 626–28 (2008) (applying *Motion Picture* to exhaustion of patent rights following an unconditional sale of a patented product); Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 480 (1974) (observing that patent laws promote "progress by offering a right of exclusion for a limited period as an incentive to inventors to risk the often enormous costs in terms of time, research, and development" and thus engender a "positive effect on society through the introduction of new products and processes of manufacture into the economy, and the emanations by way of increased employment and better lives for our citizens"); Mercoid Corp. v. Mid-Continent Inv. Co., 320 U.S. 661, 665 (1944) ("It is the public interest which is dominant in the patent system."); *cf.* Mazer v. Stein, 347 U.S. 201, 219 (1954) ("The copyright law, like the patent statutes, makes reward to the owner a secondary consideration." (internal quotation marks omitted)).

59. *See* Menell, *supra* note 5, at 130 ("[T]he principal philosophical theory applied to the protection of utilitarian works—that is, technological inventions—has been utilitarianism."); *see also* Alan Devlin & Neel Sukhatme, *Self-Realizing Inventions and the Utilitarian Foundations of Patent Law*, 51 WM. & MARY L. REV. 897, 901 (2009) ("Almost all commentators and judges agree that utilitarian considerations enjoy hegemonic status in patent jurisprudence, such that the purpose of the patent system is to induce the creation and commercialization of technology that otherwise could be easily appropriated.").

60. 383 U.S. 1 (1966).

61. *Id.* at 9; *see also* Devlin & Sukhatme, *supra* note 59 ("As a patent monopoly carries with it potentially large social welfare costs, it can be condoned only when necessary to incentivize the creation and dissemination of disproportionately valuable information.").

62. 62 U.S. (21 How.) 322 (1858).

63. *Id.* at 327–28.

purely incidental to the aim of patent law in promoting innovation.[64] By implication, the vindication of patent law rights via litigation is not meant to remedy a personal wrong inflicted on the inventor.[65]

Thus, patent law rights stand in stark contrast to traditional private law rights, such as those that sound in real property, contract, or tort.[66] At least historically, the vindication of these kinds of private law rights protected *individual* interests, and these rights grew out of a common law tradition not so much concerned with utilitarian values,[67] but instead personal autonomy and liberty.[68] So when a trespasser invaded an owner's land, the owner—to the extent he had Blackstonian property rights that afforded him

---

64. *See* La Belle, *supra* note 19, at 50 ("Simply put, patents are a means to an end. And it is that end—the promotion of innovation for the public good—that is paramount." (footnotes omitted)); *cf.* Wendy J. Gordon, *Of Harms and Benefits: Torts, Restitution, and Intellectual Property*, 21 J. LEGAL STUD. 449, 450 (1992) ("Uncompensated use of an inexhaustible good is worth discouraging only as a means to an end: obtaining adequate incentives for the good's initial production and maintenance.").

65. Of course, one might argue that the patent laws should as a normative matter protect private interests. *See* ROBERT P. MERGES, JUSTIFYING INTELLECTUAL PROPERTY 14–19 (2011) (espousing a theory of intellectual property based on the labor desert theory of Locke and deontological approach of Kant); Adam D. Moore, *A Lockean Theory of Intellectual Property*, 21 HAMLINE L. REV. 65, 66 (1997) (arguing "that incentives based rule-utilitarian arguments fail to justify anything remotely close to modern Anglo-American copyright, patent, and trade secret institutions"); *see also* Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287, 296–330 (1988) (describing intellectual property according to a Lockean labor theory). And, of course, these approaches would generally counsel in favor of retaining make-whole remedies. However, consistent with the views of the vast majority of courts and scholars, including my own views, I only consider utilitarian approaches herein. *See* ROBERT P. MERGES, PETER S. MENELL & MARK A. LEMLEY, INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE 11 (5th ed. 2010) [hereinafter MERGES ET AL., INTELLECTUAL PROPERTY] (remarking that utilitarianism is "the dominant paradigm for analyzing and justifying the various forms of intellectual property protection"); John M. Golden, *Principles for Patent Remedies*, 88 TEXAS L. REV. 505, 509 (2010) ("For purposes of simplicity and, in many quarters, plausibility, I generally assume a utilitarian goal that is standard in modern accounts: the patent system should act to promote the development, disclosure, and use of new technologies, ideally in a way that maximizes social welfare."); Sichelman, *supra* note 35, at 379 (explaining that the foundation of U.S. patent law is utilitarian in nature).

66. Even Merges admits as much. *See* ROBERT P. MERGES ET AL., INTELLECTUAL PROPERTY, *supra* note 65, at 133–34 (noting that patent rights are modeled on utilitarian, not deontological, principles).

67. *See* Francis H. Bohlen, *Voluntary Assumption of Risk*, 20 HARV. L. REV. 14, 14 (1907) (noting "the individualistic tendency of the common law . . . proceeding from the people and asserting their liberties"); *cf.* Roscoe Pound, *Do We Need a Philosophy of Law?*, 5 COLUM. L. REV. 339, 346 (1905) ("Men have changed their views as to the relative importance of the individual and of society; but the common law has not. . . . We no longer hold that society exists entirely for the sake of the individual. . . . The common law, however, is concerned . . . with individual rights.").

68. *See* 1 WILLIAM BLACKSTONE, COMMENTARIES *139 ("So great . . . is the regard of the law for private property, that it will not authorise the least violation of it; . . . not even for the general good of the whole community."); Thomas Merrill, *Property and the Right to Exclude*, 77 NEB. L. REV. 730, 752 (1998) ("[T]he hallmark of property . . . is an individual entitlement grounded in state law, which cannot be removed except for cause." (internal quotation marks omitted)).

"sole and despotic dominion" over his land[69]—was entitled to exclude the trespasser under all circumstances in order to vindicate his *individual* ownership interests.[70] Any sort of exception to this rule—for instance, in the event the trespasser crossed onto the owner's land in order to avoid death[71]—was typically a narrow gloss to the individualistic nature of the underlying right.[72]

More recently, private law scholars have embraced more public-oriented, utilitarian approaches to contract, property, and tort law, particularly in the field of law and economics.[73] Yet, as I shall describe, patent law has oddly clung to the more Blackstonian notion of private law. Oddly—unlike the recent law-and-economics accounts of private law—from its constitutional origins, patent law in the United States has been decidedly utilitarian.[74] In this sense, the patentholder is more akin to a

---

69. 2 WILLIAM BLACKSTONE, COMMENTARIES *1–2 ("There is nothing which so generally strikes the imagination, and engages the affections of mankind, as the right of property; or that sole and despotic dominion which one man claims and exercises over the external things of the world, in total exclusion of the right of any other individual . . . ."); *see also* OLIVER WENDELL HOLMES, JR., THE COMMON LAW 222 (Harv. Univ. Press 2009) (1881) ("The owner is allowed to exclude all, and is accountable to no one."). Jeremy Bentham, a leading founder of utilitarianism, decidedly criticized the Blackstonian approach. Bentham wrote of "the universal inaccuracy and confusion which seemed to my apprehension to pervade the whole" of the work. JEREMY BENTHAM, A FRAGMENT ON GOVERNMENT 4 (J.H. Burns & H.L.A. Hart eds., Cambridge Univ. Press 1988) (1776).

70. *See* Carl J. Circo, *Does Sustainability Require a New Theory of Property Rights?*, 58 U. KAN. L. REV. 91, 105 (2009) ("Blackstone's famous reference to an owner's absolute dominion generally comports with the centrality of individual rights in property as a common law tenet."); Shyamkrishna Balganesh, *Demystifying the* Right *to Exclude: Of Property, Inviolability, and Automatic Injunctions*, 31 HARV. J.L. & PUB. POL'Y 593, 598 (2008) ("For quite some time, the right to exclude in the context of both tangible and intangible property has come to be associated with an entitlement to exclusionary (injunctive) relief. Thus, interferences with an owner's interests are thought to entitle the owner to obtain a permanent injunction restraining such interferences.").

71. *See* Eduardo Moisés Peñalver & Sonia K. Katyal, *Property Outlaws*, 155 U. PA. L. REV. 1095, 1172 (2007) ("The doctrine of necessity permits nonowners to trespass on, and under certain circumstances even to appropriate, the property of others in order to avoid a grave harm."); *see also* WILLIAM B. STOEBUCK & DALE A. WHITMAN, THE LAW OF PROPERTY 411 (3d ed. 2000) ("[O]thers [may] enter with the rightful possessor's permission but [also] in a few cases, against his will.").

72. *See* Peñalver & Katyal, *supra* note 71, at 1173 ("Many courts . . . have interpreted the necessity defense in an artificially narrow way that would restrict it to extremely unusual circumstances . . . ."). *But cf.* Adam Mossoff, *The Trespass Fallacy in Patent Law* 14–16 (George Mason Law & Econ. Research Paper No. 12-54, 2012), *available at* http://ssrn.com/abstract=2126595 (noting that historical trespass and related disputes were sometimes subject to substantial uncertainty).

73. Abraham Bell & Gideon Parchomovsky, *A Theory of Property*, 90 CORNELL L. REV. 531, 542 (2005) ("[M]ost scholars today base their understandings of property on a model where property is justified by utilitarianism and defined by positive law rather than upon natural rights theories.").

74. *See* Tun-Jen Chiang, *Fixing Patent Boundaries*, 108 MICH. L. REV. 523, 553–54 (2010) (arguing that the proper mechanism for analyzing the problem of claim amending, and determining on whom to place the burden of loss, is found in the utilitarian purpose of patent law);

private attorney general, paid via the enforcement of his right as a reward for benefitting the *public*, than a vindicator of his own *private* rights.[75] As many scholars have rightly recognized, instead of conferring private rights on inventors, a public prize system—whereby innovators are merely paid by the State for their endeavors—could, at least in theory, function equally well to promote innovation.[76] It is for economic reasons, rather than the protection of individual interests, that the United States decided that a patent system was generally more effective than a prize system, patronage, or similar means to encourage innovation.[77]

Yet, judges and scholars came to view the means as the ends. Thus, patents are now viewed not merely as a mechanism the State uses to induce innovation but as a "contract between the inventor and the public," affording private rights and remedies to the inventor–contractee.[78]

Devlin & Sukhatme, *supra* note 59, at 897 ("Unlike other forms of intellectual property, patents are universally justified on utilitarian grounds alone."); John F. Duffy, *Inventing Invention: A Case Study of Legal Innovation*, 86 TEXAS L. REV. 1, 5 (2007) ("[U]nlike many other areas, such as tort and criminal law, the patent system has long been based on utilitarian considerations rather than considerations of fairness or justice."); Adam Mossoff, *Rethinking the Development of Patents: An Intellectual History, 1550-1800*, 52 HASTINGS L.J. 1255, 1315 (2001) ("One common interpretation of eighteenth-century patent doctrine is that patents evolved according to a utilitarian framework with the inventor being granted a reward in exchange for the benefit that accrued to the public."); David S. Olson, *Taking the Utilitarian Basis for Patent Law Seriously: The Case for Restricting Patentable Subject Matter*, 82 TEMP. L. REV. 181, 182–83 (2009) ("There is widespread agreement that the reason we have a patent system is utilitarian—to solve a market failure problem. . . . Accordingly, a properly crafted patent law should provide enough property rights to incentivize the socially desirable (efficient) level of innovation, and no more.").

75. *See* Olatunde C.A. Johnson, *Beyond the Private Attorney General: Equality Directives in American Law*, 87 N.Y.U. L. REV. 1339, 1347 (2012) (explaining that the private attorney general model enables "Congress [to] vindicate important public policy goals by empowering private individuals to bring suit"); *see also* Daniel Guttman, *Public Purpose and Private Service: The Twentieth Century Culture of Contracting Out and the Evolving Law of Diffused Sovereignty*, 52 ADMIN. L. REV. 859, 917–18 (2000) ("While the tradition of third party government traced here dates to the early twentieth century, there is a much older tradition of the use of private attorneys general or bounty hunters to vindicate public rights, acting in the name of the king or the sovereign.").

76. *See, e.g.*, Michael Abramowicz, *Perfecting Patent Prizes*, 56 VAND. L. REV. 115, 236 (2003) ("With a carefully designed prize system, even if an individual decisionmaker makes an error or is influenced by political considerations in calculating a prize, these flaws will not affect the decisions that matter—those made in anticipation of the eventual governmental awards."); Joseph E. Stiglitz, *Economic Foundations of Intellectual Property Rights*, 57 DUKE L.J. 1693, 1721 (2008) ("[I]n areas where there are well-defined needs . . . both the prize system and the patent system can provide comparable incentives to undertake research.").

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

78. Davis Airfoils, Inc. v. United States, 124 F. Supp. 350, 352 (Ct. Cl. 1954) ("A patent is a contract between the inventor and the public, the terms of which are formulated by the United States Patent Office. The inventor in such a contract gives as a consideration to the public a new and useful art, machine or composition of matter, and, in return, the public gives as a consideration to the inventor a monopoly expressed by the claims of the patent of a period limited by statute . . . ."); *see also* Ghosh, *supra* note 54, at 1316 ("Patents are commonly understood as a hypothetical contract between the inventor and the government resulting in a quid pro quo of innovation for exclusivity.").

Similarly, instead of treating patents as property-like rights used by inventors to collect payment from society as inducement for their innovative efforts, patents are viewed at root as property rights.[79] And, last, patent infringement is not just a mode of cataloguing which third parties should provide recompense to the inventor so as to guarantee the appropriate level of R&D and commercialization incentives but rather a species of common law tort.[80] Judge Frank Easterbrook encapsulates this sentiment when he writes, "Patents give a right to exclude, just as the law of trespass does with real property."[81]

This reification of a patent right as fundamentally a private law right, with all its attendant remedies, is not only conceptually flawed but leads to pernicious outcomes. Several scholars have recognized as much for the contract and real property law models of patents. Shubha Ghosh has challenged the soundness of treating patents as contracts with the State.[82] In particular, he argues that the "patent law consists of more than a bargain between the inventor and the state; it includes other parties, such as consumers of the invention, follow-on inventors, students of the knowledge created by the innovation, and disseminators of the invention through market and non-market processes."[83] Mark Lemley has convincingly dismissed patent law's reliance on real property, particularly the notion that patentholders should be able to exclude infringers at-will through injunctive relief.[84]

---

79. *See* Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 535 U.S. 722, 730 (2002) ("The patent laws 'promote the Progress of Science and useful Arts' by rewarding innovation with a temporary monopoly. The monopoly is a property right . . . ." (quoting U.S. CONST., art. I, § 8, cl. 8)); Zoltek Corp. v. United States, 58 Fed. Cl. 688, 696 (Fed. Cl. 2003) ("The Federal Circuit, its predecessor court, the Court of Claims, and the U.S. Supreme Court have repeatedly recognized that patent rights are property rights."); John F. Duffy, Comment, *Intellectual Property Isolationism and the Average Cost Thesis*, 83 TEXAS L. REV. 1077, 1078 (2005) ("I, along with others, believe that intellectual property should be treated as a species of property."); Adam Mossoff, *Exclusion and Exclusive Use in Patent Law*, 22 HARV. J.L. & TECH. 321, 322 (2009) ("The status of patents is undisputed: patents are property."); Adam Mossoff, *Patents as Constitutional Private Property: The Historical Protection of Patents Under the Takings Clause*, 87 B.U. L. REV. 689, 690 (2007) ("Patents are property."); Mossoff, *supra* note 74, at 1255–58 (tracing the origins of the incorporation of private law concepts, such as property and contract, into patent law).

80. Mars, Inc. v. Coin Acceptors, Inc., 527 F.3d 1359, 1365 (Fed. Cir. 2008) ("Patent infringement is a tort.").

81. Frank H. Easterbrook, *Intellectual Property Is Still Property*, 13 HARV. J.L. & PUB. POL'Y 108, 109 (1990); *see also* Orin S. Kerr, *Rethinking Patent Law in the Administrative State*, 42 WM. & MARY L. REV. 127, 132–53 (2000) (setting forth a "private law theory" of patent law).

82. *See* Ghosh, *supra* note 54, at 1339.

83. *Id.*; *see also* La Belle, *supra* note 19, at 43 ("[P]atent validity challenges are complaints about government conduct that implicate important public interests and potentially affect many parties not before the court. Validity disputes, therefore, fit within [the] paradigm of public law litigation.").

84. *See* Lemley, *supra* note 25, at 1046–65 (concluding that granting intellectual property rights imposes complex economic costs, which can only be justified to the extent necessary to

Despite critiques such as Ghosh's and Lemley's, scholars have not recognized—at least in any systematic fashion—that classifying infringement as a common law tort is just as problematic as viewing patents as private contracts or a form of real property.[85] Indeed, courts and scholars have promoted infringement as a form of tort.[86] An illustrative example is the Supreme Court's view of patents in 1915 from *Dowagiac Manufacturing Co. v. Minnesota Moline Plow Co.*,[87] in which it remarked that "[a]s the exclusive right conferred by the patent was property and the

---

provide incentives to innovate); *see also* Daniel A. Crane, *Intellectual Liability*, 88 TEXAS L. REV. 253, 253–54 (2009) (arguing intellectual property should not be treated as real property under the law); Mark A. Lemley, *Romantic Authorship and the Rhetoric of Property*, 75 TEXAS L. REV. 873, 895–96, 902 (1997) (book review) (describing the emergence of the property view of intellectual property and arguing against the propertization of intellectual property rights).

85. Lemley has rejected tort law as a model for intellectual property, not because of its compensatory aims but because of a fear that it would lead to restitutionary remedies that focus on the alleged infringer, rather than the IP rights holder. He explains:

> In one sense, treating intellectual property as a form of tort law is consistent with the economic lessons of the previous Parts. A focus on harm to the intellectual property owner, rather than on the benefit conferred on the infringer, is consistent with optimal intellectual property policy. . . .
>
> . . . . [But] my fear . . . is that drawing too close an analogy to the tort system will encourage the courts to focus attention on how the defendant was enriched, not on the need for compensating intellectual property owners.

Lemley, *supra* note 25, at 1072. So Lemley in essence embraces the traditional compensatory aim of tort law as it applies to patent remedies, which is consistent with his views in his later work. *See infra* subpart II(B). Indeed, he rejects a public regulatory model of IP. Lemley, *supra* note 25, at 1074 ("Regulation is out of vogue, and those who talk about intellectual property as regulation usually do so in order to denigrate it.").

86. *See, e.g.*, *In re* Cambridge Biotech Corp., 186 F.3d 1356, 1371 (Fed. Cir. 1999) ("Patent infringement is properly classified as a tort . . . ."); Blair & Cotter, *supra* note 19 (arguing that patent infringement is best characterized as a tort and therefore proximate cause is an appropriate way to determine damages); Roger D. Blair & Thomas F. Cotter, *Strict Liability and Its Alternatives in Patent Law*, 17 BERKELEY TECH. L.J. 799, 800 (2002) ("Patent infringement is often characterized as a strict liability tort . . . ."); Kenneth W. Dam, *The Economic Underpinnings of Patent Law*, 23 J. LEGAL STUD. 247, 255 (1994) ("Remedies for infringement of a patent are, with limited exceptions, those appropriate for property."); Heald, *supra* note 19 ("In establishing what constitutes infringement and what remedies apply, patent law's secondary function looks like tort law."); Christopher M. Newman, *Patent Infringement as Nuisance*, 59 CATH. U. L. REV. 61, 67–68, 115–18 (2009) (comparing in economic terms patent infringement to traditional nuisance); Michael L. Rustad, *Torts as Public Wrongs*, 38 PEPP. L. REV. 433, 514 (2011) ("Patent and copyright have long been regarded as creatures of federal statute, but infringement was historically considered to be a tort. For instance, the infringement of a patent was classifiable as a tort—namely, trespass on the case." (footnotes omitted)); Sandrik, *supra* note 4, at 99 (proposing an "infringement-as-trespass framework" as a basis for understanding patent infringement); *cf.* Dan L. Burk, *The Trouble with Trespass*, 4 J. SMALL & EMERGING BUS. L. 27, 53 (2000) (arguing that nuisance is an appropriate framework for addressing certain cyberspace disputes); Avihay Dorfman & Assaf Jacob, *Copyright as Tort*, 12 THEORETICAL INQUIRIES L. 59 (2011) (proposing a tort framework for understanding copyright infringement); Mossoff, *supra* note 72, at 11–12 (rejecting the analogy of infringement to pure trespass but proposing a complex analogy based on the violation of various real property interests).

87. 235 U.S. 641 (1915).

infringement was a tortious taking of a part of that property, the normal measure of damages was the value of what was taken."[88]

As I argue in more detail in Part II, a major problem with viewing patent infringement as a tort is that the private law remedies usually associated with tort law—injunctions and compensatory damages—are not always sensible for optimally encouraging innovation. Rather, I show that in many situations, denying injunctions—even to practicing patentees—and limiting damages to levels below that needed for full compensation may better promote the utilitarian aims of patent law.[89] Akin to Lemley's critique of the view that a patent is a form of real property that should allow its owner to internalize the full social value of the invention,[90] my critique contends that a patent should not necessarily allow its owner to collect the full value of the damages inflicted by an infringer.[91] In yet other situations, I argue that limiting the patentee's remedy merely to make-whole damages may be insufficient—such as when the overall social value of an invention far exceeds its economic value to private actors in the marketplace.[92] In these cases, more than make-whole damages—supplied in part from the public fisc—may be a more sensible means of optimizing innovation incentives.[93] In order to arrive at these conclusions, I first critique the patent system's current remedy scheme.

II. Reconfiguring Patent Law Remedies to Promote Innovation

The standard approach to patent law remedies has followed the dominant view that patent law is a form of private law.[94] Thus, a patentee's "right to exclude" has historically been treated by courts as akin to a landowner's right to keep out trespassers, leading judges to grant injunctive relief as a matter of course to prevent ongoing infringement.[95] Relatedly,

---

88. *Id.* at 648.

89. *See infra* section II(C)(1).

90. Lemley, *supra* note 25, at 1046 ("The assumption that intellectual property owners should be entitled to capture the full social surplus of their invention runs counter to our economic intuitions in every other segment of the economy. We do not permit producers to capture the full social value of their output. Nor do we permit the owners even of real property to internalize the full positive externalities associated with their property.").

91. *See infra* section II(C)(1).

92. *See infra* section II(C)(2).

93. *See infra* section II(C)(2).

94. *See, e.g.*, Carbice Corp. of Am. v. Am. Patents Dev. Corp., 283 U.S. 27, 33 (1930) ("Infringement, whether direct or contributory, is essentially a tort, and implies invasion of some right of the patentee."); Pall Corp. v. Micron Separations, Inc., 66 F.3d 1211, 1221 (Fed. Cir. 1995) ("[P]atent infringement is a continuing tort . . . ."); Rite-Hite Corp. v. Kelley Co., 56 F.3d 1538, 1578 (Fed. Cir. 1995) (Newman, J., concurring in part and dissenting in part) ("Patent infringement is a commercial tort, and the remedy should compensate for the actual financial injury that was caused by the tort.").

95. *See, e.g.*, Richardson v. Suzuki Motor Co., 868 F.2d 1226, 1246–47 (Fed. Cir. 1989) ("Infringement having been established, it is contrary to the laws of property, of which the patent

like an injured party in tort, Congress has legislated, and courts have held, that patentees are entitled to be "made whole" for any past infringement, with damages compensating them for the full extent of the "harm" inflicted upon them by an infringer.[96] Thus, manufacturing patentees are entitled under current law to a full measure of "lost profits"—and licensing patentees to "reasonable royalties"—in order to compensate them as they would have been paid in a market scenario in which the patent was not infringed.[97]

In the past decade, there has been an assault on the practice of treating a patent as a form of absolute property entitling its holder to injunctive relief.[98] In large part, this reaction has been driven by the rise of nonpracticing entities (NPEs), particularly the proliferation of lawsuits filed on patents perceived to be "weak," because they are likely to be held not infringed, invalid, or unenforceable.[99] Despite the seemingly questionable nature of these patents, critics argue that many of these suits have led to relatively large nuisance-value settlements given the high costs and risks of

---

law partakes, to deny the patentee's right to exclude others from use of his property."); Easterbrook, *supra* note 81 ("Patents give a right to exclude, just as the law of trespass does with real property. Intellectual property is intangible, but the right to exclude is no different in principle from General Motors' right to exclude Ford from using its assembly line . . . .").

96. *See* Landers, *supra* note 18 ("[P]atent damages are a make-whole remedy, intended to restore the patentee to the same position as before infringement."); John W. Schlicher, *Measuring Patent Damages by the Market Value of Inventions—The* Grain Processing, Rite-Hite, *and* Aro *Rules*, 82 J. PAT. & TRADEMARK OFF. SOC'Y 503, 503 (2000) ("[P]atent damages should award patent owners the market values or economic benefits of the inventions *that they were unable to capture* due to infringement." (emphasis added)).

97. *See* Aro Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 507 (1964) ("[T]hat question [is] primarily: had the Infringer not infringed, what would Patent Holder-Licensee have made?" (second alteration in original)); Yale Lock Mfg. Co. v. Sargent, 117 U.S. 536, 552 (1886) (stating that a patentee's damages are "the difference between his pecuniary condition after the infringement, and what his condition would have been if the infringement had not occurred"). To be certain, there is some circularity in determining a reasonable royalty. *See* Shapiro, *supra* note 12, at 308 ("[C]ircularity arises because reasonable royalties are often based on the royalties actually negotiated in the shadow of litigation, and these negotiated royalties depend . . . upon the magnitude of damages that courts are expected to award if the parties are unable to sign a licensing deal and instead engage in patent litigation.").

98. *See, e.g.*, eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388, 392–93 (2006) (rejecting "invitations to replace traditional equitable considerations with a rule that an injunction automatically follows a determination that a copyright has been infringed"); Landers, *supra* note 19 (contending that patent remedies should differ as between practicing and nonpracticing entities because each incurs different kinds of harm); Lemley & Shapiro, *supra* note 19 (arguing that courts should consider withholding injunctive relief for nonpracticing entities when only one of many different components that make up an infringing product is the subject of the patent-in-suit).

99. *eBay*, 547 U.S. at 397 (Kennedy, J., concurring) (describing the "potential vagueness and suspect validity" of some of "the burgeoning number of patents over business methods"); *see also Patent Quality Improvement: Hearing Before the Subcomm. on Courts, the Internet, and Intellectual Prop. of the H. Comm. on the Judiciary*, 108th Cong. 21 (2003) (testimony of David Simon, Chief Patent Counsel for Intel Corporation) (categorizing NPEs as "patent system bottom feeders" who buy "improvidently-granted patents from distressed companies for the sole purpose of suing legitimate businesses").

litigation for defendants.[100] Commentators have reasoned that if NPEs only licensed their patents in the marketplace—by definition, they do not manufacture and sell products—then returning them to the *status quo ante* via a patent remedy need only compensate them for lost royalties.[101] Thus, the standard remedy of providing injunctive relief to prevent ongoing infringement to NPEs is, on this view, unnecessary to fulfill the private law vision of patent remedies as "making the patentee whole."[102] On this ground, Justice Kennedy concurred in *eBay* to argue that NPEs typically should not be entitled to injunctive relief.[103]

This widely followed concurrence has led to a substantial reduction in the grant of injunctive relief to NPEs.[104] Yet, as I argued earlier, Justice Kennedy relied on the faulty assumption that make-whole private law remedies optimally promote innovative activity.[105] As I explain in more detail in this Part, in many situations, make-whole remedies may impede efficient levels of innovation.[106] For instance, when a patent merely covers a component of a larger product and the infringer's switching costs are high, I contend that not only NPEs *but also manufacturing entities* should regularly be denied injunctive relief.[107] Thus, in order to fully excise the

---

100. *See* James Bessen & Michael J. Meurer, Essay, *The Direct Costs from NPE Disputes*, 99 CORNELL L. REV. 387, 388–91 (2014) (contending that NPEs resulted in approximately $29 billion in direct costs for accused infringers in 2011); Brian J. Love & James C. Yoon, *Expanding Patent Law's Customer Suit Exception*, 93 B.U. L. REV. 1605, 1607–09 (2013) (discussing "bottom feeder" NPEs that file "nuisance-value patent suits"); Michael J. Meurer, *Controlling Opportunistic and Anti-Competitive Intellectual Property Litigation*, 44 B.C. L. REV. 509, 512–16 (2003) (arguing that NPEs use the threat of injunction and high litigation costs to extract exorbitant settlement or licensing fees).

101. *See infra* subpart II(B).

102. *See* Lemley, *supra* note 6, at 661 ("[W]hat it takes to make the patentee whole is very different if the patentee's only interest is in licensing the patent than if the patentee's only interest is in excluding competition and maintaining a monopoly price." (internal quotation marks omitted)); FTC, *supra* note 15, at 38–40 (noting that NPEs do not produce and sell products and thus are not vulnerable to countersuit, presenting a potential "hold up" threat to the accused infringer).

103. *See eBay*, 547 U.S. at 396 (Kennedy, J., concurring) ("[A]n industry has developed in which firms use patents not as a basis for producing and selling goods but, instead, primarily for obtaining licensing fees . . . . For these firms, an injunction, and the potentially serious sanctions arising from its violation, can be employed as a bargaining tool to charge exorbitant fees to companies that seek to buy licenses to practice the patent.").

104. *See id.* (noting that prior cases established a pattern of granting an injunctive remedy "almost as a matter of course"); FTC, *supra* note 13 (finding that since *eBay*, NPEs have been granted injunctive relief in approximately 45% of the cases in which remedies were awarded and the practicing nature of the party was mentioned).

105. *See supra* notes 11–19 and accompanying text.

106. *See infra* subpart II(B).

107. *See infra* subpart II(B). Recently, in *Apple, Inc. v. Motorola, Inc.*, 869 F. Supp. 2d 901 (N.D. Ill. 2012), Judge Posner denied injunctive relief to Motorola, a practicing entity. *Id.* at 904, 913–15. However, he did so because Motorola had agreed to license the patent-in-issue on fair, reasonable, and nondiscriminatory (FRAND) terms as part of its membership in the standards-setting organization (SSO). *See id.* at 911, 913–14. Thus, denying Motorola injunctive relief

private law specter from patent law remedies, it is not only necessary to remove real property concepts, but also tort concepts, from the doctrinal mix.

In this Part, I begin by recounting the traditional model of patent law remedies. Next, I describe the recent attack on this model for incorporating contract and real property notions. Finally, I argue that in order to complete the attack, it is necessary to counter patent law's reliance on common law tort.

A. The Standard Model of Patent Law Remedies

The usual theory explaining why inventors or their employers seek patents assumes that patentees can potentially generate greater-than-average profits on their patented products[108] by preventing others from making, using, and selling those products.[109] According to this theory, society benefits because these supernormal returns compensate for a market defect—namely, that the copying and selling of innovative products by others can often be achieved cheaply and quickly,[110] which can suboptimally thwart innovation.[111] Adopting exclusionary rights via patents

---

would not thwart its return to the *status quo ante* because Apple had an *ex ante* right to practice Motorola's patent in the marketplace in exchange for a fair and reasonable royalty. *See id.* at 914 ("By committing to license its patents on FRAND terms, Motorola committed to license the '898 to anyone willing to pay a FRAND royalty and thus implicitly acknowledged that a royalty is adequate compensation for a license to use that patent."); *cf.* Crane, *supra* note 84, at 289 ("For example, the rule might be that participation in an SSO automatically leads to the waiver of the right to exclude."). My proposal goes well beyond FRAND situations because it advocates denying injunctions even when the infringer has no right to practice the patent whatsoever. *See infra* subpart II(B).

108. More specifically, patents can also cover methods, processes, and services. *See* 35 U.S.C. § 101 (2006). The reference to "products" in the text is illustrative.

109. *See id.* § 271(a) (2006) (preventing the making, using, and selling of patented products and methods); WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 294–300 (2003) ("The standard rationale of patent law is that it is an efficient method of enabling the benefits of research and development to be internalized, thus promoting innovation and technological progress."); F.M. SCHERER & DAVID ROSS, INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE 621–24 (3d ed. 1990) (explaining that patent protection prevents competition and allows the inventors to recoup investment).

110. *See* Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, *in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609, 615 (1962) ("In the absence of special legal protection, the owner cannot, however, simply sell information on the open market. Any one purchaser can destroy the monopoly, since he can reproduce the information at little or no cost."); Douglas Gary Lichtman, *The Economics of Innovation: Protecting Unpatentable Goods*, 81 MINN. L. REV. 693, 701–02 (1997) (describing the traditional IP assumption that in the absence of intellectual property protection, new inventions could be easily copied).

111. *See, e.g.*, Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017, 1024–25 (1989) ("The incentive to invent theory holds that too few inventions will be made in the absence of patent protection because inventions once made are easily appropriated by competitors of the original inventor who have not shared in the costs of invention."); Edmund W. Kitch, *The Nature and Function of the Patent System*, 20

and attendant liability for infringement makes copying costly.[112] Thus, the exclusionary rights afforded by patents promote a more optimal level of innovation by providing greater incentives to innovators to invent, market, and sell innovative products,[113] as well as to disclose the knowledge underlying those innovations in the form of published patent documents.[114] Importantly, "[t]his traditional conception requires exclusivity; the value of a patent is accordingly commensurate with the value of the market or market niche it controls."[115]

Taking the traditional account of patent law on its face, the standard view holds that patent damages should return the patentee to the hypothetical state of affairs that would have obtained but for the infringement.[116] For infringement prior to suit, of course, the only possible remedy is some form of monetary payment sufficient to "compensate patentees for their losses."[117] Presuit damages divide into two camps: lost profits and reasonable royalties.[118] Lost profits are available to

---

J.L. & ECON. 265, 276–77 (1977) (arguing that the patent system promotes innovation by allowing inventors to invest in R&D without the fear of their discoveries being copied without their permission); Lichtman, *supra* note 110 (noting that if others could easily copy inventions, "few would want to be innovators, preferring instead to wait and free-ride on someone else's good idea").

112. *See* Lichtman, *supra* note 110, at 700–02 (explaining that under the traditional view, exclusionary rights increase the cost of copying by eliminating the ability for copiers to avoid the development costs incurred through innovation by the original creator).

113. *See* SCHERER & ROSS, *supra* note 109, at 622–23 (using economic models to illustrate how a lack of patent protection leads to innovators internalizing less of the social benefit from innovation); Kitch, *supra* note 111, at 265–67 (espousing a view that patents provide post-patenting incentives to market and distribute inventions).

114. *See* Martin J. Adelman, *Property Rights Theory and Patent-Antitrust: The Role of Compulsory Licensing*, 52 N.Y.U. L. REV. 977, 982 (1977) (observing that secrecy deprives the public of the full benefit of the invention and may also lead to duplicative R&D efforts); Jason Rantanen, *Peripheral Disclosure*, 74 U. PITT. L. REV. 1, 4 (2012) ("Given the critical role information plays in invention, it is unsurprising that the patent system has long been justified on the ground that it encourages the disclosure . . . .").

115. Lemley, *supra* note 6, at 657; *see also* David W. Opderbeck, *Patent Damages Reform and the Shape of Patent Law*, 89 B.U. L. REV. 127, 185 (2009) ("The degree of market power possessed by the owner of a patent on a method or process is a measure of the potential rents the patentee can control by virtue of the patent.").

116. Lemley, *supra* note 6 ("Patent damages are supposed to compensate patent owners for their losses, putting them back in the world they would have inhabited but for infringement."); *see also* BLAIR & COTTER, *supra* note 9, at 113–14 ("Irrespective of the fact that the infringer took measures to avoid patent infringement, if she infringed a valid patent, the infringer will be liable for the full economic injuries that the infringement caused. In principle, the patentee will be fully compensated for any injury due to the infringement."); Blair & Cotter, *supra* note 19, at 44–70 (advocating a "but for" standard of patent remedies).

117. *See* Lemley, *supra* note 6, at 655 (citing Pall Corp. v. Micron Separations, Inc., 66 F.3d 1211, 1223 (Fed. Cir. 1995)); *see also* 35 U.S.C. § 284 (2006) ("[T]he court shall award the claimant damages adequate to compensate for the infringement.").

118. *See* 35 U.S.C. § 284; Lemley, *supra* note 6, at 655 ("Courts interpreting this provision have divided patent damages into two groups—lost profits . . . and reasonable royalties."). Disgorgement of the infringer's profits was another available remedy, but Congress eliminated it

manufacturing patentholders who would have made sales—and associated profits—but for the infringement.[119] When the patentee does not practice its patent, or lost profits cannot be proved, reasonable royalties are provided.[120] To prevent *post*-suit infringement, prior to the Supreme Court's decision in *eBay*,[121] the standard approach was to grant an injunction.[122] So while nonpracticing entities licensed their inventions in the marketplace, they did so under the threat of an injunction in the event of infringement.[123]

B. *Removing Real Property: A Partially Revised Model of Patent Law Remedies*

With the rise in suits from nonpracticing entities (NPEs) in the early 2000s, the standard model of patent remedies was impugned by several leading scholars as too often providing "windfalls" to patentees that typically performed no R&D, and certainly did nothing to commercialize their patents.[124] In a widely cited article, *Property, Intellectual Property,*

---

in 1946 because of the perceived difficulty and high costs of determining these profits. *See* 35 U.S.C. § 70 (1940), *amended by* Act of Aug. 1, 1946, ch. 726, 60 Stat. 778 (codified at 35 U.S.C. § 70 (1946)); H.R. Rep. No. 79-1587, at 1–2 (1946). Interestingly, disgorging the infringer's profits would not return the patentee to the *status quo ante* because the patentee's profits in the absence of infringement could be substantially less than the infringer's profits.

119. *See* Panduit Corp. v. Stahlin Bros. Fibre Works, Inc., 575 F.2d 1152, 1156 (6th Cir. 1978) (Markey, C.J.) ("[Damages] have been said to constitute 'the difference between his pecuniary condition after the infringement, and what his condition would have been if the infringement had not occurred.'") (quoting Aro Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 507 (1964) (quoting Yale Lock Mfg. Co. v. Sargent, 117 U.S. 536, 552 (1886))).

120. *See id.* at 1157 ("When actual damages, e.g., lost profits, cannot be proved, the patent owner is entitled to a reasonable royalty."); *see also* Lemley, *supra* note 6, at 671 ("Patentees whose harm is based on a loss of market exclusivity—those who reasonably could have expected to make additional sales, or sales at a higher price, absent infringement—should be entitled to lost profits damages. Patentees whose harm is lost licensing revenue, but who could not plausibly claim to have lost sales as a result of the infringement, should be entitled to reasonable royalties . . . .").

121. *See* eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388, 394 (2006) (holding that injunctions are not mandatory but should instead be awarded on the basis of a set of equitable factors).

122. *See id.* at 394–95 (Roberts, C.J., concurring) (noting the historical practice of granting injunctions in most patent cases); Lemley & Shapiro, *supra* note 19; *cf.* Smith Int'l., Inc. v. Hughes Tool Co., 718 F.2d 1573, 1577–78 (Fed. Cir. 1983) (stating that without the ability to issue injunctions, the "right to exclude granted by the patent would be diminished, and the express purpose of the Constitution and Congress, to promote the progress of the useful arts, would be seriously undermined").

123. *See supra* notes 100–03 and accompanying text.

124. *See, e.g.*, Amy L. Landers, *Let the Games Begin: Incentives to Innovation in the New Economy of Intellectual Property Law*, 46 SANTA CLARA L. REV. 307, 308–09 (2006) ("At their worst, patent trolls threaten companies with baseless lawsuits that seek product shutdowns and large monetary judgments."); Lemley & Shapiro, *supra* note 19, at 2009 ("These companies are paying holdup money to avoid the threat of infringement. That is not a legitimate part of the value of a patent; it is a windfall to the patent owner that comes at the expense not of unscrupulous copyists but of legitimate companies doing their own R&D.").

and Free Riding, Mark Lemley derides the real property foundation implicit in patent law's exclusion principle.[125] Specifically, Lemley begins his critique by returning to the fundamental goal of patent law: "to give as little protection as possible consistent with encouraging innovation."[126] In this vein, he argues that "[t]he absolute protection or full-value view draws significant intellectual support from the idea that intellectual property is simply a species of real property rather than a unique form of legal protection designed to deal with public goods problems."[127] In particular, Lemley rejects the view that a patent owner should be able to internalize the full social value of his invention because an inventor need not capture this entire value in order to be optimally incentivized to invent.[128] Rather, sufficient incentive will be present as long as the inventor earns a suitable return on its investment in R&D.[129]

Yet, oddly, when it comes to specific remedies, Lemley essentially limits the implications of his critique to those entities that non-exclusively license their patents, such as nonpracticing entities.[130] For instance, he concludes that "if the baseline assumption of the law is that the intellectual property owner is entitled to capture the full social value of the invention, it is that baseline that will drive any *licensing* negotiations."[131] He advocates routinely denying injunctions to nonpracticing entities in situations involving small components of complex products (so-called "component patents") and high switching costs, but he does not do so for practicing entities, instead suggesting that injunctive relief is generally appropriate.[132] However, as I argue below, Lemley's same baseline assumption and related critique hold true for the value the manufacturing patentee is entitled to capture because third parties will gauge the financial cost of infringing based upon the expected damages they will pay in court and the related costs of complying with an injunction.[133]

---

125. Lemley, *supra* note 25, at 1031–32.

126. *Id.* at 1031.

127. *Id.* at 1031–32.

128. *See id.* at 1065 ("Economic theory tells us that we must balance those rights if we are to achieve efficiency, granting intellectual property rights only to the extent necessary to enable creators to cover their average fixed costs. Anything more does harm and no good." (footnotes omitted)).

129. *Id.* at 1046 ("In a market economy, we care only that producers make enough return to cover their costs, including a reasonable profit.").

130. *See id.* at 1045–46.

131. *Id.* at 1046 (emphasis added).

132. *See* Mark A. Lemley, *Patenting Nanotechnology*, 58 STAN. L. REV. 601, 630 (2005) ("The ability of patent owners to . . . threaten injunctions against whole products even if a small component is found to infringe has created a thriving economy of patent 'trolls.' These patent trolls game the system in an effort to capture not only the value of their inventions, but the value of complementary assets and irreversible investments made by defendants as well.").

133. *See infra* section II(C)(1).

In a later and highly influential article by Lemley and Carl Shapiro, *Patent Holdup and Royalty Stacking*, these conceptual limitations are even more apparent. Lemley and Shapiro begin their analysis with a convincing argument that the granting of injunctions to nonpracticing entities on component patents can provide a windfall to the patentee that distorts optimal innovation incentives.[134] Specifically, Lemley and Shapiro determine a "benchmark royalty," which represents the optimal level of licensing fees that a nonpracticing licensor would earn in the "ideal patent system."[135] Although there is no need here to recount all of the mathematical details of how the benchmark royalty is precisely calculated to illustrate my point, importantly, the royalty is a fraction of the economic value of the invention to the licensee (though one might suitably recharacterize it as based on the value of the invention to society).[136]

Unlike Lemley's earlier work, Lemley and Shapiro do not posit that the benchmark royalty should be the licensee's pro rata portion of the *minimum* amount necessary to induce invention on the part of the patentee.[137] However, because the benchmark royalty is generally a fraction of the full economic value of the invention to the licensee, we would expect it to roughly approximate such an amount. Conversely, if the patentee were able to garner substantially *more than* the full economic value of the invention to the licensee, then we would expect that this amount would generally be far greater than that needed to induce invention. Thus, although Lemley and Shapiro do not reject the make-whole damages assumption per se, their theoretical analysis showing that injunctions can sometimes produce gross windfalls for the patentee implicitly supports the thesis that injunctions can thwart optimal innovation incentives.

Suppose, for instance,[138] that a patent allows a manufacturer to cut its production costs by $1 per product sold by reducing the costs of

---

134. *See* Lemley & Shapiro, *supra* note 19, at 2009 ("These companies are paying holdup money to avoid the threat of infringement. That is not a legitimate part of the value of a patent; it is a windfall to the patent owner that comes at the expense not of unscrupulous copyists but of legitimate companies doing their own R&D."). The same kind of analysis applies to patent claims covering an entire complex product, the only novel elements of which are components of the product. *See* Mark A. Lemley, Essay, *Point of Novelty*, 105 NW. U. L. REV. 1253, 1260 (2011) ("[W]e don't distinguish between the novel elements of the claim and the preexisting ones.").

135. Lemley & Shapiro, *supra* note 19, at 1999.

136. *See id.* (using the value of the use of the patented invention to the downstream firm (licensee) as the base for determining the benchmark royalty rate).

137. *See id.* at 1999–2000.

138. Here, I provide several numerical examples to support my case. In so doing, I do not intend to show that the traditional analysis is incorrect in *every* case; rather, I do so to show that it is incomplete in a subset of cases that are likely to occur in practice. Moreover, although I present numerical examples, I follow Lemley and Shapiro's formal algebraic model, which applies in a wide set of circumstances. *See id.* at 1996–2005.

manufacturing a particular component in the product.[139] In this event, Lemley and Shapiro's benchmark royalty is some fraction of the $1.[140] For simplicity, if we suppose the licensee is the only potential producer of the product,[141] and it sells one million products per year, the total economic value produced by the innovation is $1 million per year. Also suppose that prior to the inventive technology, the licensee sold the product for $100 (resulting in $100 million in annual sales), whereby each sale returned $9 in cash profits (resulting in $9 million in annual profits). This profit may derive, for example, from the licensee's patents, know-how, marketing, or other competitive advantages.[142] Thus, by using the patented process, the licensee can increase its profits to $10 million per year, less any payments to the patentee.

In this example, it is relatively straightforward to see how a mandatory injunction can systematically skew licensing rates or settlement fees. Suppose that the licensee-seller independently invents the cost-saving technology, implements it, and is then sued by the patentee for infringement.[143] If the patentee can shut down all sales of the infringing

---

139. Importantly, standards-essential patents—those patents that must be practiced in order to implement an industry-wide standard—often cover small components of complex products. As Suzanne Michel has noted:

> Recent, very large damage awards, some of which have been based on one feature of a complex product covered by hundreds of patents and increases in IT industry patent litigation raise questions of whether damages law is sufficiently economically grounded, at least in that subset of cases in which the invention is one component of a complex IT product. Of course, these are the types of cases most likely to involve a patent covering a standard.

Suzanne Michel, *Bargaining for RAND Royalties in the Shadow of Patent Remedies Law*, 77 ANTITRUST L.J. 889, 899 (2011).

140. *See* Lemley & Shapiro, *supra* note 19, at 1999. Specifically, they find that the benchmark royalty rate, assuming an ideal patent system and parties of equal bargaining power, would be half the gains from using the invention adjusted by the strength of the patent. *Id.* Thus, if there is an 80% chance the patent would be found valid, enforceable, and infringed, and the value of using the patented invention is $1, the benchmark royalty would be $0.40 (half of $1 × 80% probability of infringement). *See id.*

141. For instance, the licensee may own patents covering other components of the product that prevent its manufacture by third parties, including the licensor. *See* Sichelman, *supra* note 35, at 354 ("[M]any patented products contain components that infringe earlier patents."); Robert P. Merges, *Intellectual Property Rights and the New Institutional Economics*, 53 VAND. L. REV. 1857, 1859 (2000) ("Complex, multi-component products are the norm in many industries (e.g., autos and consumer electronics), and individual patents often cover only a single component or sub-component.").

142. *See* Lemley, *supra* note 6, at 663 ("Even if there are no other relevant patents, the defendant's know-how, materials, and marketing efforts almost always contribute some value, and usually the most significant part of the value of an infringing product."); Brian J. Love, Note, *Patentee Overcompensation and the Entire Market Value Rule*, 60 STAN. L. REV. 263, 278 (2007).

143. Recent empirical work indicates that infringers tend not to "copy" from the patentee in the course of infringement. *See* Christopher A. Cotropia & Mark A. Lemley, *Copying in Patent Law*, 87 N.C. L. REV. 1421, 1424 (2009) (finding that "copying" is proven by accused infringers in less than 2% of all cases and alleged in less than 11%). These figures very likely understate the

product, which it typically can—because the product was made using the infringing production method and the resulting component[144]—the seller is faced with two scenarios.[145] First, the seller may be able to profitably redesign the product in a noninfringing manner—for example, by returning to the prior production technique and component. If the redesign is costly—because, for instance, it requires significant retooling of manufacturing processes—say $2 million amortized over each year of product sales (which includes lost sales of products during the period of redesign), the seller will negotiate a license rate at far more than the optimal one of some fraction of $1 per product. In particular, the seller reasons that if it redesigns, then it stands to earn $7 million per year in profit (the original $9 million per year in profits less the $2 million per year in retooling costs), whereas if it does nothing, it could lose all profits if it is enjoined from using the patented method. The seller also reasons that if it continues to use the patented production method, it will stand to earn $10 million per year (the original $9 million plus the $1 million in added savings) less any licensing or settlement fees.

If the patent at issue is "ironclad"—meaning it would certainly be found infringed, and not invalid or unenforceable[146]—and the parties are of

---

amount of copying present in the marketplace. First, copying probably more often results from the copying of others' products, rather than patents, and mere product copying is unlikely to find its way into the kinds of litigation documents Cotropia and Lemley examined. *See* Golden, *supra* note 65, at 589 n.456 (providing an example of such). Second, based on my personal experience, documentary evidence of direct copying is usually scant, and litigators tend to do an outstanding job in preparing witnesses to avoid disclosing any copying in depositions. Nonetheless, there are surely many cases of independent invention. Moreover, as I argue below, the analysis here can, under the right conditions, apply in the event the accused infringer copied the patented technology. So, although I do not agree that lack of copying inherently implies independent invention, it is likely to do so in a significant number of cases. Moreover, I argue below that injunctions may distort innovation incentives even in the presence of copying. *See infra* note 149.

144. The patentee can typically draft patent claims covering not only the inventive production process but also the resulting component and even the entire product, offering numerous routes to proving infringement. *See* Landers, *supra* note 124, at 360 ("[P]atentees drafting improvement claims may be encouraged to include additional components in combination claims to sweep additional products within their scope. For example, a patentee who claims an improvement to one component of a computer networking system can mention other components of the system in the language of the claim to later support an argument that the royalty base should include the entire system."). Nonetheless, even if only the component is covered by the applicable patent claim, an injunction over the component will in many cases effectively result in an injunction over the entire product. *See, e.g.*, Michael A. Carrier, *Innovation for the 21st Century: A Response to Seven Critics*, 61 ALA. L. REV. 597, 619 (2010) (citing *z4 Technologies, Inc. v. Microsoft Corp.* as an example in which "Microsoft infringed [because of] a 'very small component' . . . not related to the 'core functionality' of Windows in a setting in which injunctive relief would have required Microsoft to release new versions of its Windows software in 600 variations in more than 40 languages").

145. I have slightly adapted Lemley and Shapiro's analysis in the following exposition. *See* Lemley & Shapiro, *supra* note 19, at 2001–05.

146. Lemley & Shapiro, *supra* note 19, at 1999 (implying that an "ironclad" patent is one that is clearly infringed, valid, and enforceable).

equal bargaining power, then the licensee will pay half the expected marginal profits from using the patented production technique relative to the old technique.[147] Here, the marginal difference is $3 million—namely, $10 million per year (i.e., profits using the new technique) less $7 million per year (i.e., profits from retooling to the old technique). So, in an equal bargaining power scenario, the patentee secures $1.5 million per year, or $1.50 per product, in royalties. This amount, of course, is more than the entire cost-reducing value of the patented invention. In other words, the threat of an injunction coupled with high switching costs can enable the patentee to extract more than the social value of its invention in rents from the potential user.[148]

The situation is even worse if the seller cannot design around the patent. Suppose, for instance, that the seller had just launched its business, only to learn that one of its key product components was patented.[149] Because the patentee can shut down the seller's operations entirely, assuming the parties are of equal bargaining strength, and the patent is ironclad, the marginal difference will be the potential profits of operating (here, $10 million per year) less the profits of being shut down (of course, nothing—perhaps even negative profits). In this case, the seller must forgo $10 million in potential profit, and licenses the patent for $5 million per year, or $5 per product, in royalties—much more of a distortion from the optimal royalty rate.

In the previous example, only the licensee-seller—but not the patentee—manufactured the relevant product. Yet all of Lemley and

147. *See supra* note 140 and accompanying text.

148. Here, I have assumed that the social value of the patent equals its total cost-saving potential. Of course, the social value of a patent can be greater than its private economic value, in which case we might justify the greater reward to the patentee. *See infra* section II(C)(2). Yet, in the standard make-whole remedies approach, courts assume that the profits or royalties lost by the patentee—i.e., its *private market value* to the patentee—is sufficient to reward the patentee. As a first cut, here I do the same. I relax this assumption below. *See infra* section II(C)(2).

149. For standards-essential patents, it is typically impossible to design around the patent and still practice the standard. *See* Patrick J. Flinn, *Why FRAND Matters*, INTELL. PROP. & TECH. L.J., Feb. 2013, at 8, 11 (explaining that "when the patent is a standards-essential patent . . . the question is now whether to enter the market at all because license to every standards essential patent is needed"). Thus, even in the absence of a FRAND commitment—and in the presence of direct copying by the accused infringer—for the reasons presented herein, the presumption should generally be to deny injunctions to practicing and nonpracticing entities alike for infringement of standards-essential patents covering minor components of complex products. *See* Joseph Scott Miller, *Standard Setting, Patents, and Access Lock-In: RAND Licensing and the Theory of the Firm*, 40 IND. L. REV. 351, 378 (2007) ("If the owner of a standard-essential patent can enjoin (or threaten to enjoin) would-be adopters from practicing the standard, the very enterprise of adopting a standard fails to meet its basic purpose."); *cf.* Colleen V. Chien et al., *RAND Patents and Exclusion Orders: Submission of 19 Economics and Law Professors to the International Trade Commission* 2 (Santa Clara Univ. Sch. of Law Legal Studies Research Paper Series, Paper No. 07-12, 2012), *available at* http://ssrn.com/abstract=2102865 (recommending that injunctions generally be denied in International Trade Commission proceedings on all standards-essential patents subject to a FRAND commitment).

Shapiro's arguments and mathematical modeling apply with similar force to practicing entities that compete directly with the infringer. For instance, if a practicing patentee is granted an injunction, just as with a nonpracticing entity, the infringing firm, as Lemley and Shapiro recognize, "cannot sell the infringing product and must withdraw it from the market unless and until the firm can introduce a redesigned version that does not contain the patented feature, or until the patent expires."[150]

To illustrate the consequences of this observation, take the same patented invention as the earlier example, but now suppose the patentee also sells the product. Suppose prior to the patented invention, the product sells at $98, instead of $100, because the duopolistic market is more competitive than the monopolistic market of the previous example.[151] In this event, the infringer earns $7 per product in profit (as the marginal cost per product prior to the cost-saving invention is $91). Suppose further the infringer sells half the number of products per year (500,000 instead of the original million). Also suppose that the patentee earns the same profit and sells the same number of products per year as the infringer.[152] Like the original example, this duopoly may arise from other patents owned by the parties or from other barriers to entry.[153]

When the patented cost-saving component comes along, if both parties use it—assuming for simplicity that the market is fully competitive at the margins[154]—then the price of the product drops from $98 to $97, resulting in $1 million per year in additional consumer surplus. In the event the patentee is granted injunctive relief, it has two choices. First, it can shut down sales of the competitor's product entirely until the competitor redesigns it.[155] During this period, because the patentee is the only seller in

150. Lemley & Shapiro, *supra* note 19, at 1996.

151. As with the earlier example, this example is generalizable across a host of variables. *See generally* George Symeonidis, *Comparing Cournot and Bertrand Equilibria in a Differentiated Duopoly with Product R&D*, 21 INT'L J. IND. ORG. 39 (2003) (offering a general model of duopolistic behavior in the presence of R&D).

152. Because the firms sell products that are slightly different in characteristics, *see infra* note 154, under a Bertrand model for differentiated products, firms would typically choose to price at above marginal cost. *See* DAVID BESANKO & RONALD BRAEUTIGAM, MICROECONOMICS 563 (4th ed. 2011) ("In a Bertrand equilibrium with differentiated products, equilibrium prices generally exceed marginal cost.").

153. *See* Harold Demsetz, *Barriers to Entry*, 72 AM. ECON. REV. 47, 47–57 (1982) (discussing generally the various kinds of barriers to entry into an economic market); *supra* note 142 and accompanying text.

154. For instance, the $5 per product profit earned by each party may result from differences in product features attributable to their separate patents, leading to market differentiation, whereas the cost-saving feature merely leads both sellers to cut their product prices by $1. *See* Greg Vetter, *Patenting Cryptographic Technology*, 84 CHI.-KENT L. REV. 757, 774–75 (2010) (noting the use of patents for product differentiation).

155. In *Apple Inc. v. Samsung Electronics Co.*, 695 F.3d 1370 (Fed. Cir. 2012), the Federal Circuit denied a manufacturing patentee a preliminary injunction over the accused infringer's product containing a patented component because the patentee failed to show a "causal nexus"

the market, it can sell the product at the monopoly price (here, $100), leading to increased deadweight losses (here, some fraction of $3 million per year) that far exceed the social value of the patented invention. When the competitor redesigns its production process, as in the earlier example, it must spend $2 million per year on an amortized basis, which is more than the value of the patent, and hence a net social loss. Of course, the competitor has a financial incentive to redesign in order to avoid being shut down and earning nothing.

Second, the patentee could decide to license the competitor once it learns of the infringement. In this event, an analysis similar to that recounted in the nonpracticing entity example above applies. Namely, the patentee reasons that the infringer has one of two viable choices: (1) retool and earn $1.5 million per year in profits ($3.5 million in gross profits less the $2 million in amortized retooling costs); or (2) license and earn $3.5 million per year in profits (or perhaps more if the license allows the parties to effectively fix prices) less any licensing fees. If the patent is ironclad and the parties are of equal bargaining power, then the licensor will demand annual royalties of half the difference between $3.5 million or more and $1.5 million (i.e., at least $1 million), which is equal to or more than the social value of the invention.[156]

Thus, just like nonpracticing patentees, following Lemley and Shapiro's otherwise insightful analysis, manufacturing patentees with injunctive threats can—when switching costs are abnormally high—effectively extract or cause systematic deviations from the optimal profits or payment for a given invention, as well as unnecessarily raise consumer deadweight losses if the infringer cannot design around the patent or the design-around requires a long period of time. Lemley and Shapiro, however, provide no significant analysis of the manufacturing entity situation. Instead, they contend without extended discussion that:

> [W]e stress that our analysis in this Article is expressly limited to

---

between the alleged irreparable harm from the accused infringer's sales of the entire product and its infringement of the specific patented component. *Id.* at 1374–75. In other words, because "it may very well be that the accused product would sell almost as well without incorporating the patented feature," the Federal Circuit found preliminary relief unwarranted. *Id.* In a later ruling in the same case, the Federal Circuit held that the causal nexus rule applies to permanent injunctions. Apple Inc. v. Samsung Elecs. Co., 735 F.3d 1352, 1361 (Fed. Cir. 2013). Nonetheless, in the example in the text above, the rule would not prevent the patentee from securing an injunction because the cost-cutting patented feature here would enable the patentee to garner 100% of the sales of the patented product in the event of no infringement. *See Apple*, 695 F.3d at 1375 (stating that the causal nexus requirement is met when the patentee "show[s] that the infringing feature drives consumer demand for the accused product"). Thus, although the "nexus" requirement adopted in the *Apple v. Samsung* line of cases ameliorates some of the deleterious effects of injunctions in the context of multicomponent products, it does not fully rectify overdeterrence problems.

156. *Cf.* Heald, *supra* note 19, at 1184 ("In fact, if the infringing firm's switching costs are high enough, it may be willing to agree to a license that results in a net loss.").

> situations in which the patent holder's predominant commercial interest in bringing a patent infringement case is to obtain licensing revenues. Our policy recommendations here pertain only to this type of situation, where the patent holder can claim reasonable royalties but not lost profits, and not to settings in which the patent holder suffers significant lost profits as a result of the allegedly infringing activities of the downstream firm and seeks to use the patent to exclude a competitor from the market in order to preserve its profit margins. In cases involving significant lost profits, we favor a presumption that the patent holder will be granted a permanent injunction, perhaps with a stay to allow the infringing firm to redesign its product. The presumptive right to a permanent injunction in these cases is justified in part for reasons of equity and in part because of the grave difficulties associated with calculating and awarding lost profits on an ongoing basis.[157]

Lemley and Shapiro's categorical distinction between licensing and manufacturing entities is not economically justified. Specifically, their suggestion of a stay to allow the infringer to redesign the product does not always solve holdup concerns. As I showed earlier, the costs of redesign may exceed the value of the patented component, clearly a net social loss. Additionally, there may be situations for which no redesign is available. In essence, Lemley and Shapiro's resort to "reasons of equity" to justify presumptive injunctions for manufacturing patentees sounds in the same reasoning as that in tort law because it aims to provide an injunction "only when the remedy of compensatory damages will not suffice to restore the status quo ante."[158] But as a matter of pure economics, the simple examples above—which are easily generalizable—show that traditional tort law remedies do not always promote optimal outcomes in the patent law context.

An additional argument commentators have made to distinguish practicing entity suits from NPE suits is that accused infringers can typically assert their own patents back against a practicing patentee in a counterclaim or separate suit, thereby preventing holdup.[159] Although many direct competitors cross-license one another or settle in the face of a potential countersuit, there are numerous disputes in which an accused infringer cannot credibly threaten to shut down a practicing patentee.[160]

---

157. Lemley & Shapiro, *supra* note 19.

158. Fellmeth, *supra* note 4.

159. *See* FTC, *supra* note 15 ("The potential for hold-up to result in mutually assured destruction means firms actively participating in the industry—patent practicing entities (PPEs)—are unlikely to employ this hold-up strategy against each other.").

160. *See, e.g.*, Ted Sichelman, *The Vonage Trilogy: A Case Study in "Patent Bullying"* 7 (San Diego Legal Studies, Paper No. 11-057, 2013), *available at* http://ssrn.com/abstract=

Specifically, the accused infringer may not own any patents that it can assert against the patentee or may simply own far fewer applicable patents than the patentee plaintiff, causing insufficient threat to the patentee.[161]

Another simple hypothetical illustrates why the distinction between nonpracticing and practicing entities is problematic—at least if we take at face value the dominant reward theory of patent law, which justifies awarding exclusionary rights on the basis of allowing inventors to recoup their costs with a suitable profit.[162] In one state of the world, inventor A spends $1 million in R&D, designing, building, and testing a particular invention. The inventor then manufactures the invention and sells one million units at $100 each, earning $10 in profit, of which $1 is attributable to the patented invention (like the earlier example).

Assume, further, that—consistent with the reward theory of patents—manufacturing and commercialization of the product does not incur substantial cost or risk that requires additional patent protection to ensure optimal levels of investment.[163] In this regard, it is important to recognize that Justice Kennedy decidedly did not adopt a commercialization theory of patents in *eBay*—namely, that NPEs should be punished merely because they fail to practice their patents.[164] Rather, his decision stemmed from his view that NPEs gain "undue leverage" from potential injunctive relief in negotiating with infringers, primarily because NPEs only garner money damages in the marketplace and cannot be countersued in an infringement suit.[165] So while commercialization goals might very well play an important role in patent law—and I have advocated as much elsewhere[166]—taking Justice Kennedy and the vast majority of scholars at face value,[167] I assume in this example that commercialization is not an interest of concern.

In the event of infringement—using the same base assumptions as the earlier example—assume further that the patentee's sales drop in half (from

---

1856703 ("Vonage's lack of a patent portfolio (by definition) precluded it from the preferred tactic of 'defending' against suit by asserting its patents back against Sprint.").

161. *See id.*

162. *See* Sichelman, *supra* note 35, at 357–58 (discussing reward theory); Mark A. Lemley, *Ex Ante Versus Ex Post Justifications for Intellectual Property*, 71 U. CHI. L. REV. 129, 129 (2004) ("The traditional economic justification for intellectual property is well known. Ideas are public goods: they can be copied freely and used by anyone who is aware of them without depriving others of their use.").

163. *See* Lemley, *supra* note 162, at 148–49 (contending that at least one *ex post* justification for intellectual property asserts that commercialization of inventions will tend to proceed efficiently following the invention phase).

164. *See supra* notes 11–16 and accompanying text.

165. *See supra* notes 11–16 and accompanying text.

166. *See* Sichelman, *supra* note 35, at 354 (arguing that commercialization efforts typically entail supernormal costs and risks that independently justify some form of *ex post* IP protection).

167. *See id.* at 344 ("The dominant 'reward' theory of patenting, which undergirds much of today's law, perceives little to no need to protect risky and costly post-invention development and commercialization efforts.").

one million to 500,000 units) and its profits drop by more than half (from $10 to $4 per unit), resulting in only $2 million in annual profits (500,000 units × $4 per unit). Under current law, the inventor is entitled to lost profits, and is able to recoup $8 million in annual damages for past infringement, returning it to the $10 million per year in total profits.[168] The net result is that the patentee earns the entire $1 million per year in additional profits attributable to the patented invention. Yet, in another state of the world, imagine the patentee decides to license its invention. The would-be licensee rejects the offer and decides instead to infringe. The patentee wins at court and now collects a reasonable royalty. Even assuming a high royalty of 25% of the additional profits generated by the patented invention,[169] the patentee only collects $250,000 in annualized damages.[170]

So the practicing inventor spends the $1 million on R&D and reaps $1 million per year in benefits, while the nonpracticing inventor spends the same $1 million and collects $250,000 in annual royalties. This disparity has no economic basis in the generally accepted view of patent law to provide optimal incentives to invent, rather than commercialize. At least for this hypothetical, the two inventors should earn *exactly the same return* on their efforts. As John Golden properly recognizes in a response to Shapiro and Lemley's article, "A per se rule of discrimination based on a patent holder's business model could act as an undesirable drag on the efficiency and competitiveness of markets for innovation."[171] In a reply to Golden, Lemley and Shapiro—instead of engaging in economic analysis—

168. For simplicity, suppose that $7 million of the $8 million is awarded for damages relating to other patents. However, as noted below, the patentee is often entitled to damages on the entire market value of a product merely due to a single patented component. Here, the cost-saving component would have put the patentee in a position to make 100% of the sales of the product in the market because it could price below all competitors in the event of no infringement. As such, the patentee could show a nexus between the component and the purchase of the product by consumers, justifying imposition of the entire market value rule. *See infra* note 183 and accompanying text.

169. Under current law, reasonable royalties are typically apportioned to reflect a royalty on the value of the patented component. *See* Lucent Techs., Inc. v. Gateway, Inc., 580 F.3d 1301, 1339 (Fed. Cir. 2009) (holding that "when the patented invention is a small component of a much larger commercial product," the royalty base may be the sale price of the entire product "so long as the [royalty] multiplier accounts for the proportion of the base represented by the infringing component or feature"). So, here, the patentee would only collect 25% of the $1 million in profits attributable to the patented invention.

170. Until the Federal Circuit's recent decision in *Uniloc USA, Inc. v. Microsoft Corp.*, 632 F.3d 1292 (Fed. Cir. 2011), patentees were generally entitled to 25% of the accused infringer's profits as a "rule of thumb" in reasonable royalty determinations, making that figure today more of a ceiling than a floor. *See id.* at 1312–18 (holding that evidence relying on the 25% rule of thumb is generally inadmissible).

171. John M. Golden, Commentary, *"Patent Trolls" and Patent Remedies*, 85 TEXAS L. REV. 2111, 2117 (2007); *see also* Golden, *supra* note 65, at 512 (advocating for an "antidiscrimination" principle in patent law remedies).

again fall back on the flawed conceptual tort law model of patent damages: "There is no unjustified discrimination here; in both cases, the goal of patent remedies is properly to ensure that patent owners are compensated for any unauthorized uses made by others."[172]

In their reply, Lemley and Shapiro further contend that:

> While injunctions in lost-profits cases may also cause holdup problems, we do not see a practical and general way of avoiding these problems while providing adequate compensation to patent holders for infringement. There is a fundamental difference between cases involving reasonable royalties and those involving lost profits. In reasonable-royalty cases, the joint profits of the patent holder and the infringing firm are *increased* by the infringing firm's use of the patented invention, so the hypothetical ex ante negotiation concept makes good economic sense. In lost-profits cases, the joint profits of the patent holder and the infringing firm may well be *decreased* by the infringing firm's use of the patented invention. In such cases, the hypothetical ex ante negotiations concept is not helpful for assessing damages and it is not possible to find licensing terms that are commercially acceptable to the downstream firm and adequately compensate the patent holder.[173]

Although Lemley and Shapiro are correct that in lost-profits cases, joint profits can be decreased by the infringing firm's use, one need not rely on a hypothetical *ex ante* negotiation to calculate the proper level of ongoing damages. Rather, as in the simple numerical examples presented earlier, the question becomes what measure of monetary award would properly reward the patentee's invention in an incentive-based model. Because, as Lemley has recognized elsewhere, a patentee need not reap the entire social value of its invention in order to be sufficiently incentivized, it is not necessary to imagine a hypothetical set of "licensing terms that are commercially acceptable . . . and [that] adequately compensate the patent holder."[174] Rather, the court merely needs to award the patentee some portion of the available profit stemming from infringement adequate to *incentivize the invention.*[175]

---

172. Lemley & Shapiro, Reply, *Patent Holdup and Royalty Stacking*, 85 TEXAS L. REV. 2163, 2171 (2007). In a later article Lemley argues that compensation for nonpracticing entities "should be based on the value that the patented invention actually contributes as a proportion of the defendant's product, taking into account the other patents, know-how, raw materials, and labor that also contribute to the value of that product and the existence of possible alternatives to the patented technology." Lemley, *supra* note 6, at 670. This more nuanced view of compensation should be applied with equal force to practicing entities.

173. Lemley & Shapiro, *supra* note 172, at 2173 n.49 (emphasis omitted).

174. *Id.*

175. Below, I advocate also considering incentives to commercialize. *See infra* notes 196–200, 240 and accompanying text.

Last, returning to their original article, Lemley and Shapiro try to escape the debate by conjecturing that ongoing lost profits are difficult to calculate.[176] Yet, the same kinds of difficulties in calculating and awarding lost profits on an ongoing basis beset calculating ongoing reasonable royalties.[177] Indeed, in one of his later articles, Lemley argues that calculating lost profits is generally more straightforward than calculating reasonable royalties:

> But where the problem is imprecision in calculating lost profits, it is important to keep in mind that the alternative to denying lost profits is a less precise, and more distorted, reasonable royalty measure. . . . But under a strict divide approach, a patentee who can show that it is more likely than not that an infringer's sales cut into its own should be entitled to the court's best estimate of the patentee's lost profits. That estimate may not be perfect, but it is likely to be at least as accurate as the alternative reasonable royalty measure . . . .
>
> Indeed, it is somewhat ironic that courts have insisted on strict compliance with the elements of proof of a lost profits claim, given that the reasonable royalty alternative involves at least as much uncertainty and approximation.[178]

Like ongoing reasonable royalties, courts could assume that ongoing lost profits generally mimic historical profits, subject to adjustment for increased or decreased future sales.[179] As such, Lemley and Shapiro's claim that ongoing lost profits are too difficult to calculate is dubious. In any event, patent law should not aim to award a manufacturing patentee lost profits on an ongoing basis. Rather, the core question is what level of damages should be awarded to sufficiently induce the patentee to engage in

---

176. Lemley & Shapiro, *supra* note 19 ("The presumptive right to a permanent injunction in these cases is justified in part for reasons of equity and in part because of the grave difficulties associated with calculating and awarding lost profits on an ongoing basis.").

177. *See* Paul M. Janicke, *Implementing the "Adequate Remedy at Law" for Ongoing Patent Infringement After* eBay v. MercExchange, 51 IDEA 163, 166–67 (2011) ("[I]n the course of straying away from damages precedents and into the realm of judicially forced periodic payouts, the courts have brought about a number of difficulties for themselves, for the litigants, and for the patent system. . . . The mandatory ongoing royalty approach thus weakens the patent system as a whole, at a time when it is viewed, at least by some observers, as critical to the nation's economic wellbeing."); Stephen M. Ullmer, Note, Paice *Yourselves: A Basic Framework for Ongoing Royalty Determinations in Patent Law*, 24 BERKELEY TECH. L.J. 75, 98 (2009) ("The Federal Circuit's decisions in *Paice* and *Amado* show the uncertainty involved in an ongoing royalty determination. Though the court in each case acknowledged the potential validity of an ongoing royalty, neither decision set forth a comprehensive framework for the lower courts to follow.").

178. Lemley, *supra* note 6, at 672 & n.79 (footnote omitted).

179. *See* Mark A. Lemley, *The Ongoing Confusion over Ongoing Royalties*, 76 MO. L. REV. 695, 701–02 (2011) ("[A] court setting an ongoing royalty after a finding of infringement is not writing on a blank slate. A jury has already set a reasonable royalty for past damages."); Ullmer, *supra* note 177, at 95 ("The reasonable royalty should act as a baseline value for the courts, which can then alter the reasonable royalty in accordance with the other essential *Georgia-Pacific* considerations.").

innovation. On this ground, as I explain more fully below, courts need not look solely to tort-based compensation—that is, lost profits or reasonable remedies—in fashioning an appropriate measure of damages.[180]

In sum, the premise of promoting the *status quo ante* should be jettisoned as a fundamental tenet of patent law remedies. Neither nonpracticing nor manufacturing entities should routinely be entitled to an injunction as a matter of equity, other than in situations for which the costs of determining damages (including error costs from wrong determinations) exceed the costs of granting the injunction (an issue I address in more detail in the next subpart).[181] Indeed, on the same reasoning, patentees should *not* conclusively be entitled to damages sufficient to fully compensate for the "harm" inflicted on the patentee by the infringement. Instead, patentees should only be entitled to the level of damages that promotes optimal innovation incentives, which—as I explain in the next subpart—may often be less, and sometimes more, than the profits or royalties that would have been earned in the marketplace but for the infringement.

C. *Fully Rejecting Tort Law Remedies in Patent Law*

In this subpart, I extend my argument that tort law principles should be excised from patent law remedies. As I recounted earlier, tort law concepts immediately lead to a distinction in the types of remedies that are awarded to manufacturing and nonpracticing entities that is not always economically justified. In this regard, tort law concepts have caused courts and commentators to improperly overlook holdup problems caused when injunctions are granted to practicing entities asserting patents on components of complex products and switching costs are high. In this subpart, I contend that injunctions and full compensatory damages may also thwart optimal innovation incentives when they generate large consumer deadweight losses, result in substantial duplicated costs during the pre-invention R&D process, or create transaction costs far in excess of the value of the invention. On the other hand, patentees may be systematically undercompensated by mere make-whole damages when infringement is difficult to detect or when private incentives to innovate from make-whole damages are insufficient to generate optimal levels of social benefits from the innovation.[182]

---

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

181. *See* Golden, *supra* note 65, at 512 ("[D]etailed tailoring of patent remedies must have limits and is only justified insofar as it promises benefits that justify its costs.").

182. In what follows, I limit my discussion to standard modes of injunctive and monetary relief—namely, granting (or denying) injunctive relief and awarding (or not) some measure of monetary damages to the patentee. In a recent article, Dan Burk offers an original, nearly dizzying, array of nonstandard remedies that revolve around various "real options" that the court can award either to the patentee or the accused infringer. Dan Burk, *Intellectual Property in the Cathedral*, *in* ACCESS TO INFORMATION AND KNOWLEDGE: 21ST CENTURY CHALLENGES IN

1. Correcting Patentee Overcompensation.—In the last subpart, I argued that injunctions involving patents covering components of complex products should typically be denied—at least when switching costs are high—not only to nonpracticing entities, as commentators have recognized, but also to practicing entities. In this section, I examine other areas in which make-whole damages may overcompensate patentees.

First, for reasons similar to those presented earlier for denying injunctions for component patents, overall damages for practicing entities should not always reflect the entirety of the patentee's lost profits attributable to the infringement. Suppose the patentee sells a product for $100 with $2 of marginal profit per product, which reflects the minimum profit necessary for competitors to stay in the product market. In other words, competitors could earn slightly less than 2% profit margins by making other products. Again, suppose that the invention allows the patentee to reduce its costs of manufacturing a component of the product by $1. If competitors cannot use the patented cost-saving method or a suitable design-around with similar cost-savings, the patentee could lower its price to just below $100, potentially forcing its competitors out of the product market.

In the event of infringement by a competitor, under the "entire market value" rule of current law, the patentee would be entitled to full lost profits because the competitors' customers would have bought the patentee's product in the absence of infringement precisely because of its patented production method and resulting cheaper component, which is the only reason for the lower price of the product.[183] If nine competitors each sell

---

INTELLECTUAL PROPERTY AND KNOWLEDGE GOVERNANCE 95 (Dani Beldiman ed., 2013). For instance, Burk describes a sort of "put" rule under which "the alleged infringer continues to use the intellectual property, but can choose to stop infringing and receive damages from the intellectual property owner." *Id.* at 106. Such a rule potentially presents a more beneficial alternative than merely denying injunctive relief or awarding less than lost profits. However, in order to simplify the exposition here, I leave such considerations for future work.

183. For example, the Federal Circuit upheld a finding of lost profits on patents reducing the cost of producing VCRs under the entire market value rule because "[t]he evidence at trial portrayed general industry demand for smaller, cheaper, faster, and more reliable VCRs, and [the patentee] presented evidence that the patented technology furthers these goals." Funai Elec. Co., v. Daewoo Elecs. Corp., 616 F.3d 1357, 1362, 1375–76 (Fed. Cir. 2010) ("The patented inventions are described as improvements that lower the cost of producing VCRs while maintaining product quality."); *see also* Garretson v. Clark, 111 U.S. 120, 121 (1884) ("The patentee . . . must show . . . that the profits and damages are to be calculated on the whole machine, for the reason that the entire value of the whole machine, as a marketable article, is properly and legally attributable to the patented feature." (internal quotation marks omitted)); LaserDynamics, Inc. v. Quanta Computer, Inc., 694 F.3d 51, 67 (Fed. Cir. 2012) ("If it can be shown that the patented feature drives the demand for an entire multi-component product, a patentee may be awarded damages as a percentage of revenues or profits attributable to the entire product."); TWM Mfg. Co. v. Dura Corp., 789 F.2d 895, 901 (Fed. Cir. 1986) ("The entire market value rule allows for the recovery of damages based on the value of an entire apparatus containing several features, when the feature patented constitutes the basis for customer demand.").

100,000 products per year prior to the patented invention, assuming for simplicity that all competitors infringed, the patentee would garner just less than $300,000 per year (that is, just less than $3 per product, the original $2 in profit plus the extra $1 afforded by the patent) in damages from each infringing competitor, even though the value of the infringing component was $100,000 per year (that is, $1 per product) for each competitor. This excessive level of damages, which is triple the social value of the invention, is only justifiable on tort law grounds as a matter of compensation for the "harm" caused by infringement. Like the arguments in favor of apportionment for reasonable royalty damages,[184] full lost profits will often be inappropriate for patents merely covering components of more complex products.

Second, in many contexts besides component patents, the current system of patent remedies can generate excessive innovation incentives. This is especially so because R&D and commercialization costs and risks are substantially more in some industries than others. For instance, the costs of invention in the software industry are usually far below those in the pharmaceutical industry.[185] Yet, the duration of software and pharmaceutical patents are exactly the same (in practice, software patents last longer), and the effective scope of software patents often exceeds the scope of pharmaceutical patents.[186] If the broad scope and long duration afforded software patents is unnecessary to fully incentivize innovation in that industry, then the rewards provided by the patent system are excessive.[187] This is so even if the social value of an invention far exceeds the R&D costs necessary to produce it because society should only pay the inventor the minimum amount necessary to generate the innovation.[188] As such, paying full damages in situations where R&D and commercialization costs and risks are low creates windfalls for innovators.[189]

---

184. *See* Lemley & Shapiro, *supra* note 19, at 2023–24 (suggesting that reasonable royalty damages should be solely based on the value added by the patented component and that total damages should be accordingly apportioned).

185. *See* Burk & Lemley, *supra* note 26, at 1581–83 (noting that R&D costs for software are considerably lower than those for pharmaceuticals).

186. *See* Morris, *supra* note 27 (noting that the effective patent term for pharmaceuticals, even with patent term restoration under the Hatch-Waxman Act, is fourteen years while the effective term for nonpharmaceuticals patents is eighteen and a half years).

187. *See* Burk & Lemley, *supra* note 26, at 1687 ("While most biotechnological and chemical inventions require broad patent protection because of their high cost and uncertain development process, the opposite is true in the case of software development. Software inventions tend to have a quick, cheap, and fairly straightforward post-invention development cycle.").

188. *See* Lemley, *supra* note 25, at 1031 (arguing that the fundamental goal of patent law is "to give as little protection as possible consistent with encouraging innovation").

189. *See* Cohen & Noll, *supra* note 29, at 458 ("[I]f IP protection is too strong, it can reduce economic welfare by inhibiting some kinds of technological change and excessively encouraging other types relative to the progress that would be forthcoming with weaker IP protection."); *cf.* Tun-Jen Chiang, *A Cost-Benefit Approach to Patent Obviousness*, 82 ST. JOHN'S L. REV. 39, 73

Additionally, awarding injunctive relief and full damages can create deadweight losses that stem from (1) the patentee's ability to price its patented goods far over the competitive price;[190] and (2) excessive rents generated from duplicated R&D in "racing" for too large a patent prize.[191] Thus, when R&D costs and risks are low and deadweight losses and rent-seeking costs are high, injunctions should be routinely denied and damages reduced when the potential error costs of doing so are low.[192] Indeed, Ian Ayres and Paul Klemperer[193]—as well as one of my papers[194]—have shown that allowing for imperfect enforcement of the patent can result in superior welfare outcomes to full enforcement, particularly under these conditions.[195]

Third, in commercial situations involving large transaction costs in negotiating licenses, traditional injunctive relief may diminish the commercialization and improvement of patented inventions.[196] Specifically, when a patentee and a third-party commercializer or improver substantially differ over whether a given patent is infringed, valid, or enforceable, it may be economically efficient for the third party to forgo large transaction costs in negotiating a license, instead infringing. Like the theory of efficient breach in contract law,[197] efficient infringement occurs when the transaction costs of negotiation dwarf the value of the innovation at issue, which can occur when there is large uncertainty in the underlying patent rights,[198] or simply when the economic value of the innovation is

---

(2008) ("[P]atents may create such enormous incentives that research funds are diverted from other places where they could have been more socially efficiently invested . . . .").

190. *See* Lemley, *supra* note 25, at 1059 ("By definition . . . the intellectual property system permits owners to raise price above marginal cost, creating deadweight losses by raising the price to consumers.").

191. *See* Grady & Alexander, *supra* note 31 ("The defect in the system is that if multiple inventors expend resources in competition for the patent monopoly, the benefit to society of having the invention will be dissipated by the cost of numerous, redundant, development efforts.").

192. *Cf.* Lederman & Sichelman, *supra* note 42, at 1685 (proposing a measured enforcement approach whereby a tax enforcer intentionally engages in imperfect enforcement so as to reduce deadweight losses).

193. *See* Ayres & Klemperer, *supra* note 46, at 993.

194. *See* Sichelman, *supra* note 47.

195. Reducing overall damages can be achieved not only through reductions on damages for sales of products covered by the patent but also by limiting damages on "convoyed" sales of related products. *See* Lemley, *supra* note 6, at 660 (discussing convoyed sales jurisprudence).

196. *See* Sichelman, *supra* note 35, at 384 ("Like the ex ante licensing context . . . strong, property-like rights will exacerbate the effects of strategic, ex post litigation by increasing the risk of infringement, making the damage award or injunction value to the patentee inefficiently high relative to the social value of commercialization.").

197. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 119–21 (7th ed. 2007); Macneil, *supra* note 33.

198. *See* Depoorter, *supra* note 8, at 67–70 (describing how "innovation uncertainty" can frustrate licensing deals).

fairly minimal.[199] In these circumstances, it may be more optimal to preclude injunctions, even for practicing patentees.[200]

2. Correcting Patentee Undercompensation.—Eliminating the presumption of tort law that a patentee should be made whole need not merely lead to eliminating injunctions and reducing overall compensation. Here, I describe two situations in which the patentee may need *more than* injunctive relief and compensatory damages to generate optimal innovation incentives.

First, infringement is often hard to detect.[201] For instance, if a competitor uses a patented process in a manufacturing process, it may be very difficult, even impossible, to determine whether a given product was made by the process.[202] Even in the make-whole realm of tort law, when violations are not easily detected, standard law-and-economics analysis mandates that for optimal deterrence, damages must be multiplied by the inverse of the probability of detection.[203] Thus, if detection of infringement is difficult, damages should be enhanced to provide optimal innovation incentives.[204] As Roger Blair and Thomas Cotter have recognized, "an award that merely renders the infringer no better off as a result of the infringement may be an ineffective deterrent, because only a portion of all possible infringements are susceptible of detection. This insight suggests that a substantial damages multiplier often may be necessary to achieve adequate deterrence."[205] Because detection may be difficult irrespective of

---

199. *Cf.* Jeffrey L. Harrison, *A Positive Externalities Approach to Copyright Law: Theory and Application*, 13 J. INTELL. PROP. L. 1, 49 (2005) ("When transaction costs are high and the fair user attributes greater value to the use of the work than the loss suffered by the original author, it can be efficient to permit the fair use.").

200. *Cf.* Sichelman, *supra* note 35 (advocating that reducing effective remedies for the patentee may be efficient when the patentee fails to commercialize the patented invention); Katherine J. Strandburg, *Patent Fair Use 2.0*, 1 U.C. IRVINE L. REV. 265, 299–300 (2011) (proposing a fair use defense when, *inter alia*, transaction costs prevent a failure to license the patented technology); Julie S. Turner, *The Nonmanufacturing Patent Owner: Toward A Theory Of Efficient Infringement*, 86 CALIF. L. REV. 179, 204 (1998) (advocating "efficient infringement[]" solely for "unused" patents).

201. *See* Stuart J.H. Graham & Ted Sichelman, *Why Do Start-Ups Patent?*, 23 BERKELEY TECH. L.J. 1063, 1086 n.101 (2008) ("[T]he ability to detect infringement in the first instance . . . may be very costly itself . . . .").

202. *See, e.g.*, Jochen Pagenberg, *The WIPO Patent Harmonization Treaty*, 19 AIPLA Q.J. 1, 14 (1991) (noting the difficulty of detecting process patent infringement).

203. *See* Richard Craswell, *Deterrence and Damages: The Multiplier Principle and Its Alternatives*, 97 MICH. L. REV. 2185, 2186 (1999) ("[T]he ideal penalty (insofar as deterrence is concerned) equals the harm caused by the violation multiplied by one over the probability of punishment."). *See generally* Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968) (describing the theory of punishment multipliers).

204. *See* Heald, *supra* note 19, at 1182 ("[O]ne can justify an award of augmented damages in order to account for the likelihood that the infringement would go undetected.").

205. *See* Blair & Cotter, *supra* note 37.

the infringer's behavior, this condition applies regardless of whether the infringer was willful, unless the willfulness somehow increases the odds of nondetection.

A potential counterargument is that courts will themselves have difficulty determining when infringement is difficult to detect. One possible administrative solution is for the courts to assess whether the accused product or method was sold or used in public. If not, then courts should presume, rebuttably, that the accused product or method is difficult to detect. Assuming the presumption holds, then the question becomes, "How difficult?" This is an even thornier inquiry. Interestingly, as Paul Heald has recognized, the oft-discredited "disgorgement" remedy—whereby the accused infringer is required to pay no less than the profits it earned from infringement—may become an effective remedy because it achieves some multiplier on ordinary damages, yet is much less costly to administer than an exact-multiplier remedy.[206] Of course, it would be ideal for courts to fashion a low cost remedy that roughly determined the appropriate multiplier.

Second, make-whole damages may sometimes be too low to generate sufficient positive incentives for any private actor to invest in R&D and commercialization for a given invention, but the net social welfare of the invention may be sufficiently high that society desires the invention to be made and commercialized.[207] Alternatively, the invention may be justified purely on distributive grounds, such as drugs for rare diseases that are not otherwise profitable.[208] In these situations, infringers, the government, or some other entity would need to supplement make-whole damages with additional funds.

Presumably, other than the possibility of disgorgement of profits, infringers should not pay these additional amounts because assessing such liability would lead to allocative distortions in which potential infringers take excessive caution.[209] As I explained earlier, in many circumstances, infringement can be a kind of efficient breach that diminishes the deadweight losses imposed by patents.[210] In these cases, if we raise the costs of infringement, we lower the benefits from otherwise efficient

206. *See* Heald, *supra* note 19, at 1197–98 ("[I]t suggests the infringer's profits should sometimes be part of a remedy for patent infringement (as it is for trademark and copyright infringement), supporting arguments that it should be reinstated.").

207. *Cf.* Shavell & Ypersele, *supra* note 38 ("[I]ncentives to invest in research are inadequate because monopoly profits are less than the social surplus created by an innovation.").

208. *See* Sichelman, *supra* note 35, at 385–88 (discussing distributive values in patent law).

209. *See* Daniel A. Crane, *Optimizing Private Antitrust Enforcement*, 63 VAND. L. REV. 675, 715 (2010) ("For example, granting an award of treble damages . . . entails a very serious risk of chilling . . . innovation, one of the very things that both antitrust and patent law seek to stimulate.").

210. *See supra* notes 196–200 and accompanying text.

infringement. Rather, a more appropriate approach would be for the government to step in and determine whether current R&D and commercialization incentives, which stem in part from patent remedies, are sufficient for particular kinds of inventive activity. If incentives are too low, the government could—as it does now—subsidize certain categories of inventive activity (e.g., drugs for rare diseases, assistive devices for the disabled, and renewable-energy technologies) both prior to and after invention.[211]

III. Some Possible Costs and Criticisms of Judicially Measured Remedies

In arguing in favor of measuring remedies so as to provide more or less to the patentee than make-whole remedies, I generally ignored several potential costs and criticisms of my proposal. First, as I explained, changing remedies may effectively alter the underlying substance of the patent laws—not only in terms of incentives provided to inventors and commercializers but also to potential infringers. To the extent judges ascribe to themselves a power that may effectively contravene statutory or regulatory law, one might argue that doing so impermissibly treads on constitutional separation of powers principles and related democratic concerns. Second, deviating from the baseline of make-whole damages may increase overall uncertainty, potentially deterring future innovation. Relatedly, juries and judges may not have the ability to accurately award damages and determine injunctive relief other than to make plaintiffs whole. In this event, moving to a system of measured remedies—whereby judges and juries have significant discretion to depart from the tort law baseline—would merely exacerbate present errors, leading to inferior results. Here, I address these concerns in turn.

A. Contravening Democracy and the Separation of Powers

Because the current patent-remedies scheme contemplates making the patentee whole in the event of infringement, attempts to deviate from that baseline—even for awarding reasonable royalties or equitable remedies[212]—would contravene legislative authority.[213] Instead, as I argue

---

211. *See, e.g.*, Orphan Drug Act, Pub. L. No. 97-414, 96 Stat. 2049 (1983) (codified as amended in scattered sections of 21, 26 & 42 U.S.C.) (providing short-term, seven-year market-exclusivity rights similar to patent rights; subsidies for clinical trials; tax incentives; and exemption from certain FDA registration fees for companies developing drugs for rare diseases).

212. Recently, in *Apple v. Motorola*, Judge Posner denied Apple injunctive relief, in part because:

> [I]t would be likely to impose costs on the alleged infringer disproportionate both to the benefits to it of having infringed and to the harm to the victim of infringement, and would thus be a windfall to the patentee and a form of punitive rather than compensatory damages imposed on the infringer. . . . [A]n injunction could force Motorola to remove lucrative products from the market for as long as it took to remove the infringing features—minor features in complex devices most features of

below, modification of the existing statutory framework by Congress would be essential to carrying out my proposal. Even still, would congressionally authorized judicial discretion somehow compromise separation of powers or democratic principles?[214] Specifically, because Congress is slated with "making" the law and the courts with "interpreting" the law, one might argue that judicially "dialing up" or "dialing down" remedies on the basis of the economic factors perverts the notion of substantive law to such a great extent that it unconstitutionally impedes on Congress's ability to legislate.[215] Although such a view may have some force, the same could be said of executive discretion to enforce the law and judicial discretion to fashion remedies in other areas of the law (e.g., criminal law). Any broad

---

> which are not alleged to infringe—from its products, or to invent around the infringing features.

See Apple, Inc. v. Motorola, Inc., 869 F. Supp. 2d 901, 917 (N.D. Ill. 2012). Although Judge Posner's actions are consistent with the normative views presented here, they are unlikely to pass judicial muster on appeal. Specifically, despite Judge Posner's argument that the disproportionate relief that would be enjoyed from an injunction shows the absence of irreparable harm, *see id.*, generally the courts have interpreted irreparable harm in a measure consistent with restoring the *status quo ante*, *see supra* Part I. Because Apple practices the patent-in-issue—and, unlike Motorola, did not make any *ex ante* licensing commitments, *see supra* note 107—the *status quo ante* here would bar the practicing of Apple's patent by Motorola. Thus, under the present case law—which is consistent with Congress's intent—there would be no irreparable harm, despite the economic windfall to Apple. For instance, in *Robert Bosch LLC v. Pylon Manufacturing Corp.*, 659 F.3d 1142 (Fed. Cir. 2011), the Federal Circuit reversed a denial of injunctive relief, relying in part on the facts that the parties were direct competitors and the patentee faced loss of market share. *Id.* at 1152–54. Although the Federal Circuit affirmed a denial of a preliminary injunction in *Apple v. Samsung*, the court found that "Apple ha[d] presented no evidence that directly ties consumer demand for the [infringing product] to its allegedly infringing feature." Apple Inc. v. Samsung Elecs. Co. 695 F.3d 1370, 1375 (Fed. Cir. 2012). Thus, in *Apple v. Samsung*, no injunction was necessary to restore the *status quo ante* because on the record essentially the same number of infringing products would have been sold even if the product had not incorporated the infringing component. *See id.* at 1377 (holding that no irreparable harm occurred because Apple failed to establish a causal nexus between the allegedly infringing features and consumer demand); *see also* Apple Inc. v. Samsung Elecs. Co., 735 F.3d 1352, 1364–65, 1375 (Fed. Cir. 2013) (remanding for further findings regarding customer demand in the context of a permanent injunction determination). Thus, there seems little support in Supreme Court or Federal Circuit case law for Judge Posner's normatively admirable, but judicially questionable, approach.

213. *See, e.g.*, Gen. Motors Corp. v. Devex Corp., 461 U.S. 648, 653–55 (1983) ("When Congress wished to limit an element of recovery in a patent infringement action, it said so explicitly. . . . Congress' overriding purpose [is] affording patent owners complete compensation.").

214. *See generally* Jack M. Beerman, *An Inductive Understanding of Separation of Powers*, 63 ADMIN. L. REV. 467, 472–81 (2011) (describing separation of powers principles).

215. *See* Lederman & Sichelman, *supra* note 42, at 1739 n.268 ("In theory, a measured enforcement regime, particularly in the federal context, may raise concerns that it could override congressional intent, violate separation of powers, or violate due process requirements."); *see also* BLAIR & COTTER, *supra* note 9, at 6 ("More fundamentally . . . this view threatens to undermine the rule of law by substituting judges' idiosyncratic views of the merits of (say) the patent system for the view expressed by Congress in enacting the Patent Act."); *id.* ("We believe that, as a general matter, courts *ought* to operate as if the intellectual property laws embody the proper balance when deciding how to best calculate damages . . . and leave it to other branches of government to decide whether the underlying assumption is true or false.").

discretion can drastically affect the incentives for legal actors to engage in or refrain from particular behavior and, in turn, substantially affect the economic sphere.[216] Discretion in patent law remedymaking seems no different in kind from discretion in other areas of the law, which have existed before and after the time of the Constitution.[217] As such, as long as Congress provides the courts with such discretion, separation-of-powers and democratic concerns seem satisfactorily addressed.[218]

B. The Costs of Uncertainty

Of course, with greater discretion comes greater uncertainty. In general, legal uncertainty can create significant public and private costs in complying with the law and bargaining in its shadow.[219] Yet, uncertainty is relative and the costs it creates must be measured against the costs avoided by affording courts the leeway to fashion economically sensible remedies. Although the current system revolves around a baseline of making patentees whole in the event of the infringement, awarding remedies is anything but an exact science.[220] Deviating from current damages calculations to take into account the economic incentives necessary to motivate invention and innovation—and to avoid duplicated costs and deadweight losses—seems unlikely to generate significant additional uncertainty in the long term. This is especially so if such deviation is limited to cases in which make-whole remedies are clearly economically inferior, as I advocate in the following subpart.

Of greater concern is providing courts even more leeway to deny injunctions in favor of ongoing monetary damages. As I explained, *eBay* eliminated the nearly irrebuttable presumption of injunctive relief.[221] Scott

---

216. *Cf.* Lederman & Sichelman, *supra* note 42, at 1739–40 ("The reality is that agencies already have enforcement discretion.").

217. *See* Emily Sherwin, *Introduction: Property Rules as Remedies*, 106 YALE L.J. 2083, 2085 (1997) ("[T]he great body of property rules are enforced by courts through equitable remedies such as injunctions and specific performance—remedies that historically have been surrounded by a cloud of judicial discretion."); Michael Barone, Jr., Comment, *Delegation and the Destruction of American Liberties: The Affordable Care Act and the Contraception Mandate*, 29 TOURO L. REV. 795, 804 n.72 (2013) ("The Court has a long history of allowing Congress to grant Executive agencies broad discretion in determining if Congress's vague standards have been met.").

218. *See, e.g.*, Mistretta v. United States, 488 U.S. 361, 412 (1989) (affirming judicial discretion in criminal sentencing); Wagner Seed Co. v. Daggett, 800 F.2d 310, 316 (2d Cir. 1986) ("[T]here is no constitutional violation if the imposition of penalties is subject to judicial discretion . . . .").

219. *See* Richard E. Levy & Robert L. Glicksman, *Agency-Specific Precedents*, 89 TEXAS L. REV. 499, 574 (2011) ("For private parties, legal uncertainty increases information costs, requires additional planning, and creates risk. . . . For courts, legal uncertainty leads to litigation and makes settlement more difficult because parties may entertain substantially different assessments of the likely outcome of litigation.").

220. *See supra* notes 176–180 and accompanying text.

221. *See supra* notes 11–19 and accompanying text.

Kieff, Henry Smith, Richard Epstein, and others have stringently criticized this move as introducing substantial uncertainty, as well as increasing potential error costs, thereby decreasing incentives to strike deals outside litigation.[222] Although I agree these costs may be large, the costs avoided in denying economically unjustified injunctions are likely to be even larger in many situations.[223] Additionally, although private parties may sometimes be able to contract around inefficient injunctive remedies,[224] high bargaining costs and information asymmetries will often prevent such deals, leaving a substantial residuum of inefficient cases.[225] Ultimately, the relative balancing of costs in these cases will turn on the ability of judges to

---

222. *See* F. Scott Kieff & Henry E. Smith, *How* Not *to Invent a Patent Crisis*, *in* REACTING TO THE SPENDING SPREE: POLICY CHANGES WE CAN AFFORD 55, 67 (Terry L. Anderson & Richard Sousa eds., 2009) ("[A]n injunction forces the infringer to stop and enforces the delegation of valuation questions to patentees and their contractual partners, with a view toward markets for inputs and products, rather than officials, courts, and experts for hire."); *see also* Richard A. Epstein, *The Disintegration of Intellectual Property? A Classical Liberal Response to a Premature Obituary*, 62 STAN. L. REV. 455, 488 (2010) (arguing that injunctive relief "keeps open the channels for voluntary exchange"); F. Scott Kieff, *Removing Property from Intellectual Property and (Intended?) Pernicious Impacts on Innovation and Competition*, 19 SUP. CT. ECON. REV. 25, 41 (2011) ("A central and underexplored problem with liability rules is that they seriously frustrate the ability for the patentee to attract and hold the constructive attention of a potential contracting party while preserving the option to terminate the negotiations in favor of striking a deal with a different party. This comparative effectiveness of property rules in achieving these goals is the mechanism by which property rules facilitate both innovation and competition."); Sandrik, *supra* note 4, at 136 ("[P]roperty rules in patent law encourage parties to interact and reach a mutually beneficial bargain."); *cf.* BLAIR & COTTER, *supra* note 9, at 40 ("[A] court setting the terms of the exchange would have a difficult time doing so quickly and cheaply, given the specialized nature of the assets and the varied and complex business environments in which the assets are deployed."); Blair & Cotter, *supra* note 19, at 48 (arguing that a property rule "encourage[s] would-be users of the patented invention to negotiate with the patent owner, the theory being that the outcome of these private negotiations . . . more accurately reflects the value of the invention than would a compulsory-licensing fee set by the government"); Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CALIF. L. REV. 1293, 1299 (1996) [hereinafter Merges, *Contracting into Liability Rules*] ("[P]rivate, voluntary organizations . . . would be superior to state-mandated compulsory licenses."); Robert P. Merges, *Of Property Rules, Coase, and Intellectual Property*, 94 COLUM. L. REV. 2655, 2662 (1994) [hereinafter Merges, *Of Property Rules*] (arguing that in a property rule regime, IP owners will have strong incentives to form institutions that are "designed to streamline the exchange of property rights" by modifying "the strong property rule baseline of intellectual property law by *contracting into* liability rules").

223. Even Smith recognizes that in some situations, money damages may be more appropriate. *See* Mark P. Gergen, John M. Golden & Henry E. Smith, *The Supreme Court's Accidental Revolution? The Test for Permanent Injunctions*, 112 COLUM. L. REV. 203, 249 (2012) (noting that denying injunctive relief can sometimes perform an important "safety valve" function that "both prevents ex post waste and, more importantly, discourages opportunism").

224. *See* Merges, *Contracting into Liability Rules*, *supra* note 222 (noting that private collective rights organizations are well-attuned to the needs of specific industries and provide robust mechanisms for contracting around compulsory licenses and injunctive remedies with relatively low transaction costs).

225. Additionally, parties can also contract around inefficient liability rules. Mark A. Lemley, *Contracting Around Liability Rules*, 100 CALIF. L. REV. 463, 464 (2012).

recognize (or not) those specific situations in which denying injunctions is economically sound, to which I turn next.

C. *Institutional and Cognitive Limitations of Courts and Judges*

Clearly, courts may find it challenging to award an appropriate level of ongoing money damages, regardless of how damages are calculated,[226] and there is reason to believe courts may systematically undercompensate patentholders.[227] An injunction, on the other hand, effectively forces the parties to enter into a private negotiation to determine price terms—assuming, of course, that they can come to some agreement.[228] More specifically, injunctions (i.e., property rules) are more optimal than damages (i.e., liability rules) when the marginal judicial costs of determining damages and the error costs from wrong determinations exceed the marginal costs of granting an injunction (relative to awarding damages).[229]

---

226. *See* Janicke, *supra* note 177 (arguing that numerous uncertainties are introduced under an ongoing royalties approach); Stewart E. Sterk, *Property Rules, Liability Rules, and Uncertainty About Property Rights*, 106 MICH. L. REV. 1285, 1313 (2008) (contending that a pure damages regime results in a "loss of all protection for subjective value, together with increased need for judicial assessment of damages"); Thomas F. Cotter, *The Comparative Law and Economics of Standard-Essential Patents and FRAND Royalties* 31 (Univ. of Minn. Law Sch. Legal Studies Research Paper Series, Paper No. 13-40, 2013) ("[T]he principal economic justification for awarding permanent injunctions in patent infringement cases is a matter of information costs: that is, the assumption that the parties . . . are in a better position than is a court or agency to determine what the patent is worth . . . ."); *cf.* PHILLIP E. AREEDA & HERBERT HOVENKAMP, 3 ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION ¶ 707j, at 293 (3d ed. 2008) (stating that patent "value is almost impossible to determine, apart from such an obvious case as an improved process that reduces everyone's production costs by, say, 10 percent").

227. Blair & Cotter, *supra* note 19, at 69 (contending that reducing damages from a make-whole level could result in too little invention); Peter Camesasca et al., *Injunctions for Standard-Essential Patents: Justice is Not Blind*, 9 J. COMPETITION L. & ECON. 285, 288, 290 (2013) (arguing that injunctions protect against the risk of "reverse holdup," namely opportunistic behavior by potential licensees to "propose rates that are significantly below the fair [market] value"); Einer Elhauge, *Do Patent Holdup and Royalty Stacking Lead to Systematically Excessive Royalties?*, 4 J. COMPETITION L. & ECON. 535, 557 (2008) (contending that hindsight bias, inaccurate damages-estimation techniques, and the difficulties involved in communicating to juries typically leads to undercompensatory damages); Richard A. Epstein, *A Clear View of* The Cathedral*: The Dominance of Property Rules*, 106 YALE L.J. 2091, 2093 (1997) ("The risk of undercompensation in such situations is pervasive . . . .").

228. *See* Merges, *Of Property Rules*, *supra* note 222, at 2664 ("Because each [asset covered by an IP Right] is in some sense unique . . . it is difficult for a court in an infringement case to properly value the right holder's loss. Hence, the parties should be left to make their own deal.").

229. *See* Mark A. Lemley & Philip J. Weiser, *Should Property or Liability Rules Govern Information?*, 85 TEXAS L. REV. 783, 840 (2007) ("Ultimately, policymakers must evaluate the comparative costs and benefits of relying on a property rule regime enforced by courts and accompanied by antitrust oversight against a liability rule regime superintended either by courts or agencies."). The same reasoning holds when the costs of determining less than make-whole damages, as well as related error costs, exceed the costs from awarding make-whole damages.

As an empirical matter, it is difficult to measure how often judicial determination and error costs outstrip the costs from granting injunctions (and make-whole damages).[230] Of course, it can be even more troublesome for a judge to determine potential error costs in any given case.[231] For instance, in my hypotheticals above, I assumed that a host of variables were easily measurable. In actuality, it will often be exceedingly difficult to determine even approximate private economic values of patented technologies—and associated revenues, profits, and costs—much less social values.[232]

Yet, the same kinds of evidence required to determine social values are available under today's damages regime.[233] Using this and related evidence, courts can arguably determine at least when the social costs from injunctions and make-whole damages are substantial compared with the social benefits of the invention. For instance, courts can hear evidence on R&D, testing, and commercialization costs (including the cost of failures); technological and market risk; increased profits versus baseline profits; the value of other patented components; the value of noninfringing alternatives; and so forth, in order to determine when injunctions and make-whole damages might lead to grossly excessive awards.[234] Upon such an examination, courts would routinely find without too much labor or error that—for the very reasons presented earlier—awarding injunctions and make-whole damages in certain situations can lead to drastically suboptimal outcomes.[235] These scenarios include component patents involving high switching costs; high transaction cost, low value inventions; and low R&D

230. *See generally* Cotter, *supra* note 226, at 34–37 (explaining the variety of complex factors that play a role in determining whether "the social harm from [holdup] is greater than the social harm resulting from [reverse holdup]").

231. Blair & Cotter, *supra* note 19, at 69 (criticizing Ayres and Klemperer's partial damages theory for presenting "daunting empirical questions" in practice).

232. *See* Golden, *supra* note 171, at 2115–16 (questioning whether the economic value of a patent can be precisely defined).

233. *See, e.g.*, Roy J. Epstein, *The Market Share Rule With Price Erosion: Patent Infringement Lost Profits Damages After* Crystal, 31 AIPLA Q.J. 1, 32–33 (2003) (recounting the types of "complex evidence" reviewed in a particular patent case, including "[the patentee's] market share, volume discount policy, strategy of dropping the [product's] effective price by adding . . . features but not increasing prices; [the accused infringer's] market share, sales, profits, growth, and selling costs; the level of prices and competition for different product models; entry by new competitors; and expert testimony").

234. For a more comprehensive list of the types of relevant evidence courts should consider, see *infra* subpart IV(A).

235. Ben Depoorter has aptly noted:

> [P]roperty-rights oriented scholars have traditionally presumed that the transaction costs involved in patent license negotiations are negligible compared with the information costs involved when courts apply compulsory licenses (liability rule protection) to patents. Accordingly, property rules provide better footing for consensual agreements in the area of patents, without having courts impose prices on innovation. This conventional wisdom is outdated.

Depoorter, *supra* note 8, at 66.

cost, low commercialization cost, low risk industries.[236] At least in these situations, patent law remedies typically should be modified to deny injunctions and reduce overall damages. In closer cases, especially given the historical practice—and more in line with Kieff and Smith's suggestion—a default rule of injunctive relief and make-whole damages seems more appropriate, at least until we are quite confident in the abilities of adjudicators.

IV. Patent Law as a Public Regulatory Mechanism

In this last Part of the Article, I argue that the previous analysis implies that the current structure of remedies under the Patent Act is fundamentally flawed because it rests on a make-whole notion of damages that often fails to optimally promote innovation. Instead, I suggest that patent law remedies should be viewed not as a means to remedy private wrongs but as part of a larger regulatory mechanism grounded in public law concepts.

A. The Fundamentally Flawed Foundation of the Remedies Provisions in the Patent Act

In Part II, I showed that make-whole damages can often be too high or too low to incentivize optimal levels of innovation. As I explained in Part I, the reason courts impose make-whole damages is because they are mandated to do so by the Patent Act itself. First, 35 U.S.C. § 284 states that "[u]pon finding for the claimant the court shall award the claimant damages

236. Kieff and Smith criticize the view that "[t]he ideal benchmark would be an all-knowing planner who would hand out checks in the exact minimum amount to induce the invention in question—and in any event no greater than the invention's social value" on the ground that it ignores the coordinating function of patents, which ultimately leads to commercialization. Kieff & Smith, *supra* note 222, at 62 ("Patents help achieve this socially constructive coordination by allowing . . . various actors to interconnect with each other like modules of a larger system."). Although Kieff and Smith are correct to include the larger aim of commercialization within patent law's ambit, liability rules need not necessarily lead to less commercialization. Indeed, injunctive remedies not only lead to beneficial coordination but also to detrimental transaction costs that can prevent bargaining among commercial actors, especially for products involving multiple components. *See* Sichelman, *supra* note 35, at 382–88 (explaining the varieties of transaction costs resulting from overly strong and broad patent protection); Bell & Parchomovsky, *supra* note 73, at 590 (suggesting that liability rules may be useful in overcoming strategic obstacles to successful negotiations); Depoorter, *supra* note 8 (arguing that "boundary costs, fragmentation costs, and the costs of bundling necessities" that arise from property rules "have substantially increased the burden on patent licensing markets"); *see also* Ian Ayres & Eric Talley, *Solomonic Bargaining: Dividing a Legal Entitlement to Facilitate Coasean Trade*, 104 YALE L.J. 1027, 1036–72 (1995) (suggesting that liability rules tend to induce actors to disclose more information otherwise withheld under a property-rule regime, increasing the odds that deals are consummated). Thus, Kieff and Smith's claim that "the fear of weak enforcement creates a disincentive for the necessary parties to work together at the outset" is not a universal fact. Kieff & Smith, *supra* note 222, at 63. Rather, the truth lies somewhere in the middle, and it will vary with the facts in each case—facts that an adjudicator can, and should, take into account in fashioning an appropriate remedy.

adequate to compensate for the infringement but in no event less than a reasonable royalty for the use made of the invention by the infringer, together with interest and costs as fixed by the court."[237] In other words, the Patent Act enshrines the tort law compensatory rationale right into the statutory framework for damages.[238]

Based on the analysis I provided earlier, such an approach is misguided because it does not provide for optimal remedies in many circumstances. Rather, the Patent Act would encourage more optimal levels of innovation by giving the court discretion to adjust damages upward or downward from make-whole compensatory levels, particularly when (1) the novel aspect of the invention is but a mere component of a complex invention and switching costs are high; (2) full compensatory damages substantially exceed or fall below the amount needed to incentivize research, development, and commercialization costs and risks of the invention, including opportunity costs; and (3) infringement is especially difficult to detect.[239] Of course, there are very likely many other situations that warrant damages adjustment; ultimately, a more general approach that determined remedies on the basis of innovation incentives per se would be ideal.

As I outlined earlier, such a test would take into account a variety of factors, such as total R&D costs, testing costs (including regulatory approval), commercialization costs, increased profits versus baseline profits, comparable royalties, the value of other patented components, invention risk, commercialization risk, competition risk (including the risk of free riding), available profit from use or licensing, potential R&D duplication costs, potential consumer deadweight losses, the social benefit of the invention relative to noninfringing alternatives, the ease of detecting infringement, other barriers to entry (including other forms of IP), and the like in order to appropriately gauge remedies for infringement.[240] Formulating such a test is an ambitious project that is well beyond the scope of this Article. In the meantime, however, courts should not sit idle when it

---

237. 35 U.S.C. § 284 (2006) (emphasis added).

238. *See* Aro Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 507 (1964) (describing damages analysis as asking, "had the Infringer not infringed, what would the [Patentee] have made?"); Brooktree Corp. v. Advanced Micro Devices, Inc., 977 F.2d 1555, 1579 (Fed. Cir. 1992) ("In patent cases, as in other commercial torts, damages are measured by inquiring: had the tortfeasor not committed the wrong, what would have been the financial position of the person wronged?").

239. *See supra* Part II.

240. *Cf.* Anna B. Laakman, *An Explicit Policy Lever for Patent Scope*, 19 MICH. TELECOMM. & TECH. L. REV. 43, 77 (2012) ("The Federal Circuit should expressly incorporate contextual factors into its patent scope determinations. This includes the cost of R&D, the ratio of R&D costs to imitation costs, technological risk, and the availability of non-patent alternatives for capturing the social value of inventions.").

is all too evident that traditional remedies go far afield in certain classes of cases.[241]

Relatedly, the injunctive relief provision, 35 U.S.C. § 283, should be similarly amended. As it stands, it states, "courts . . . may grant injunctions in accordance with the principles of equity to prevent the violation of any right secured by patent, on such terms as the court deems reasonable."[242] Again, this provision should not aim to "prevent the violation of any right secured by patent," but instead merely to promote optimal innovation incentives. Such a shift would imply that the current four-factor tests for preliminary and permanent injunctions should be suitably modified.[243] Like make-whole damages, the irreparable injury, inadequate compensation, and hardship prongs of the test focus on the private injury to the parties, while only one prong—the public interest factor—concerns society at large. Under a proper model of patent remedies, even if the patentee might experience irreparable harm not curable by monetary damages, which in turn creates a burden that outweighs any experienced by an infringer, an injunction could still be socially detrimental. Ultimately, injunctions should turn only on an analysis that solely concerns the public interest—and not just the interest of the public in using the invention but also its interest in the creation and commercialization of the invention.[244]

Finally, the notion of willful infringement does not seem to accord well with promoting innovation. In tort law, punitive damages are often awarded for intentional, willful conduct in order to provide greater deterrence of harmful behavior. But infringement of a patent is not harmful per se; rather, infringement is only harmful to the extent it denies the patentee an opportunity to be compensated an amount sufficient to induce it

---

241. John Golden argues that a remedies approach focused on costs would be riddled with measurement difficulties. Golden, *supra* note 65, at 538. Yet, the current approach is also riddled with such difficulties. *See supra* notes 176–180 and accompanying text. Thus, the proper determination is a comparative one, and while most cases today will not justify deviating from the baseline, many will. *See supra* subpart III(C).

242. 35 U.S.C. § 283.

243. *See* eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388, 391 (2006) (holding that injunctive relief is appropriate for a patentee when the following factors weigh in its favor: "(1) that it has suffered an irreparable injury; (2) that remedies available at law, such as monetary damages, are inadequate to compensate for that injury; (3) that, considering the balance of hardships between the plaintiff and defendant, a remedy in equity is warranted; and (4) that the public interest would not be disserved by a permanent injunction"); PHG Techs., LLC v. St. John Cos., 469 F.3d 1361, 1365 (Fed. Cir. 2006) (noting that the factors for a preliminary injunction are "(1) the likelihood of the patentee's success on the merits; (2) irreparable harm if the injunction is not granted; (3) the balance of hardships between the parties; and (4) the public interest").

244. Of course, the "public interest" certainly should not presume that granting injunctions promotes innovation, as the Federal Circuit has been wont to do—even after *eBay*. *See, e.g.*, Celsis In Vitro, Inc. v. CellzDirect, Inc., 664 F.3d 922, 931 (Fed. Cir. 2012) ("We have long acknowledged the importance of the patent system in encouraging innovation." (quoting Sanofi–Synthelabo v. Apotex, Inc., 470 F.3d 1368, 1383 (Fed. Cir. 2006))).

to engage in innovative activity.[245] As I explained earlier, in many situations, some level of infringement will promote overall welfare.[246] In other situations, as long as the patentee can recover suitable damages in a lawsuit, whether the infringement was willful becomes a meaningless investigation.

Instead, willful infringement should only play a role in determining damages when it (1) increases the difficulty with which the patentee can detect infringement; or (2) substantially raises the costs of the patentee collecting appropriate compensation for its innovative activity, i.e., by creating the need for a costly lawsuit relative to the value of the innovation. By framing infringement in terms of "wrongful" behavior, Congress and the courts ill-advisedly eradicate some third-party activity that would otherwise serve a salutary purpose.[247] In this sense, willfulness damages solely inflicted as "punishment" for "wrongful" behavior should be eliminated from patent doctrine. Rather, infringement should be encouraged just up to the level that suboptimally diminishes incentives to innovate.

B. Towards a Regulatory Model of Patent Law Remedies

So far, I have argued that all private law notions—contract, real property, and tort—should play no fundamental role in how we think about patent rights and remedies.[248] A more appropriate model is one that draws on what David Rosenberg terms a "public law vision" of the tort system's aims.[249] Instead of viewing patents as conferring a private right on the patentee, under a public law model, a patent confers a kind of private attorney general status on the patentee that allows it to collect payment *on behalf of society*, which is then immediately remitted in full to the patentee solely in order to optimize overall social innovation incentives.[250] Indeed, on a public law model, one could imagine the government bringing patent infringement actions, either in place of or alongside private patentees, then

245. *See supra* section II(C)(2).

246. *See supra* section II(C)(2).

247. *See supra* section II(C)(2).

248. In making this claim, I do not dispute the view that historically the patent system has rested on private law framework. *See* Kerr, *supra* note 81 (setting forth a "private law theory" of patent law); *cf.* John M. Golden, *Patent Privateers: Private Enforcement's Historical Survivors*, 26 HARV. J.L. & TECH. 545, 547–50 (2013) (proposing a "patent privateering" model of enforcement whereby "patents . . . are devices to harness private law enforcement to advance public policy," namely the constitutionally sanctioned goal of "promot[ing] the Progress of Science and useful Arts."). Rather, my point is a normative one.

249. *See generally* David Rosenberg, *The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L. REV. 849 (1984) (positing a "public law" approach to tort law grounded in regulatory aims).

250. *Cf.* Golden, *supra* note 248, at 612 (limiting the "private attorney general" notion in patent law to validity challenges); La Belle, *supra* note 19, at 50–51 (proposing a "private attorney general" public law model of patent validity challenges but not patent litigation as a whole).

remitting damages to the patentee.[251] However, because the patentee typically plays an active role in providing factual evidence in any given case—as well as enjoys the direct benefits of a judgment—arguably it is the least-cost enforcer of patent rights. As such, a patentholder will generally be more effective than the government (or some third party for that matter) to collect damages on society's behalf. But, importantly, the choice of patentee as money collector does not stem from any private right of the patentee to be compensated for harm caused to it.[252]

In this regard, Shubha Ghosh has offered an insightful public regulatory model of patent law. In Ghosh's view, the patent system should not rest on a social contract or private property foundation but rather "a system of regulation."[253] In other words, "patents serve as a way to administer and organize markets."[254] Yet, Ghosh draws particular conclusions about the structure of the patent system that do not in my view necessarily follow from a regulatory model. Specifically, he contends that a regulatory approach counsels in favor of viewing intellectual property law as involving an "assurance game"—namely, one in which "each player independently can be assured that the other[s] will not imitate."[255]

However, there is no a priori reason why a regulatory model mandates any different game theoretic view of patent law than a private law model. The underlying economic problems associated with patent law are the same regardless of the lens through which we view the rights and remedies designed to solve the problems. Rather, the aim of the regulatory approach should be more general: to optimize incentives for invention and commercialization, taking into account potential costs from diminished

---

251. *See* William Kingston, *The "Thesis" Chapters*, *in* DIRECT PROTECTION OF INNOVATION, xi, 20–27 (William Kingston ed., 1987) (suggesting that the government bring infringement actions in certain classes of cases).

252. Of course, we may wish to incorporate principles of substantive fairness and distributive justice into a public law system—for instance, to protect independent inventors whose life's work is appropriated—but conferring traditional private rights is unnecessary to adequately promote these aims. *See* LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE 43–44 (2002) (discussing how deontological "fairness" constraints can impose limits on otherwise purely consequentialist systems).

253. Ghosh, *supra* note 54, at 1387. Lemley rejects a public regulatory model on the grounds that "[r]egulation is out of vogue, and those who talk about intellectual property as regulation usually do so in order to denigrate it." Lemley, *supra* note 25, at 1074. Of course, this is not a substantive rejection but merely one of form. Here, I do not propose a regulatory model as a means of political reform but as the proper theoretical lens from which to view intellectual property.

254. Ghosh, *supra* note 54, at 1350; *see also* LANDES & POSNER, *supra* note 109, at 35–36 (addressing the role of the state in regulating intellectual property law).

255. *See* Ghosh, *supra* note 54, at 1335–36.

follow-on invention, duplicated development and rent-seeking, and consumer deadweight losses, as well as distributive interests.[256]

Conclusion

The fundamental premise of patent law remedies derives from tort law and counsels that courts should always attempt to make the patentee whole in the event of infringement. Here, I have rejected this premise, arguing that the more fundamental goal of promoting innovation sometimes requires deviations from make-whole remedies. Specifically, when a patent covers a mere component of a patented product, and an infringer's switching costs are high, it will often be more desirable to deny injunctive relief and reduce make-whole damages, regardless of whether the patentee practices its patents. Additionally, an injunction may thwart optimal innovation incentives when it generates large consumer deadweight losses, results in substantial duplicated costs during the pre-invention R&D process, or creates transaction costs far in excess of the value of the invention. Specifically, in these situations, some infringement may be efficient. In other situations, such as when infringement is difficult to detect, increasing damages beyond the make-whole level may be optimal.

More generally, I have argued that the notion that patentees should be returned to the *status quo ante* is a private law concept that should play no essential role in determining optimal remedies in patent law. On this ground, the statutory remedies provisions of the Patent Act rest on a flawed foundation. Instead of correcting for "wrongs" inflicted on private parties, patent law remedies should be tailored to optimize innovation incentives, taking into account the various social costs of patenting. Such a public law vision of the patent system is essentially regulatory in nature. By fully removing the private law moorings of patent law remedies, patents will better fulfill their fundamental role in promoting innovation.

256. Indeed, Ghosh recognizes that "[f]rom the perspective of patent law, the relevant interests include pioneer inventors, follow-on inventors, marketers, consumers, and the innumerable other interests reflected in the research and development processes." *Id.* at 1355.

* * *

Understanding Behavioral Antitrust

Avishalom Tor*

Behavioral antitrust—the application to antitrust analysis of empirical evidence of robust behavioral deviations from strict rationality—is increasingly popular and hotly debated by legal scholars and the enforcement agencies alike. This Article shows, however, that both proponents and opponents of behavioral antitrust frequently and fundamentally misconstrue its methodology, treating concrete empirical phenomena as if they were broad hypothetical assumptions. Because of this fundamental methodological error, scholars often make three classes of mistakes in behavioral antitrust analyses: first, they fail to appreciate the variability and heterogeneity of behavioral phenomena; second, they disregard the concrete ways in which markets, firms, and other institutions both facilitate and inhibit rational behavior by antitrust actors; and, third, they erroneously equate all deviations from standard rationality with harm to competition. After establishing the central role of rationality assumptions in present-day antitrust and reviewing illustrative behavioral analyses across the field—from horizontal and vertical restraints, through monopolization, to merger enforcement practices—this Article examines the three classes of mistakes, their manifestation, and their consequences in antitrust scholarship. Besides providing guidance to future behavioral antitrust scholarship, this Article concludes by discussing two sets of essential lessons that the behavioral approach already can offer to advance antitrust law and policy: one concerning the value of case-specific evidence in antitrust adjudication and enforcement, the other showing how antitrust law can and should account for systematic and predictable boundedly rational behavior that is neither constant nor uniform.

---

* Professor of Law and Director, Research Program on Law and Market Behavior, Notre Dame Law School. Earlier versions of this Article benefited from comments and criticisms of Louis Kaplow, Bruce Kobayashi, Steve Salop, Daniel Sokol, Josh Wright, and participants at the 2012 Annual Conference of the American Association of Law Schools (AALS), the 2012 ABA/NYU Law School Next Generation of Antitrust Scholarship Conference, the 60th Annual Meeting of the ABA Section of Antitrust Law, the 29th Annual Meeting of the European Law and Economics Association (EALE), the 2013 Annual Meeting of the American Law and Economics Association (ALEA), and the 2013 Annual Meeting of the International Society for New Institutional Economics (ISNIE), as well as seminar participants at the Copenhagen Business School Law Department, the Max Planck Institute for Research on Collective Goods, the Hebrew University of Jerusalem Faculty of Law, and the Harvard Law School Law and Economics Workshop. Christopher Kieser, John Lindermuth, Geoffrey Miller, Lauren Riley, and Jessica Williams provided excellent research assistance at different stages of this project.

Introduction

The behavioral approach[1] to antitrust law draws on a large body of empirical behavioral evidence to inform antitrust doctrine and policymaking.[2] In particular, behavioral antitrust focuses on findings that

1. For general reviews of the behavioral approach to law, see Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471 (1998); Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CALIF. L. REV. 1051 (2000); and Avishalom Tor, *The Methodology of the Behavioral Analysis of Law*, 4 HAIFA L. REV. 237 (2008) [hereinafter Tor, *Behavioral Methodology*].

2. *See, e.g.*, Amitai Aviram & Avishalom Tor, *Overcoming Impediments to Information Sharing*, 55 ALA. L. REV. 231 (2004); Amanda P. Reeves & Maurice E. Stucke, *Behavioral Antitrust*, 86 IND. L.J. 1527 (2011); Maurice E. Stucke, *Behavioral Economists at the Gate: Antitrust in the Twenty-First Century*, 38 LOY. U. CHI. L.J. 513 (2007) [hereinafter Stucke, *At the Gate*]; Avishalom Tor, *A Behavioural Approach to Antitrust Law and Economics*, 14 CONSUMER POL'Y REV. 18, 18–19 (2004) [hereinafter Tor, *Behavioural Approach*] (noting that this approach

reveal how the judgment and decision behaviors of actual antitrust actors are likely to systematically and predictably deviate from the strict rationality that antitrust law currently assumes.[3]

Perhaps due to the dominance in antitrust of rationality-based law and economics[4]—from the field's jurisprudence and enforcement policies to its legal and economic scholarship—behavioral findings took far longer to garner broad attention in antitrust law than in many other legal fields.[5] In fact, until a few years ago, antitrust discourse largely neglected those behaviorally informed analyses offered by a small number of legal scholars.[6]

---

is "grounded in empirical observations of human behaviour" and "based on scientific findings regarding actual human behaviour, which can often provide better descriptions of market dynamics and thus more effective prescriptions for competition policy"); Avishalom Tor, *The Fable of Entry: Bounded Rationality, Market Discipline, and Legal Policy*, 101 MICH. L. REV. 482 (2002) [hereinafter Tor, *Entry*]; Avishalom Tor, *Illustrating a Behaviorally Informed Approach to Antitrust Law: The Case of Predatory Pricing*, ANTITRUST, Fall 2003, at 52 [hereinafter Tor, *Predatory Pricing*].

3. For discussion of the centrality of the rationality assumption in antitrust, see PHILLIP E. AREEDA & HERBERT HOVENKAMP, 1 ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION ¶113 (4th ed. 2013); HERBERT HOVENKAMP, THE ANTITRUST ENTERPRISE: PRINCIPLE AND EXECUTION 134–36 (2005) [hereinafter HOVENKAMP, ANTITRUST ENTERPRISE]; KEITH N. HYLTON, ANTITRUST LAW 226 (2003); RICHARD A. POSNER, ANTITRUST LAW vii–x (2d ed. 2001); and also see Christopher R. Leslie, *Rationality Analysis in Antitrust*, 158 U. PA. L. REV. 261 (2010), in which Professor Leslie challenges the judicial use of rationality theory in antitrust cases and argues that judges often employ an overly narrow conception of rationality.

4. *See, e.g.*, AREEDA & HOVENKAMP, *supra* note 3 ("As a general proposition business firms are (or must be assumed to be) profit maximizers, which means that they constructively 'intend' to take the course of action that maximizes their returns, given the physical and legal environment in which they find themselves . . . ."); Stucke, *At the Gate*, *supra* note 2, at 536–44 (discussing the Chicago School's continuing influence on antitrust policy); Tor, *Behavioural Approach*, *supra* note 2, at 18 (discussing the impact of the Chicago School on antitrust law and economics); Tor, *Predatory Pricing*, *supra* note 2, at 52 ("One of the core assumptions of the traditional economic approach to antitrust law is that competitors are perfectly rational, profit-maximizing, decision makers.").

5. Compare, for instance, the statements offered already more than a decade ago with respect to behavioral–legal applications more generally, such as Donald C. Langevoort, *Behavioral Theories of Judgment and Decision Making in Legal Scholarship: A Literature Review*, 51 VAND. L. REV. 1499 (1998), in which Professor Langevoort described, fifteen years ago, the many applications of behavioral analysis to legal fields other than antitrust, and Cass R. Sunstein, *Behavioral Law and Economics: A Progress Report*, 1 AM. L. & ECON. REV. 115, 115 (1999), in which Professor Sunstein described a "flood" of behaviorally oriented legal research already in 1999, with recent statements about behavioral antitrust, such as Stucke, *At the Gate*, *supra* note 2, at 514, where Professor Stucke notes, quite colorfully, that "[w]hile tossed against the rocks elsewhere, within the quiet waters of antitrust these rational choice theories stand largely unchallenged," and Luca Arnaudo, *The Quest for Behavioural Antitrust: Beyond the Label Battle, Towards a Cognitive Approach*, DOVENSCHMIDT Q., June 2013, at 77, 79 where Professor Arnaudo states that "when considering the growing fortunes of [behavioral economics], the process towards a [behavioral antitrust] could have been expected to occur much faster . . . ." (footnote omitted).

6. *See, e.g.*, Arnaudo, *supra* note 5 (noting that attempts by pioneering scholars in the '80s and '90s to interject behavioral insights into the antitrust field "basically fell into the void" and

Yet now behavioral antitrust clearly is in vogue: Numerous recent articles by lawyers and economists debate the merits and demerits of behavioral antitrust generally and its specific application to issues spanning horizontal and vertical restraints, monopolization, mergers, and more.[7]

---

that "it is only in the last few years that a massive behavioural takeover of antitrust studies has been experienced"); Max Huffman, Neo-Behavioralism? 3 (Dec. 23, 2010) (unpublished manuscript), *available at* http://ssrn.com/abstract=1730365 ("Behavioral Antitrust has the feel of being something quite new. The earliest article explicitly proposing a behavioral approach to antitrust was written in 2002." (citing Tor, *Entry*, *supra* note 2)). Early antitrust scholarship making use of behavioral evidence includes Aviram & Tor, *supra* note 2; Albert A. Foer, *The Third Leg of the Antitrust Stool: What the Business Schools Have to Offer to Antitrust*, 47 N.Y.L. SCH. L. REV. 21 (2003); Harry S. Gerla, *A Micro-Microeconomic Approach to Antitrust Law: Games Managers Play*, 86 MICH. L. REV. 892 (1988); Harry S. Gerla, *The Psychology of Predatory Pricing: Why Predatory Pricing Pays*, 39 SW. L.J. 755 (1985); Tor, *Behavioural Approach*, *supra* note 2; Tor, *Entry*, *supra* note 2; and Tor, *Predatory Pricing*, *supra* note 2. For an early discussion among prominent antitrust economists that briefly touches on the potential of behavioral economics to inform antitrust law, see also Philip Nelson et al., *Economists' Roundtable*, ANTITRUST, Spring 2003, at 8, 15–16.

7. *See, e.g.*, Matthew Bennett et al., *What Does Behavioral Economics Mean for Competition Policy?*, COMPETITION POL'Y INT'L, Spring 2010, at 111, 120–32; James C. Cooper & William E. Kovacic, *Behavioral Economics and Its Meaning for Antitrust Agency Decision Making*, 8 J.L. ECON. & POL'Y 779 (2012); Eliana Garcés, *The Impact of Behavioral Economics on Consumer and Competition Policies*, COMPETITION POL'Y INT'L, Spring 2010, at 145; Thomas J. Horton, *The Coming Extinction of* Homo Economicus *and the Eclipse of the Chicago School of Antitrust: Applying Evolutionary Biology to Structural and Behavioral Antitrust Analyses*, 42 LOY. U. CHI. L.J. 469 (2011); Max Huffman & Daniel Heidtke, *Behavioral Exploitation Antitrust in Consumer Subprime Mortgage Lending*, 4 WM. & MARY POL'Y REV. 77 (2012); Max Huffman, *Marrying Neo-Chicago with Behavioral Antitrust*, 78 ANTITRUST L.J. 105 (2012) [hereinafter Huffman, *Neo-Chicago with Behavioral Antitrust*]; Leslie, *supra* note 3; Amanda P. Reeves, *Behavioral Antitrust: Unanswered Questions on the Horizon*, ANTITRUST SOURCE, June 2010, art. 3, at 1, *available at* http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/Jun10_Reeves6_24f.authcheckdam.pdf; Stucke, *At the Gate*, *supra* note 2; Maurice E. Stucke, *Money, Is That What I Want?: Competition Policy and the Role of Behavioral Economics*, 50 SANTA CLARA L. REV. 893 (2010); Maurice E. Stucke, *Reconsidering Competition*, 81 MISS. L.J. 107 (2011) [hereinafter Stucke, *Reconsidering*]; Maurice E. Stucke, *New Antitrust Realism*, GCP MAG., Jan. 2009 [hereinafter, Stucke, *New Antitrust Realism*]; Huffman, *supra* note 6; Nicolas Petit & Norman Neyrinck, Behavioral Economics and Abuse of Dominance: A Fresh Look at the Article 102 TFEU Case-Law (May 15, 2010) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1641431; *see also* Mark Armstrong & Steffen Huck, *Behavioral Economics as Applied to Firms: A Primer*, COMPETITION POL'Y INT'L, Spring 2010, at 3; Douglas H. Ginsburg & Derek W. Moore, *The Future of Behavioral Economics in Antitrust Jurisprudence*, COMPETITION POL'Y INT'L, Spring 2010, at 89; Alison Oldale, *Behavioral Economics and Merger Analysis*, COMPETITION POL'Y INT'L, Spring 2010, at 139; Vivien Rose, *The Role of Behavioral Economics in Competition Law: A Judicial Perspective*, COMPETITION POL'Y INT'L, Spring 2010, at 103; Michael A. Salinger, *Behavioral Economics, Consumer Protection, and Antitrust*, COMPETITION POL'Y INT'L, Spring 2010, at 65; Maurice E. Stucke, *Am I a Price-Fixer? A Behavioural Economics Analysis of Cartels*, *in* CRIMINALISING CARTELS: CRITICAL STUDIES OF AN INTERNATIONAL REGULATORY MOVEMENT 263, 289 (Caron Beaton-Wells & Ariel Ezrachi eds., 2011); Maurice E. Stucke, *Behavioral Antitrust and Monopolization*, 8 J. COMPETITION L. & ECON. 545 (2012) [hereinafter Stucke, *Monopolization*]; Maurice E. Stucke, *Greater International Convergence and the Behavioral Antitrust Gambit* [hereinafter Stucke, *Gambit*], *in* RESEARCH HANDBOOK ON INTERNATIONAL COMPETITION LAW 155 (Ariel Ezrachi ed., 2012); Maurice E. Stucke, *Is Intent Relevant?*, 8 J.L. ECON. & POL'Y 801 (2012); Avishalom Tor & William J. Rinner, *Behavioral Antitrust: A New Approach to the Rule of Reason*

Antitrust journals dedicate issues to behavioral antitrust,[8] and the professional associations of practitioners and legal academics in the field devote panels at their meetings to discussing it.[9] Perhaps most telling, even some enforcement agencies and officials now study and discuss the policy implications of this new approach.[10]

After Leegin, 2011 U. ILL. L. REV. 805; Avishalom Tor, *The Market, the Firm, and Behavioral Antitrust*, *in* THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW (Eyal Zamir & Doron Teichman eds., forthcoming 2014) [hereinafter Tor, *The Market*]; Gregory J. Werden et al., *Behavioral Antitrust and Merger Control*, 167 J. INSTITUTIONAL & THEORETICAL ECON. 126 (2011); Joshua D. Wright & Judd E. Stone II, *Misbehavioral Economics: The Case Against Behavioral Antitrust*, 33 CARDOZO L. REV. 1517 (2012).

8. *See, e.g.*, ANTITRUST SOURCE, June 2010, *available at* http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/Jun10_FullSource6_24.authcheckdam.pdf; *The Jevons Colloquium: Behavioral Economics in Consumer Protection and Competition Law*, COMPETITION POL'Y INT'L, Spring 2010, at 89; *A Symposium on Antitrust and Behavioral Economics*, COMPETITION POL'Y INT'L, Spring 2010, at 3.

9. *See, e.g.*, Agenda, Am. Antitrust Inst., 9th Annual Conference: The Next Antitrust Agenda (June 18, 2008), *available at* http://www.antitrustinstitute.org/files/agenda2008_062320081121.pdf; Brochure, Am. Bar Ass'n, Section of Antitrust Law, Spring Meeting (Mar. 28–30, 2012), *available at* http://www.americanbar.org/content/dam/aba/events/antitrust_law/2012/03/spring_meeting/spring_2012_brochure.authcheckdam.pdf; Program, Ass'n Am. Law Schs., AALS 2012 Annual Meeting: Academic Freedom and Academic Duty (Jan. 4–8, 2012), http://www.aals.org/am2012/2012program.pdf; Meeting Announcement, Brit. Inst. Int'l & Comp. L., Competition Law Forum Meeting: Behavioural Economics (July 1, 2009), http://www.biicl.org/events/view/-/id/401/. Moreover, at the *Next Generation of Antitrust Scholarship Conference* held at the New York University School of Law in January 2010, three out of the twelve papers applied behavioral economics to antitrust policy. Agenda, N.Y. Univ. Sch. of Law, Next Generation of Antitrust Scholarship Conference (Jan. 29, 2010), http://www.law.nyu.edu/conferences/nextgenantitrust/.

10. *See, e.g.*, Amelia Fletcher, Chief Economist, U.K. Office of Fair Trading, Address at the European Commission Consumer Affairs Conference: What Do Policy-Makers Need from Behavioural Economists? (Nov. 28, 2008) (transcript available at http://ec.europa.eu/consumers/dyna/conference/programme_en.htm); Interview with Joseph Farrell, Dir. Bureau Econ., FTC, and Carl Shapiro, Deputy Assistant Att'y Gen. Econ. Analysis, U.S. Dep't Justice, Antitrust Div. (Jan. 22, 2010) [hereinafter Farrell & Shapiro Interview], *available at* http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/Feb10_FullSource2_25.authcheckdam.pdf; *see also* OFFICE OF FAIR TRADING, THE IMPACT OF PRICE FRAMES ON CONSUMER DECISION MAKING (2010) (U.K.), *available at* http://www.oft.gov.uk/shared_oft/economic_research/OFT1226.pdf; Bennett et al., *supra* note 7; Oldale, *supra* note 7; Press Release, Eur. Union Comm'n for Consumers, Why Consumers Behave the Way They Do: Commissioner Kuneva Hosts High Level Conference on Behavioural Economics (Nov. 28, 2008), *available at* http://europa.eu/rapid/press-release_IP-08-1836_en.htm; J. Thomas Rosch, Comm'r, FTC, Remarks at the Vienna Competition Conference: Behavioral Economics: Observations Regarding Issues that Lie Ahead (June 9, 2010) [hereinafter Rosch, Issues that Lie Ahead] (transcript available at http://www.ftc.gov/speeches/rosch/100609viennaremarks.pdf); J. Thomas Rosch, Comm'r, FTC, Remarks at the Conference on the Regulation of Consumer Financial Products: Managing Irrationality: Some Observations on Behavioral Economics and the Creation of the Consumer Financial Protection Agency (Jan. 6, 2010) [hereinafter Rosch, Managing Irrationality] (transcript available at http://www.ftc.gov/speeches/rosch/100106financial-products.pdf). Similarly, in June 2012, the Organization for Economic Co-Operation and Development (OECD) held discussions on competition and behavioral economics. *Competition Issues Under Discussion*, OECD, http://www.oecd.org/daf/competition/mergers/workinprogress.htm.

Some particularly enthusiastic proponents of behavioral antitrust depict it as an attractive wholesale alternative to the traditional economic approach to antitrust law.[11] Other supporters view the evidence on systematic deviations from strict rationality marshaled by the behavioral approach, at the very least, as confirming their longstanding suspicions of the soundness of the rationality assumptions underlying the accepted economic methodology in the field.[12] At the same time, however, some commentators have been quick to criticize behavioral antitrust on numerous grounds, challenging the robustness and validity of its empirical evidence as well as the clarity and coherence of its legal implications.[13] In fact, certain detractors have gone so far as to argue that behavioral antitrust cannot possibly benefit the law.[14]

Nevertheless, a closer analysis reveals that both extreme positions in the behavioral antitrust debate are mistaken. Proponents are correct in holding that the behavioral approach can advance antitrust policy and doctrine based on a better understanding of market behavior. Yet those

---

11. *See* Stucke, *Gambit*, *supra* note 7, at 180–81 (viewing behavioral antitrust as an alternative to price theory and effects-based analysis); Rosch, Managing Irrationality, *supra* note 10, at 8–9. Commissioner Rosch stated:

> By engaging in a fact-bound analysis of the conduct and its anticompetitive effect rather than, as the Chicago School would have it, assuming that certain conduct is inherently pro-competitive, I believe the Commission could incorporate insights from the behavioral economics literature in a way that would still put firms on notice of the type of conduct that is anticompetitive.

Id.; *see also* Horton, *supra* note 7, at 473–76 (criticizing traditional antitrust law and economics from an evolutionary biology and behavioral perspective and making the far-reaching assertion that "*Homo economicus* will become extinct [and as] *Homo sapiens* replaces *Homo economicus* in antitrust analysis, the Chicago School's antitrust dominance will come to a timely end" (footnotes omitted)); Werden et al., *supra* note 7, at 127 (claiming, while criticizing it, that the behavioral approach questions "the assumption of profit maximization at the core of neoclassical economic theory").

12. *See, e.g.*, Stucke, *At the Gate*, *supra* note 2, at 513 (characterizing behavioral economics as questioning "neoclassical economic theories' unrealistic and simplifying assumptions about human nature"); Stucke, *Reconsidering*, *supra* note 7, at 109 ("Antitrust policy is built on a flawed assumption of rationality.").

13. For example, in Werden et al., *supra* note 7, the authors discuss the limits of traditional economics in the merger context and note that

> To the extent such departures [from profit maximization] are mistakes, proponents of behavioral antitrust propose to inject paternalism into competition policy, but that is antithetical to the fundamental idea of competition policy. To the extent these departures result from pursuit of non-profit objectives, proponents might identify good reasons for concern about particular forms of anticompetitive conduct, *but they offer nothing to improve the identification of anticompetitive conduct.*

Id. at 138 (emphasis added); *see also* Wright & Stone, *supra* note 7, at 1553 (arguing that to fulfill its promise, behavioral economics must provide "a more robust and accurate account of both firm and consumer behavior" and concluding that "[u]ntil then, we must maintain our observation as to the tentative irrelevance of behavioral economics in antitrust").

14. *See, e.g.*, Wright & Stone, *supra* note 7, at 1526–27 (asserting their "behavioral irrelevance theorem" according to which "behavioral economics . . . fails to offer *any* clear policy implications for antitrust law").

who believe behavioral antitrust can or should provide a complete substitute for the economic analysis of antitrust law are wrong. Instead, both the nascent stage of behavioral antitrust and, more significantly, the very nature of its methodology make it an extremely valuable complement to traditional antitrust economics—capable of offering important refinements and improvements—rather than its complete substitute. Similarly, detractors correctly point out important limitations of behavioral antitrust but overstate their case, failing to recognize the potential of this approach and the essential lessons it already offers for antitrust doctrine and enforcement policy.

This Article shows that many commentators, proponents and opponents alike, reach their respective erroneous conclusions largely due to a shared, fundamental misunderstanding when they treat concrete, empirical behavioral phenomena as if they were instead broad hypothetical assumptions.[15] When this fundamental methodological error leads them astray, scholars make three distinct classes of mistakes, each of which generates its own faulty antitrust applications and policy conclusions.[16] First, analysts fail to appreciate that human judgment and decision behavior is neither constant nor uniform but rather variable and heterogeneous, the evidence of systematic and predictable patterns at the overall population

15. Reeves, for example, mistakenly states that "[b]ehavioral economics attacks the rational profit-maximizer assumption head on by *assuming* that humans have cognitive limitations that prevent them from processing information perfectly and maximizing their utility." Reeves, *supra* note 7, at 2 (emphasis added); *see also* Stucke, *Reconsidering*, *supra* note 7, at 121–22 (assuming, in turn, that firms, consumers, and the government exhibit constant and homogenous bounded rationality); Wright & Stone, *supra* note 7, at 1523, 1535–48 (arguing against what they perceive as the "irrationality hypothesis" used by "modern behavioralists" and basing much of their analysis on various hypothetical assumptions of deviations from rationality); *cf.* Cooper & Kovacic, *supra* note 7, at 780 (noting that behavioral antitrust scholars replace the "assumption of rationality with one of 'bounded rationality'"). *But see* Stucke, *New Antitrust Realism*, *supra* note 7, at 11 ("Behavioral economics at its core is empirical.").

16. Researchers also exhibit a fourth class of mistakes, common to behavioral–legal applications outside antitrust, when they exaggerate the intractability and other limitations of the behavioral evidence and thus understate its usefulness for antitrust analysis. Scholars who are used to the generality and elegance of hypothetical rational-actor models expect in vain the same "grand theory" attributes from behavioral antitrust. In addition, legal analysts sometimes are dismayed by the large number of potentially relevant behavioral findings or find the challenge of determining the ultimate effect of multiple, seemingly contradictory phenomena overwhelming. These concerns largely stem from commentators' lack of a first-hand familiarity with behavioral research. Otherwise sophisticated antitrust scholars fail to realize that the concrete, empirically driven nature of behavioral analysis mostly is incompatible with a grand-theory approach. And while concerns about multiple or conflicting phenomena do have merit, a careful study of the evidence shows these conflicts sometimes are illusory or at least less significant than they initially appear. *See generally* Jeffrey J. Rachlinski, *The "New" Law and Psychology: A Reply to Critics, Skeptics, and Cautious Supporters*, 85 CORNELL L. REV. 739, 745–52 (2000) (discussing methods for addressing seemingly conflicting behavioral phenomena and detailing how its core ideas are relevant to the law); Tor, *Behavioral Methodology*, *supra* note 1, at 301–04 (considering methods for resolving instances in which different and potentially contradictory behavioral processes co-occur).

level notwithstanding. When mistakenly assuming constancy, commentators ignore, for instance, the significance of boundary conditions—the circumstances outside of which a specific empirical regularity is not manifested[17]—and thus sometimes apply behavioral phenomena to irrelevant antitrust settings.

Second, scholars often disregard the concrete ways in which institutions—from markets and firms to enforcement agencies and courts—variously facilitate and inhibit more rational behavior on the part of antitrust actors. For example, behavioral antitrust detractors rightly cite competitive discipline as a force that promotes rational behavior by market participants. Yet these analysts neglect to examine further whether and how the processes of competitive discipline vary in their efficacy and consequences across different market conditions and behaviors. Behavioral antitrust proponents, on the other hand, occasionally commit the mirror-image mistake, failing to consider the rationality-facilitating force of markets and thereby implicitly assuming that behavioral phenomena are always robust to market discipline. In reality, however, markets differently facilitate and inhibit rationality in different circumstances. Therefore, both commentators who unquestioningly rely on markets to produce rational behavior and those who neglect their powerful effects altogether will inevitably reach some erroneous antitrust conclusions.

Third, analysts routinely and mistakenly equate deviations from strict rationality with harm to their perpetrators, to efficiency, and to competition that merits antitrust intervention. Consequently, they tend to embrace or reject the behavioral approach based on their preexisting views regarding the need for a more or less expansive antitrust policy instead of the merits of the behavioral evidence. In fact, however, certain deviations from standard rationality benefit rather than harm those actors exhibiting them.[18] Other deviations may be costly to their perpetrators yet benefit other market participants or society at large.[19] Finally, only a subset of the remaining deviations from strict rationality are comprised of behaviors that are both inefficient and also properly of antitrust concern.[20]

After exploring the three categories of common mistakes and how they can misdirect behavioral antitrust analyses, this Article discusses two essential sets of lessons the behavioral approach already can offer antitrust doctrine and enforcement policy, despite its nascent stage of development

---

17. *See* Tor, *Behavioral Methodology*, *supra* note 1, at 292–300 (examining specific areas of behavioral analysis where boundary conditions alter the effects of behavioral biases).

18. *See, e.g.*, Leslie, *supra* note 3, at 280–85 (noting that seemingly irrational behavior can make economic sense in the context of antitrust conspiracies and predatory business practices).

19. *See, e.g.*, Tor, *Entry*, *supra* note 2, at 543–45 (showing that optimistically overconfident entry can generate some social benefits even while harmful to entrants).

20. *See, e.g.*, Tor & Rinner, *supra* note 7, at 845–63 (explaining that even efficiency-reducing resale price maintenance is only of antitrust concern under specific circumstances).

and inherent limitations. One set of lessons concerns the value of case-specific evidence for antitrust adjudication in both the courts—in contrast with the doctrinal trends of recent decades—and in agency enforcement actions—where such evidence is routinely evaluated, if sometimes based on inappropriate assumptions of rationality. Another set of lessons shows how antitrust doctrine can incorporate the evidence of behavioral regularities in the market without falling prey to the fundamental methodological error of treating these empirical patterns as if they were instead broad hypothetical assumptions. All in all, this Article finds that both some detractors and certain supporters overstate their respective cases: the behavioral approach already offers valuable antitrust lessons but cannot and should not altogether replace traditional antitrust law and economics.

Organizationally, Part I defines behavioral antitrust, highlighting the basic features of the relevant empirical evidence and briefly reviewing illustrative applications from the burgeoning literature in the field. This exercise clarifies the boundaries of behavioral antitrust and reveals why scholars' pervasive methodological error is indeed fundamental. Parts II–IV study the three classes of mistakes that supporters and critics of behavioral antitrust commonly make and the legal consequences of these mistakes. Part V concludes by outlining the two essential sets of lessons that the behavioral approach already offers antitrust doctrine and policy.

I. Foundations

The behavioral analysis of law has been popular among scholars for more than fifteen years,[21] providing an explicit account of legally relevant behavior based on empirical behavioral evidence instead of either everyday intuition—like traditional legal scholarship—or the theoretical rational-actor construct of traditional law and economics.[22] In contrast to its swift endorsement in most other legal fields, until a few years ago the behavioral approach found little traction in antitrust.[23] Yet, more recently, the field's receptiveness to this approach evinced a dramatic change, with an outpour of interest from scholars, practitioners, and even enforcement officials, who all debate the merits and demerits of behavioral antitrust.[24]

This Part explains that neither the delayed reaction of antitrust scholarship to the behavioral approach nor the intensity of the current

21. *See, e.g.*, Langevoort, *supra* note 5 (reviewing the already numerous behavioral applications in law approximately fifteen years ago); Sunstein, *supra* note 5 (same).

22. *See generally* Tor, *Behavioral Methodology*, *supra* note 1 (evaluating the accomplishments, potential, challenges, and limitations of the behavioral analysis of law).

23. *See supra* note 6 and accompanying text.

24. *See supra* note 7 and accompanying text.

debate over it is mere happenstance.[25] To appreciate the forces that shape the antitrust community's reaction to behavioral antitrust, subpart I(A) outlines the role of the rationality assumption in antitrust, highlighting some of its concrete manifestations in legal doctrine and enforcement agency policies. Subpart I(B) then defines the behavioral approach, focusing on the empirical evidence of real human behavior that systematically differs from models of strict rationality. The juxtaposition of these two subparts contrasts the empirically based behavioral approach with the pervasive reliance on hypothetical rationality in antitrust and, in turn, helps explain both the delayed recognition of the behavioral approach and the intensity of the current debate over its usefulness for the field. Subpart I(C) reviews some illustrative behavioral antitrust applications, while subpart I(D) explains the fundamental methodological error that permeates much of this recent scholarship on both sides of the debate, building a foundation for the remainder of this Article.[26]

A. The Rationality Assumption

Present-day antitrust—perhaps more than any other legal area—is based on the traditional economic assumption that market participants are rational decision makers.[27] The producer firms whose conduct is the focus of the field are assumed to be perfectly rational competitors that make strictly rational judgments and whose decisions seek always and only to maximize profits.[28] Moreover, the microeconomic model of competition

---

25. *Cf.* Salinger, *supra* note 7, at 65 ("The interest in behavioral economics (and some of the resistance to it) stems from the belief that it justifies intervention that conventional economic analysis suggests is unwarranted.").

26. Subparts I(A) and I(C) and Part III draw on and develop further this Author's analysis in Tor, *The Market*, *supra* note 7.

27. *See, e.g.*, RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 3 (7th ed. 2007) ("The task of economics . . . is to explore the implications of assuming that man is a rational maximizer of his ends . . . ." (footnote omitted)); STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW 1–2 (2004) (discussing the role of the rationality assumption in descriptive analysis and noting that "the view taken will generally be that actors are 'rational'" and "maximize their expected utility"); Jolls et al., *supra* note 1, at 1481–85 (discussing the conventional law and economics assumption that market participants are strictly rational); Tor, *Behavioral Methodology*, *supra* note 1, at 239–41 (reviewing rational-actor models in law and economics). Note that traditional antitrust economics assumes that all market participants, including consumers, are rational actors who obtain an optimal amount of information, evaluate that information in an unbiased manner, and then proceed to manifest their preexisting, well-ordered preferences in their market behavior. *Cf.* GARY S. BECKER, THE ECONOMIC APPROACH TO HUMAN BEHAVIOR 14 (1976) (arguing that "all human behavior can be viewed as involving participants who maximize their utility from a stable set of preferences and accumulate an optimal amount of information and other inputs in a variety of markets").

28. *See, e.g.*, AREEDA & HOVENKAMP, *supra* note 3 ("As a general proposition business firms are (or must be assumed to be) profit maximizers . . . ."); POSNER, *supra* note 3, at ix ("[T]he issue in evaluating the antitrust significance of a particular business practice should be whether it is a means by which a rational profit maximizer can increase its profits at the expense of efficiency . . . ."); *see also* HOVENKAMP, *supra* note 3, at 134 ("The entire antitrust enterprise is

that the law relies on further assumes that consumers are rational actors as well.[29]

The rationality assumption is not merely an abstract postulate of antitrust economics, but has concrete legal manifestations throughout the field.[30] In *Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*,[31] for instance, the Supreme Court made the legal bar for allegations of illegal monopolization by predatory pricing under Section 2 of the Sherman Act nearly insurmountable by relying on the rationality assumption.[32] The Court declared that conduct will not amount to predatory pricing unless the alleged scheme involved pricing below some measure of cost and the predator "had a reasonable prospect, or, under § 2 of the Sherman Act, a dangerous probability" of recouping its losses from such below-cost predation.[33] The opinion emphasized that because "[r]ecoupment is the ultimate object of an unlawful predatory pricing scheme,"[34] a rational

---

dedicated to the proposition that business firms behave rationally."); Tor, *Entry*, *supra* note 2, at 488–90 (discussing the assumption of entrant rationality in antitrust analysis and citing additional sources); Werden et al., *supra* note 7, at 126–27 ("The tools of neoclassical economics now play a vital role in the analyses conducted by competition agencies and in the litigation of competition cases . . . .").

29. For a typical exposition of the role of consumer rationality in economic analysis see, for example, WILLIAM J. BAUMOL & ALAN S. BLINDER, MICROECONOMICS: PRINCIPLES & POLICY 85–98 (12th ed. 2012). For an informal description of the role of consumers in antitrust's model of market competition, see Bennett et al., *supra* note 7, at 115–17.

30. Because this author and others already have discussed the role of the rationality assumption in antitrust doctrine at some length elsewhere, the present Article only provides a few illustrations. For additional analysis see Leslie, *supra* note 3, at 267–73; Reeves & Stucke, *supra* note 2, at 1549–53; Tor, *Entry*, *supra* note 2; Tor, *Predatory Pricing*, *supra* note 2; and Tor & Rinner, *supra* note 7, discussing resale price maintenance. Note that the impact of the rationality assumption is not limited to judicial doctrine but also informs the enforcement policies of the antitrust agencies. *See, e.g.*, U.S. DEP'T OF JUSTICE & FTC, HORIZONTAL MERGER GUIDELINES § 1 (2010) [hereinafter 2010 MERGER GUIDELINES], *available at* http://www.justice.gov/atr/public/guidelines/hmg-2010.pdf ("In evaluating how a merger will likely change a firm's behavior, the Agencies focus primarily on how the merger affects conduct that would be *most profitable for the firm*." (emphasis added)); Rosch, Managing Irrationality, *supra* note 10, at 9 (noting that a shift to the behavioral approach might impact merger guidelines).

31. 509 U.S. 209 (1993).

32. The Court stated that to be held liable for predatory pricing under Section 2 a competitor must have "a dangerous probability[] of recouping its investment in below-cost prices." *Id.* at 224. The Court continued, "For the investment to be rational, the [predator] must have a reasonable expectation of recovering, in the form of later monopoly profits, more than the losses suffered." *Id.* (alteration in original) (quoting Matsushita Elec. Indus. Co. v. Zenith Radio Corp., 475 U.S. 574, 588–89 (1986)); *see also* Leslie, *supra* note 3, at 263–64 ("Federal judges are more frequently concluding that some types of anticompetitive conduct are facially irrational or implausible and, therefore, could not have occurred as a matter of law (because it is implausible that a business would act irrationally).").

33. *Brooke Grp.*, 509 U.S. at 224; *see also* Tor, *Predatory Pricing*, *supra* note 2, at 55 (noting that according to some economists the conditions for predatory pricing set out by the Supreme Court will rarely be satisfied).

34. *Brooke Grp.*, 509 U.S. at 224; *see also* Tor, *Predatory Pricing*, *supra* note 2, at 55, 58 n.25 (discussing the importance of the recoupment requirement for the Court).

profit-maximizing firm will not engage in such predation unless the monopoly profits it expects to charge in the future—once the competition is driven out of the market—suffice to compensate for those losses inevitably generated by its present, below-cost, predatory sales.[35]

Because it adopted the rationality assumption, the Supreme Court concluded that predatory pricing schemes only rarely are tried and even more rarely are successful.[36] According to this view, for recoupment to be likely the predator must have, *inter alia*, a very large market share that is protected by significant entry barriers.[37] However, because few alleged predators meet the former condition, and few markets meet the latter one, the Court in *Brooke Group* concluded that price predation rarely occurs.[38] Consequently, the Court declared that predatory pricing allegations can be rejected summarily in the presumably common case of unlikely recoupment.[39]

In the years since *Brooke Group*, the lower courts have followed the Court's directive, routinely rejecting predatory-pricing allegations based on the belief in their hypothetical irrationality and, hence, their assumed implausibility.[40] The same rationale was also applied by the Court more recently in *Weyerhaeuser Co. v. Ross-Simmons Hardwood Lumber Co.*[41] to reject allegations of predatory bidding because "a rational firm would not willingly suffer definite, short-run losses" without "a reasonable expectation" of recoupment.[42] More generally, the *Weyerhaeuser* Court

35. *See Brooke Grp.*, 509 U.S. at 225–26; Tor, *Predatory Pricing*, *supra* note 2, at 55, 58 n.26 (noting that the Court has adopted the view of some economic theorists that predatory pricing is typically an irrational strategy). More precisely, rational predation must bear a positive, risk-adjusted, net present value, like any other rational investment activity.

36. *See Brooke Grp.*, 509 U.S. at 226; *see also Matsushita*, 475 U.S. at 588–89 ("A predatory pricing conspiracy is by nature speculative. . . . The foregone profits may be considered an investment in the future. For the investment to be rational, the conspirators must have a reasonable expectation of recovering, in the form of later monopoly profits, more than the losses suffered."); PHILLIP E. AREEDA & HERBERT HOVENKAMP, 3A ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION ¶ 726a (3d ed. 2008) ("No rational firm would bear the losses, difficulties, and possible legal troubles of trying to exclude or discipline rivals by predatory pricing unless it is reasonably confident of a payoff that exceeds the investment . . . .").

37. Tor, *Predatory Pricing*, *supra* note 2, at 55.

38. *See id.* (elaborating on the Supreme Court's reasoning in *Brooke Group*).

39. *Id.* at 55, 59 n.27.

40. *See* Leslie, *supra* note 3, at 272 ("[L]ower courts have reasoned that predatory pricing schemes are 'unlikely to be attempted by rational businessmen.'" (quoting Stearns Airport Equip. Co. v. FMC Corp., 170 F.3d 518, 528 (5th Cir. 1999))). Leslie also points to two lower courts that used summary judgment to dismiss hypothetically irrational allegations of predatory schemes and were affirmed on appeal. *See id.* at 272 n.54 (citing Nat'l Parcel Servs., Inc. v. J.B. Hunt Logistics, Inc., 150 F.3d 970 (8th Cir. 1998) and C.B. Trucking, Inc. v. Waste Mgmt., Inc., 944 F. Supp. 66 (D. Mass. 1996), *aff'd*, 137 F.3d 41 (1st Cir. 1998) as examples).

41. 549 U.S. 312 (2007).

42. *Id.* at 319.

noted that a "rational business will rarely make th[e] sacrifice" involved in such predation.[43]

The Court's reliance on the rationality assumption to formulate antitrust doctrine is not limited to Section 2 predation. For one, allegations of a predatory horizontal conspiracy among competitors under Section 1 of the Sherman Act already were summarily rejected in *Matsushita Electric Industrial Co. v. Zenith Radio Corp.*,[44] a few years prior to *Brooke Group*, once the Court determined that the conspiracy would have required irrational behavior by the alleged conspirators.[45] In reaching this conclusion, the Court in *Matsushita* similarly noted the necessity of a rational expectation of recoupment and explained that as unlikely as it believed predatory pricing schemes to be for a single firm, it considered these schemes even more irrational and unlikely for a cartel.[46]

Notably, the rationality assumption ostensibly was relied on by the *Matsushita* Court only to establish a hurdle that plaintiffs' allegations must clear to survive summary judgment.[47] The more recent *Brooke Group* and *Weyerhaeuser* opinions, on the other hand, went further in explicitly basing a substantive legal requirement for establishing predatory pricing and bidding on the rationality assumption.[48]

Beyond their impact on predation-related doctrines, moreover, assumptions of rationality also have shaped the Court's Section 1 jurisprudence with respect to vertical restraints between manufacturers and their distributors.[49] For instance, *Leegin Creative Leather Products, Inc. v. PSKS, Inc.*[50] replaced the longstanding per se rule against minimum resale price maintenance (RPM)—a vertical restraint that forbids dealers from reselling the products they purchased from a manufacturer below a prescribed price—with a rule of reason approach.[51] In reversing its older precedents, the Court surveyed an antitrust economics literature "replete

43. *Id.* at 323.

44. 475 U.S. 574 (1986).

45. *Id.* at 588–93.

46. *See id.* at 590–91.

47. *Id.* at 587–89; *see also* Leslie, *supra* note 3, at 339–40 (discussing *Matsushita* and arguing, *inter alia*, that it blurred the line between procedural and substantive antitrust rules).

48. *See supra* notes 31–39, 41–43.

49. *See* PHILLIP E. AREEDA & HERBERT HOVENKAMP, FUNDAMENTALS OF ANTITRUST LAW §§ 16.01–.09 (4th ed. Supp. 2013) (surveying some of the relevant case law).

50. 551 U.S. 877 (2007).

51. *Id.* at 881–82. Note that maximum RPM similarly was made subject to rule of reason analysis instead of per se condemnation a decade earlier in *State Oil Co. v. Khan*, 522 U.S. 3, 22 (1997). This was partly based on the rationality assumption. *See id.* at 17 ("But [setting maximum prices too low to support essential and desired services], by driving away customers, would seem likely to harm manufacturers as well as dealers and consumers, making it unlikely that a supplier would set such a price as a matter of business judgment."); *see also* Leslie, *supra* note 3, at 273 (discussing the role of the rationality assumption in *Khan*).

with procompetitive justifications for a manufacturer's use of resale price maintenance"[52] based on an assumption of manufacturer rationality.[53]

Leegin adopted the Chicago School argument that it would be irrational for manufacturers to use RPM, which has the general tendency of raising consumer prices and, therefore, reducing profits, unless they found the practice profitable on balance.[54] According to this view, RPM must be a procompetitive means for facilitating consumer demand and increasing manufacturers' profits despite the higher prices it generates, unless it is shown to support cartelization among either manufacturers or retailers.[55] Resale price maintenance may accomplish its beneficial outcome, for example, by encouraging distributors to provide valuable services to consumers before or after the sale, to engage in brand promotion, and so on.[56]

Based on this reasoning, the *Leegin* Court found that RPM may be rationally anticompetitive in some limited settings but rationally procompetitive in many others, thereby necessitating a case-by-case rule of reason treatment instead of an automatic, per se condemnation.[57] Furthermore, after noting that the practice could be anticompetitive, the Supreme Court left lower courts the task of developing RPM's rule of reason analysis, providing them only with "certain factors" relevant to the inquiry,[58] all based on—and therefore limited by—the rationality assumption.[59] *Leegin* thus offers yet another, more recent illustration of the key role the rationality assumption plays in the formulation of substantive antitrust doctrines.

Besides its pervasive doctrinal impact, the rationality assumption also plays an important role in antitrust enforcement,[60] such as when agencies

52. *Leegin*, 551 U.S. at 889.

53. *See generally* Tor & Rinner, *supra* note 7, at 812 & n.45, 813–15 (reviewing the main procompetitive accounts of RPM and noting that they are based on the rationality assumption).

54. *See Leegin*, 551 U.S. at 896 (citing Frank Mathewson & Ralph Winter, *The Law and Economics of Resale Price Maintenance*, 13 REV. INDUS. ORG. 57, 67 (1998), for the proposition that a manufacturer will use minimum RPM only if increased demand from enhanced service would more than offset the negative demand impact of a higher retail price).

55. *Id.* at 892–93.

56. *See id.* at 890.

57. *Id.* at 894, 898–99.

58. *Id.* at 897–98; *see also* Tor & Rinner, *supra* note 7, at 854–57 (evaluating each of the factors enumerated by the Court in *Leegin*).

59. Tor & Rinner, *supra* note 7, at 854–57. For further detail see *infra* subpart I(C).

60. *See, e.g.*, Stucke, *At the Gate*, *supra* note 2, at 543–45 (discussing the Federal Trade Commission's premerger analysis and its reliance on the assumption that behavior is rational). However, the Department of Justice does not appear to base its criminal enforcement policy on assumptions of strict rationality that, for instance, would rule out the possibility of cartelization where traditional economic models predict competitors cannot maintain such arrangements. Instead, it relies on case-specific evidence, particularly the evidence generated by cartel members

evaluate whether proposed mergers are likely substantially to lessen competition under Section 7 of the Clayton Act.[61] In particular, although the antitrust agencies seek to base their merger decisions on the best available case-specific evidence,[62] various elements of their analyses rely on the rationality assumption. For instance, one category of the potentially adverse effects of a merger on competition concerns the merger's potential for generating "coordinated effects."[63] These effects occur where a merger diminishes "competition by enabling or encouraging post-merger coordinated interaction among firms in the relevant market that harms customers."[64] When predicting the likelihood of post-merger coordination, however, the agencies routinely rely on the traditional, rationality-based, economic view of the conditions necessary for effective collusion to distinguish mergers that raise coordinated-effects concerns from those that do not.[65]

In addition to the important role of the assumption of rationality with respect to producer firms, the assumption of consumer rationality also bears on antitrust doctrine and policy, as the case of aftermarket power illustrates. In *Eastman Kodak Co. v. Image Technical Services, Inc.*,[66] the Court affirmed a denial of summary judgment on claims of Section 1 tying and Section 2 monopolization.[67] The majority ruled that Kodak, a manufacturer of business copiers, could have exercised power in the aftermarket for the sale of machine parts despite competition in the primary market for copiers.[68]

Conversely, the dissent argued that a competitive market in copiers necessarily would prevent Kodak from exercising power in parts.[69] After all, if consumers who already possessed Kodak machines were "locked in"

---

that partake in the leniency program. *See id.* at 575–80, 581 & n.324 (noting differences between the agencies' approach to criminal cartels and other enforcement policies).

61. This Section is codified at 15 U.S.C. § 18 (2012).

62. *See, e.g.*, 2010 MERGER GUIDELINES, *supra* note 30, § 2 (noting that "[t]he Agencies consider any reasonably available and reliable evidence to address the central question of whether a merger may substantially lessen competition" and laying out categories and sources of evidence that agencies have found most informative).

63. *Id.* § 7.

64. *Id.*

65. *See id.* § 7.2; *see also* Interview with Alison Oldale, Deputy Dir. Antitrust, Bureau Econ., FTC (Apr. 24, 2012), *available at* http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/jun12_oldale_intrvw_6_26f.authcheckdam.pdf (stating that behavioral models "aren't there yet" and that "in the meantime our existing models give us workable approximations" but conjecturing that "the first place behavioral economic analysis might be brought to bear on antitrust enforcement will be in areas like coordinated effects or exchange of information[,] . . . areas where our existing theories are not very helpful").

66. 504 U.S. 451 (1992).

67. *Id.* at 456, 486.

68. *Id.* at 476–78.

69. *Id.* at 490–91 (Scalia, J., dissenting).

because they must use compatible parts, any exploitation of the firm's power would raise the price of parts.[70] Yet such a higher aftermarket price effectively would make Kodak's machines more costly and less attractive to rational consumers—who take into account the future costs of parts and services over the copier's lifetime—in the competitive primary market for copiers.[71] Assuming consumer rationality, therefore, the defendant—wishing to avoid damage to its copier sales in the primary market—could not engage in an anticompetitive exercise of aftermarket power.[72]

It is thus clear that the assumption of consumer rationality played a significant, if somewhat implicit, role in the disagreement between the opinions of the majority and the dissent.[73] While perfectly rational consumers in the primary market would have sufficed to deter Kodak from exploiting aftermarket power, the same does not necessarily hold for boundedly rational consumers who may systematically underestimate or fail to consider the future costs of parts.[74]

Vertical price restraints offer another example of the role of consumer rationality in doctrinal debates. As noted above, the arguments over the appropriate legal treatment of RPM have focused on the balance of its harms and benefits for strictly rational firms.[75] Interestingly, one argument for minimum RPM that manufacturers repeatedly advanced but economists summarily rejected was the "loss-leader" concern.[76] Manufacturers argued that retailers discount attractive products, selling them even below wholesale prices, to attract customers and increase sales and profits from other products at quantities that more than compensate for the retailers'

---

70. *Id.* at 496–97.

71. *Id.* at 494–95.

72. *Id.* at 495.

73. For analyses of the merits of these respective positions see, for instance, Steven C. Salop, *The First Principles Approach to Antitrust, Kodak, and Antitrust at the Millennium*, 68 ANTITRUST L.J. 187 (2000) and Carl Shapiro, *Aftermarkets and Consumer Welfare: Making Sense of* Kodak, 63 ANTITRUST L.J. 483 (1995).

74. *Cf.* Bennett et al., *supra* note 7, at 119 (arguing that "[p]assive consumers do not provide the same type of constraints on firms as active consumers do" and, in the context of aftermarkets, softened competition due to "myopic consumers who are unaware of their biases" manifests itself in overall higher prices and a loss in allocative efficiency); Werden et al., *supra* note 7, at 136 (discussing the possible implications of consumers' hyperbolic discounting to the analysis of a merger between producers of durable consumer goods whose customers must also purchase proprietary complements, which are essentially aftermarket products).

75. *See supra* notes 49–59 and accompanying text; *see also* Tor & Rinner, *supra* note 7, at 807.

76. Tor & Rinner, *supra* note 7, at 813; *cf.* Howard P. Marvel, *The Resale Price Maintenance Controversy: Beyond the Conventional Wisdom*, 63 ANTITRUST L.J. 59, 73–77 (1994) (arguing that manufacturers might employ RPM to support inventory holdings and that they object to the use of their products as "loss leaders" out of a fear that competing retailers will refuse to stock products that can only be sold at a loss). *But see* Barak Y. Orbach, *Antitrust Vertical Myopia: The Allure of High Prices*, 50 ARIZ. L. REV. 261, 277–82 (2008) (arguing that consumers sometimes value high prices in and of themselves).

losses on the former loss leaders.[77] Manufacturers oppose the use of their products as loss leaders despite the short-term wholesale profits the practice generates because they believe that frequent discounts diminish the reputation and value of both the specifically discounted products and the manufacturer's brand writ large.[78] However, even economists who favor RPM reject the loss-leader argument, arguing that discounting would not change rational consumer perceptions of the quality of standard goods.[79]

Finally, in the area of merger enforcement, both the agencies and merging parties routinely predict the unilateral effects of mergers based on the estimation of consumer demand.[80] Much like in other aspects of merger evaluation, however, some common merger simulation methods assume consumer rationality regarding the choice among competing products and services.[81]

---

77. Howard P. Marvel & Stephen McCafferty, *The Welfare Effects of Resale Price Maintenance*, 28 J.L. & ECON. 363, 375 (1985); Tor & Rinner, *supra* note 7, at 813.

78. *See, e.g.*, AM. FAIR TRADE COUNCIL, INC., RESALE PRICE MAINTENANCE BY MEANS OF FAIR TRADE LAW IN FORCE APRIL 1, 1942, at 4–5 (1942) (stressing that frequent discounts on products hurt manufacturers and cause consumers to believe that a discounted product is worth no more than is charged by the price-cutting manufacturer); *see also* Leegin Creative Leather Prods., Inc. v. PSKS, Inc., 551 U.S. 877, 883 (2007) (noting that one of the reasons Leegin gave for adopting its RPM policy was the "concern that discounting harmed [its] brand image and reputation").

79. *See* Tor & Rinner, *supra* note 7, at 813. Economists may find the argument compelling with respect to a narrow class of goods whose "luxury" value indeed derives in part from their relatively high price. *See, e.g.*, Laurie Simon Bagwell & B. Douglas Bernheim, *Veblen Effects in a Theory of Conspicuous Consumption*, 86 AM. ECON. REV. 349, 349 (1996); Harvey Leibenstein, *Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand*, 64 Q.J. ECON. 183 (1950); *see also* sources cited *infra* note 152 (discussing empirical evidence for a persistent positive correlation between perceptions of price and quality).

80. *E.g.*, 2010 MERGER GUIDELINES, *supra* note 30, § 6; PETER DAVIS & ELIANA GARCÉS, QUANTITATIVE TECHNIQUES FOR COMPETITION AND ANTITRUST ANALYSIS §§ 8.3–.5 (2009); Gregory J. Werden & Luke M. Froeb, *Unilateral Competitive Effects of Horizontal Mergers*, *in* HANDBOOK OF ANTITRUST ECONOMICS 44, 66–70 (Paolo Buccirossi ed., 2008).

81. *See, e.g.*, Elizabeth M. Bailey, *Behavioral Economics: Implications for Antitrust Practitioners*, ANTITRUST SOURCE, June 2010, art. 4, at 1, 4–5 (noting the dependence of critical loss analysis on assumptions regarding the standard shape of the demand curve); Oliver Budzinski & Isabel Ruhmer, *Merger Simulation in Competition Policy: A Survey*, 6 J. COMPETITION L. & ECON. 277 (2009) (reviewing numerous shortcomings of the different classes of models used in merger simulation, including the reliance of some on restrictive, rationality-based assumptions regarding the shape of the demand function); Daniel Hosken et al., *Demand System Estimation and Its Application to Horizontal Merger Analysis* (FTC, Working Paper No. 246, 2002), *available at* http://www.ftc.gov/sites/default/files/documents/reports/demand-system-estimation-and-its-application-horizontal-merger-analysis/wp246_0.pdf (noting that one significant limitation of logit models is their restrictive assumption of the independence of irrelevant alternatives, one of the axioms of rational choice); *see also* Oldale, *supra* note 7, at 141 (noting that, even where demand is estimated based on actual aggregate data, a behaviorally informed understanding of the factors shaping consumer demand "could highlight possible ways in which the merger might affect the demand function itself, and so suggest reasons why demand should not be treated as a given"). *See generally* ABA SECTION ON ANTITRUST LAW, ECONOMETRICS: LEGAL, PRACTICAL, AND TECHNICAL ISSUES 133–37, 269–309 (2005) (offering a nontechnical introduction to merger simulation methods and noting their many limitations, including the reliance of different models

B. *Defining Behavioral Antitrust*

In clear contrast to the hypothetical rationality assumption, the behavioral approach seeks to provide an empirically based account of the behavior of antitrust actors, from consumers, entrepreneurs, managers, and other business decision makers, to judges, juries, and enforcement officials.[82] Toward this end, behavioral antitrust draws on the extensive findings of behavioral decision research, the psychology of judgment and decision making, and related disciplines.[83]

The main findings of behavioral decision research can be classified into the two general domains of *judgment* and *decision making* (or "choice"), roughly paralleling what economists refer to as individuals' *beliefs* and *preferences*, respectively.[84] Judgment research is concerned with the intuitive formation of beliefs about the past, present, or future state of the world. Intuitive judgments involve mental processes that are neither completely automatic—like visual perception—nor elaborate and controlled—as when people solve a complex problem using a mathematical formula.[85] The study of decision making, on the other hand, examines how individuals choose among alternative courses of action—choices that economists traditionally have considered a mere manifestation of preferences[86] but psychological research proves to entail far more complex processes.[87]

---

on assumptions regarding the behavior of market participants and certain properties of consumer demand).

82. *See, e.g.*, Bennett et al., *supra* note 7, at 114–15 (discussing consumer, or "demand side," behavior); Cooper & Kovacic, *supra* note 7 (applying behavioral economics to enforcement-agency decision making); Ginsburg & Moore, *supra* note 7, at 90 (examining the potential that "judges will consult behavioral economics or literature influenced by behavioral economics with increasing regularity in the not-too-distant future"); Leslie, *supra* note 3, at 342 ("As Chicago School thinking has become entrenched, judges have dismissed and rejected antitrust claims based on narrow and inaccurate conceptions of how businesses operate."); Tor & Rinner, *supra* note 7, at 837–39 (analyzing managers' overestimation of their ability to control risks); Tor, *Entry*, *supra* note 2, at 534–36 (examining the entry judgments and decisions of entrepreneurs in manufacturing industries as well as those of financiers with respect to the ventures of those entrepreneurs).

83. Tor, *Behavioral Methodology*, *supra* note 1, at 242 & n.15.

84. *E.g.*, Nicholas Barberis & Richard Thaler, *A Survey of Behavioral Finance*, *in* 2 ADVANCES IN BEHAVIORAL FINANCE 1, 12–22 (Richard H. Thaler ed., 2005) (dividing relevant psychological findings among the domains of beliefs and preferences); Colin Camerer, *Individual Decision Making*, *in* THE HANDBOOK OF EXPERIMENTAL ECONOMICS 587, 589 (John H. Kagel & Alvin E. Roth eds., 1995) (dividing his analysis into separate sections on judgment and choice).

85. *See, e.g.*, Daniel Kahneman & Shane Frederick, *Representativeness Revisited: Attribute Substitution in Intuitive Judgment*, *in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 49, 50 (Thomas Gilovich et al. eds., rev. ed. 2003) (explaining that intuitive judgments "occupy a position . . . between the automatic parallel operations of perception and the controlled serial operations of reasoning").

86. *See, e.g.*, MARK BLAUG, THE METHODOLOGY OF ECONOMICS 141–44 (2d ed. 1992) (providing a short, nontechnical overview of revealed preferences theory).

87. *See generally* CHOICES, VALUES, AND FRAMES (Daniel Kahneman & Amos Tversky eds., 2000) (providing evidence from multiple articles that decision making involves a variety of

Notably, one of the main foci of judgment and decision research has been the study of whether human behavior accords with normative standards of rationality and—insofar as it does not—how and why it deviates from these standards.[88] Scholars compare intuitive judgments, for example, with the normative standards that probability theory offers for the formation and updating of beliefs.[89] Similarly, in the decision-making domain, researchers contrast the assumptions underlying the normative model of rational choice with actual choice behavior.[90]

By now, many of the most robust findings of behavioral decision research have been reviewed in the legal literature generally[91] and, albeit less extensively, in antitrust scholarship specifically.[92] Hence, the following paragraphs offer only a brief overview of the overarching approach of behavioral decision research.

When forming their beliefs about the world, antitrust actors routinely make legally relevant judgments, mostly under conditions of uncertainty.[93]

---

different factors). For a useful elaboration of the distinction between judgment and decision making, see, for example, Robyn M. Dawes, *Behavioral Decision Making and Judgment*, *in* 1 THE HANDBOOK OF SOCIAL PSYCHOLOGY 497, 497–99, 530–33 (Daniel T. Gilbert et al. eds., 4th ed. 1998). *See also* Daniel Kahneman, *Preface* to CHOICES, VALUES, AND FRAMES, *supra*, at ix–xi; Gregory Mitchell, *Taking Behavioralism Too Seriously? The Unwarranted Pessimism of the New Behavioral Analysis of Law*, 43 WM. & MARY L. REV. 1907, 1920 n.20 (2002) (elaborating on the distinction between judgment and decision).

88. *See* Tor, *Behavioral Methodology*, *supra* note 1, at 245–72 (providing a general examination of the main psychological research and findings in the area of deviations from rationality in judgment and decision-making outcomes); *see also* William M. Goldstein & Robin M. Hogarth, *Judgment and Decision Research: Some Historical Context*, *in* RESEARCH ON JUDGMENT AND DECISION MAKING: CURRENTS, CONNECTIONS, AND CONTROVERSIES 3, 4–6 (William M. Goldstein & Robin M. Hogarth eds., 1997) (discussing the advent of the influential von Neumann–Morgenstern axioms and how they "instigated a pattern of psychological experiments in which behavioral deviations from a presumed standard of rationality are considered the 'interesting' phenomena to be explained").

89. *See, e.g.*, Dawes, *supra* note 87, at 530–33 (comparing common cognitive biases with the Bayes Theorem of rational action).

90. *See, e.g.*, Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263 (1979) [hereinafter Kahneman & Tversky, *Prospect Theory*] (critiquing expected-utility theory as a descriptive model of decision making and presenting evidence from experiments in which subjects exhibited pervasive tendencies inconsistent with utility theory's basic tenets); Amos Tversky & Daniel Kahneman, *Rational Choice and the Framing of Decisions*, 59 J. BUS. (SPECIAL ISSUE) S251, S252 (1986) [hereinafter Tversky & Kahneman, *Rational Choice*] ("[D]eviations of actual behavior from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the normative system.").

91. *See, e.g.*, sources cited *supra* note 1.

92. *See, e.g.*, Armstrong & Huck, *supra* note 7; Bennett et al., *supra* note 7; Reeves & Stucke, *supra* note 2; Tor & Rinner, *supra* note 7; Tor, *Behavioural Approach*, *supra* note 2; Tor, *Entry*, *supra* note 2; Tor, *Predatory Pricing*, *supra* note 2.

93. The distinction between risk and uncertainty was originally made in FRANK H. KNIGHT, RISK, UNCERTAINTY AND PROFIT 19–20 (1921). For one definition of the distinction between uncertainty, risk, and certainty in a classical text, see R. DUNCAN LUCE & HOWARD RAIFFA, GAMES AND DECISIONS: INTRODUCTION AND CRITICAL SURVEY 13 (1957).

They predict the future success of a new technological platform they consider buying or the future performance of their business venture; they judge whether the evidence at trial proves that a defendant indeed participated in an illegal cartel; and they determine whether the present activities of a particular retailer violate their manufacturer's distribution policy.

Such judgments require—at least implicitly—that individuals assess the probability of different outcomes, a task for which people use cognitive heuristics (mental shortcuts), immediate affective reactions, and more.[94] These heuristic processes, which help real-world, "boundedly rational" decision makers economize on their limited cognitive resources,[95] also generate some costs. Cognitive heuristics, for instance, permit decision makers quickly to reach approximate judgments most of the time, with little conscious effort.[96] However, the same mental shortcuts also cause sometimes predictable and systematic errors known as "judgmental biases."[97]

Based on the beliefs they form through judgment,[98] antitrust actors constantly must make legally relevant decisions under uncertainty. They have to determine what course of action to take in the market when engaging in competitive and strategic interaction, making enforcement and policy decisions, and more. For the hypothetical rational actor, decision making is a straightforward matter, a mere revelation of preexisting, well-ordered preferences[99] that always maximizes subjective expected utility

---

94. *See, e.g.*, Tor, *Behavioral Methodology*, *supra* note 1, at 245–51 (discussing various judgment heuristics).

95. *See id.* at 242 & n.16, 243 (offering a brief discussion of the development of the concept of bounded rationality); *see also* Salinger, *supra* note 7, at 71 ("[B]ounded rationality means that individuals (or firms) act purposefully, but not necessarily as if they are both fully informed and perfectly rational."); Glenn Ellison, Bounded Rationality in Industrial Organization 1 (Jan. 2006) (unpublished manuscript), *available at* http://economics.mit.edu/files/904 (noting that the term "boundedly rational" has been used by different groups of economists to describe different styles of work and discussing three different "traditions" of work that use this term).

96. *See, e.g.*, Kahneman & Frederick, *supra* note 85, at 58 ("[P]eople . . . are often content to trust a plausible judgment that quickly comes to mind.").

97. *See, e.g.*, *id.* at 53 (recognizing that "[m]any judgments are made by . . . attribute substitution" and that these are sometimes "given . . . too much or too little weight" (emphasis omitted)); *see also* Rachlinski, *supra* note 16, at 755–56 (noting that people, particularly judges, typically are unaware of using heuristics); Tor, *Behavioral Methodology*, *supra* note 1, at 245 (explaining that decision makers will often answer an easier question whose answer comes readily to mind when they are confronted with a harder question).

98. In reality, of course, decision and choice do not always follow judgment, although analytically they should be based on the beliefs antitrust actors hold based on judgment.

99. The notion that choices "reveal" people's preferences originated with P.A. Samuelson, *A Note on the Pure Theory of Consumer's Behaviour*, 5 ECONOMICA 61 (1938) and P.A. Samuelson, *A Note on the Pure Theory of Consumer's Behavior: An Addendum*, 5 ECONOMICA 353 (1938), though he did not use that terminology at the time, and has received much attention and development since. For one short and highly readable discussion of the concept and its

(SEU).[100] When faced with risky or uncertain prospects, the rational actor takes into account her judgments of the value and probability of these options as well as her risk preferences.[101] Risk-neutral decision makers choose the option with the highest expected value,[102] risk-averse ones discount the expected value of risky or uncertain prospects to account for the risk involved, and risk-seeking actors find risky prospects more attractive than their mere expected value indicates.[103]

Importantly for antitrust analysis, in economic models the decision behavior of business managers is even more narrowly circumscribed: managers are tasked with maximizing the firm's profits and, therefore, should exhibit risk neutrality when making decisions on behalf of the firm.[104] While they may be risk averse under limited circumstances, rational managers would not make for their firms risk-seeking decisions—which by definition have a negative expected value and are therefore deemed irrational market behavior.[105]

However, much as in the case of belief formation through judgment, a wealth of psychological evidence reveals that real, boundedly rational individuals systematically and predictably deviate from the theoretical

---

appeal, see Amartya Sen, *Behaviour and the Concept of Preference*, 40 ECONOMICA 241, 241–44 (1973).

100. *E.g.*, BLAUG, *supra* note 86, at 229. In defining rationality, Blaug states:

> [S]ome regard . . . the most characteristic feature of neoclassical economics [as] its insistence on methodological individualism: the attempt to derive all economic behavior from the action of individuals seeking to maximize their utility, subject to the constraints of technology and endowments. This is the so-called *rationality postulate*, which figures as a minor premise in every neoclassical argument.

Id. The axiomatization of SEU was formalized by JOHN VON NEUMANN & OSKAR MORGENSTERN, THEORY OF GAMES AND ECONOMIC BEHAVIOR 617–28 (60th Anniversary ed. 2004) and LEONARD J. SAVAGE, THE FOUNDATIONS OF STATISTICS (1954); Savage developed the notion of personal, subjective probability and tied it to expected utility. For a discussion of rational choice theory in law, see Korobkin & Ulen, *supra* note 1, at 1060–75.

101. *See* ROBIN M. HOGARTH, JUDGMENT AND CHOICE: THE PSYCHOLOGY OF DECISION 89–90 (2d ed. 1987) (describing how a rational person will make a decision in an uncertain situation based on her subjective value of the possible outcomes combined with her risk attitude).

102. For example, a risk-neutral person would prefer a 50% chance of receiving $101 over the certain receipt of $50.

103. *See, e.g.*, VON NEUMANN & MORGENSTERN, *supra* note 100, at 629 (noting specifically that their axiomatization of expected utility does not require a specific risk attitude); *see also* HOGARTH, *supra* note 101 (describing these different utility functions and their interpretation).

104. *See* AREEDA & HOVENKAMP, *supra* note 3; POSNER, *supra* note 27, at 289; Korobkin & Ulen, *supra* note 1, at 1066 ("Nearly all law-and-economics literature on business organizations, following the neoclassical economic theory of firms, is built on the explicit or implicit assumption that firms seek to maximize profits." (footnote omitted)).

105. *See infra* note 365 (collecting sources discussing the concept of net present value). *But see infra* subpart III(B) (explaining that some deviations from models of strict rationality—including certain forms of risk-seeking behavior—are in fact rational for managers individually and, occasionally, for their firms as well).

model of rational choice in important respects.[106] The same sensitivity to subtle contextual cues that helps people intuitively navigate complex real-world situations also leads them predictably to violate the normative requirements for SEU maximization by acting inconsistently at different times, in different contexts, with respect to different subject matters, and so on.[107]

In sum, behavioral antitrust can be defined as the application of empirical behavioral findings to antitrust law. This approach draws upon the extensive evidence generated by researchers focusing on the processes that shape human judgment and decision making, paying particular attention to those systematic, predictable deviations of real, boundedly rational behavior from the assumptions of strict rationality.[108]

C. *Illustrative Applications*

In recent years, numerous commentators have joined the few researchers that previously offered behavioral antitrust analyses, quickly generating a sizable body of scholarship that addresses many areas of antitrust law.[109] Behavioral analyses of antitrust law typically draw on evidence suggesting that real market participants deviate systematically in some specific respect from the predictions of the rationality-based economic models that antitrust law relies on. In some cases, scholars further argue that these deviations warrant changes in enforcement policy or antitrust doctrine.[110] This subpart illustrates the range of applications commentators already offer, providing a more concrete foundation for the critical evaluation of this scholarship in the remainder of this Article.

Where horizontal restraints among competitors are concerned, scholars argue that behavioral forces make cartelization both more likely and more

---

106. *See* Tor, *Behavioral Methodology*, *supra* note 1, at 245–57 (offering examples of ways individuals tend to deviate from the theoretical model of rational choice).

107. The literature in this area is voluminous. *See generally* CHOICES, VALUES, AND FRAMES, *supra* note 87 (collecting important articles in this area); Dawes, *supra* note 87, at 499–530 (reviewing and discussing some of the basic decision-making phenomena that violate the axioms of rational choice). For one review and legal application of some basic decision-making findings, see Tor, *Behavioral Methodology*, *supra* note 1, at 258–72.

108. Note that this definition also delimits the boundaries of the behavioral approach to antitrust law, excluding, for instance, critiques of the rationality assumption that are not based on empirical behavioral evidence. Similarly, behavioral antitrust properly understood does not encompass those market features that traditional antitrust economics already account for, including asymmetric information and related phenomena, network effects, and more. *Contra* Stucke, *Monopolization*, *supra* note 7, at 552–53 (denoting as behavioral certain nonbehavioral aspects of network effects and learning processes).

109. *See* sources cited *supra* notes 6–14 and accompanying text.

110. *See, e.g.*, Reeves & Stucke, *supra* note 2, at 1532–33, 1577 (stating that "[c]onsumers are not perfectly objective and rational Bayesians" and later suggesting that behavioral economics should start to influence antitrust doctrine as a result of these deviations).

stable than traditional antitrust theories suggest.[111] This position is supported by the numerous examples of real-world cartels that were exposed and prosecuted in industries and product markets where, according to traditional economic accounts, they should not have existed and could not have thrived for extended periods.[112] These ubiquitous real-world cartels, both domestic and global in scope, spanned markets with large numbers of competitors, relatively low entry barriers, nonhomogeneous products with complex pricing and cost structures, and other characteristics that make cartelization unlikely for strictly rational actors.[113]

Explaining this evidence, commentators argue that behavioral factors, such as managers' social preferences for trust and cooperation, personal relationships, social networks, and social norms all help competing firms both establish and maintain collusive arrangements where rationality-based models that ignore such factors expect them to fail.[114] Other researchers point to additional, nonsocial phenomena, such as managers' aspiration to

111. *See, e.g.*, Armstrong & Huck, *supra* note 7, at 19–22 (noting that vengeful behavior or *esprit de corps* can sustain collusion); Leslie, *supra* note 3, at 280–84, 324–34 (arguing that seemingly irrational conduct may be a rational business decision in the context of an antitrust conspiracy); Reeves & Stucke, *supra* note 2, at 1563–67 (expressing skepticism about the assumption that the presence of big buyers in a market causes rational cartel members to defect); Stucke, *At the Gate*, *supra* note 2, at 568–69 (noting the stability of conspiracies with eleven or more participants); Marie Goppelsroeder, *Three Is Still a Party: An Experiment on Collusion and Entry* 24–26 (Amsterdam Ctr. for Law & Econ., Working Paper No. 2009–05, 2009), *available at* http://ssrn.com/abstract=1368728 (summarizing findings on entrant behavior that successfully sustained collusion); *see also* Christoph Engel & Lilia Zhurakhovska, Oligopoly as a Socially Embedded Dilemma. An Experiment 24–25 (Nov. 2011) (unpublished manuscript), *available at* http://www.coll.mpg.de/pdf_dat/2011_01online.pdf (discussing the behavioral effects of sanctions and externalities on collusion).

112. *See, e.g.*, Armstrong & Huck, *supra* note 7, at 22 (noting evidence of *esprit de corps* in the American steel cartel and the nineteenth-century UK shipping cartel); Leslie, *supra* note 3, at 324–34 (analyzing alleged price-fixing conspiracies in the tobacco, citric acid, and potash industries); Reeves & Stucke, *supra* note 2, at 1563–66 (noting that cartels in the citric acid, lysine, liquid crystal display panels, air transportation, Dynamic Random Access Memory, and graphite electrodes industries all persisted despite the existence of large, sophisticated buyers); Stucke, *At the Gate*, *supra* note 2, at 565–66 (noting that over twenty industries with moderate or low barriers to entry have been criminally prosecuted for price fixing or bid rigging).

113. More precisely, these factors diminish the likelihood of cartelization in rationality-based models only under some common specifications. For scholarship on the circumstances that lead to cartelization, see Margaret C. Levenstein & Valerie Y. Suslow, *What Determines Cartel Success?*, 44 J. ECON. LITERATURE 43, 45–49 (2006) and George J. Stigler, *A Theory of Oligopoly*, 72 J. POL. ECON. 44, 45–48 (1964).

114. *See, e.g.*, Armstrong & Huck, *supra* note 7, at 21–22 (noting the central role of trust and camaraderie, fostered through social events and meetings, in facilitating collusion); Bennett et al., *supra* note 7, at 124 ("[T]here is evidence that personal friendship and trust can play an important role in sustaining collusion, with cartel members often investing a lot of time and effort in individual relationships."); Leslie, *supra* note 3, at 281 (noting that facially irrational conduct, such as declining to make profitable sales, is often employed as a "trust-building goodwill gesture[]" in order to "maintain friendly relations among cartel participants").

obtain merely satisfactory—rather than maximal—profitability.[115] Firms whose managers exhibit such preferences find the potential for more secure profits through cartelization more attractive than the profit-maximizing firm does.[116]

Analysts also argue, however, that behavioral forces can destabilize, rather than facilitate, collusion due to processes that traditional models ignore. For example, some studies of experimental markets show that an increase in the amount of information available to competitors about rivals' output and profits—which would make easier oligopolistic coordination by rational actors—can lead in fact to less collusive, more competitive market behavior.[117] Similarly, the broader behavioral literature makes clear that individuals' concern for relative—as opposed to absolute—outcomes is ubiquitous,[118] particularly common in competitive settings,[119] and evidenced

---

115. *See, e.g.*, R.M. Cyert & James G. March, *Organizational Factors in the Theory of Oligopoly*, 70 Q.J. ECON. 44, 63 (1956) (analyzing firms' planning and budgeting processes and finding "adequate basis to justify the introduction of the concept of an acceptable-level profit norm in place of the traditional profit maximizing assumption"); Huw David Dixon, *Keeping up with the Joneses: Competition and the Evolution of Collusion*, 43 J. ECON. BEHAV. & ORG. 223, 224 (2000) (advancing a model in which "the aspiration level of all firms is to have at least normal-profits" and concluding that "cooperation is not only possible, but almost inevitable" in this economic system (emphasis omitted)); Jörg Oechssler, *Cooperation as a Result of Learning with Aspiration Levels*, 49 J. ECON. BEHAV. & ORG. 405, 406 (2002) (building on Dixon's work and showing that "imitation may result in collusion if the population average is imitated in the form of an aspiration level"); *cf.* HERBERT A. SIMON, ADMINISTRATIVE BEHAVIOR: A STUDY OF DECISION-MAKING PROCESSES IN ADMINISTRATIVE ORGANIZATION 270–75 (3d ed. 1976) (noting that in actual organizational practice, "no one attempts to find an optimal solution for the whole problem," but instead "specialized members or units of the organization. . . . find a 'satisfactory' solution for one or more subproblems").

116. Note that a similar result may obtain in the familiar case where rational, yet risk-averse, managers who pursue their self-interest believe the expected outcomes of cartelization to be more stable and less risky than competition to maximize profits and, therefore, diverge from what is optimal for the firm. *See, e.g.*, Giancarlo Spagnolo, *Managerial Incentives and Collusive Behavior*, 49 EUR. ECON. REV. 1501, 1503 (2005) (arguing that "[f]irms whose pricing policy is in the hands of managers that prefer smooth profit streams can support any collusive agreement at lower discount factors than profit-maximizing ones" and explaining that "the preference for smooth profit streams reduces managers' appreciation of short-run profits from unilaterally breaking a collusive agreement and increases that of losses from the punishment phase that follows"). *See generally* Henry Hansmann & Reinier Kraakman, *What Is Corporate Law?*, *in* REINIER KRAAKMAN ET AL., THE ANATOMY OF CORPORATE LAW: A COMPARATIVE AND FUNCTIONAL APPROACH 1, 2–3 (2d ed. 2009) (providing a brief introduction to agency problems in corporate law); John W. Pratt & Richard J. Zeckhauser, *Principals and Agents: An Overview*, *in* PRINCIPALS AND AGENTS: THE STRUCTURE OF BUSINESS 1, 2–4 (John W. Pratt & Richard J. Zeckhauser eds., 1985) (offering a general introduction to the principal–agent relationship).

117. *See, e.g.*, Steffen Huck et al., *Does Information About Competitors' Actions Increase or Decrease Competition in Experimental Oligopoly Markets?*, 18 INT'L J. INDUS. ORG. 39 (2000); Steffen Huck et al., *Learning in Cournot Oligopoly: An Experiment*, 109 ECON. J. C80, C87–89 (1999); Theo Offerman et al., *Imitation and Belief Learning in an Oligopoly Experiment*, 69 REV. ECON. STUD. 973 (2002).

118. *See, e.g.*, Jerry Suls & Ladd Wheeler, *A Selective History of Classic and Neo-Social Comparison Theory*, *in* HANDBOOK OF SOCIAL COMPARISON: THEORY AND RESEARCH 3, 3–15 (Jerry Suls & Ladd Wheeler eds., 2000) (summarizing the history of self-evaluation through social

in managerial behavior.[120] Yet the concern for relative outcomes, if manifested by managers when making decisions on behalf of their firms, can sometimes destabilize cartels and make markets more competitive.[121]

More generally, beyond showing how specific firm and market characteristics may facilitate or inhibit collusion in ways that traditional models neglect, the behavioral evidence also reveals that established patterns of market behavior—whether competitive or collusive—tend to exhibit greater stability than standard antitrust models assume.[122] Extant rivals are slower to recognize and embark on mutually profitable opportunities for coordinated behavior—whether legal, collaborative arrangements or illegal cartelization—than rationality-based accounts allow for.[123] At the same time, established collaborative or collusive arrangements are also more stable than they would have been if the parties were strictly rational actors.[124]

A number of factors combine to make market behavior "sticky." In the domain of judgment, for instance, established norms of rivalry diminish competitors' ability to identify profitable opportunities for cooperation while collusive norms can have the opposite effect.[125] Managers' risk attitudes also can lead them to overestimate the value of extant arrangements and underestimate the benefits of alternative courses of

---

comparison and discussing the importance of relative comparisons). This concern is related to social comparison, a fundamental psychological process whereby people engage in self-evaluation and self-enhancement by comparing themselves to others—"a core aspect of human experience." *Id.* at 15.

119. *See generally* Andrew E. Clarke et al., *Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles*, 46 J. ECON. LITERATURE 95 (2008) (finding that one's happiness is related to income relative to others rather than absolute income).

120. *See, e.g.*, Armstrong & Huck, *supra* note 7, at 17–18 (discussing the importance of relative performance with regard to managerial behavior); Lorenz Graf et al., *Debiasing Competitive Irrationality: How Managers Can Be Prevented from Trading Off Absolute for Relative Profit*, 30 EUR. MGMT. J. 386, 386–90 (2012) (reviewing findings for managers' frequent preference for relative position, at times even at the expense of profit maximization). *See generally* Robert Gibbons & Kevin J. Murphy, *Relative Performance Evaluation for Chief Executive Officers*, 43 INDUS. & LAB. REL. REV. 30-S (1990) (addressing the costs and benefits of relative performance evaluations for chief executive officers).

121. Armstrong & Huck, *supra* note 7, at 19. Concerns for relative position can also have the opposite effect of reinforcing cartelization under some circumstances. *See id.*

122. *See* Aviram & Tor, *supra* note 2, at 247–63 (discussing the stabilizing functions of social norms).

123. *See id.* (noting that uncertainty, social norms, managerial perception of risk, the illusion of control, and status quo bias impact market participants' judgments of information sharing).

124. *See id.* at 251–52.

125. *See, e.g.*, *id.* at 250–54 (noting that managers' aversion to seemingly uncontrolled risks erects an additional barrier to collaboration with rivals); Margaret M. Blair & Lynn A. Stout, *Trust, Trustworthiness, and the Behavioral Foundations of Corporate Law*, 149 U. PA. L. REV. 1735, 1774–75 (2001) (noting that, in experimental settings, subjects who cooperate during initial rounds tend to continue their cooperation and observing that similar patterns are seen in the business world).

interaction with rivals.[126] Moreover, some decision phenomena—including the status quo bias and the aversion to comparative ambiguity—may both lead competitors consciously to forego risky opportunities for profitable collusion and inhibit cartelists' willingness to chance potentially profitable competitive alternatives to ongoing collusive arrangements.[127]

Beyond horizontal restraints, questions of monopolization generally, and the likelihood of predatory behavior specifically, have received significant attention from behavioral antitrust scholars. This author and others identified circumstances where monopolists may engage in predatory behavior that fails to maximize expected profits and is therefore irrational according to the standard account.[128] For example, managers of a dominant firm that is losing market share may take excessive risks due to loss aversion, while those of established, profitable monopolists may exhibit the opposite pattern of excessive risk avoidance.[129] Some scholars offer further evidence from antitrust cases of predatory behavior that appears irrational,[130] while others show how even rational monopolists may find it beneficial to imitate the behavior of their irrational counterparts when market participants know that some monopolists may engage in irrational predation.[131]

At the same time, some analysts suggest that traditional models can overstate the harm from substantial market power.[132] According to this argument, real firms sometimes avoid fully exploiting their market power, charging prices they deem "fair" instead of maximizing profits.[133] For instance, when market power is generated by recent or temporary changes in market conditions, firms may not exercise it fully, whether to maintain a

126. Aviram & Tor, *supra* note 2, at 254–57.

127. *See id.* at 257–63.

128. *See, e.g.*, Tor, *Predatory Pricing*, *supra* note 2, at 55 ("In contrast with the accepted wisdom on the extreme rarity of predatory pricing, the behavioral evidence suggests that dominant firms and monopolists consciously may engage in high-risk, negative net present value predation under some circumstances."); *see also* Leslie, *supra* note 3, at 274–84 (discussing various examples of firms engaging in behavior inconsistent with perfect rationality and profit maximization).

129. Tor, *Predatory Pricing*, *supra* note 2, at 55–56.

130. *See, e.g.*, Leslie, *supra* note 3, at 274–84.

131. *See, e.g.*, Armstrong & Huck, *supra* note 7, at 30–31 (explaining how it can be rational for a firm to mimic a competitor that is engaging in predatory pricing); Leslie, *supra* note 3, at 297–301 (describing how credible threats to engage in predatory behavior can be rational); *see also* Aaron Edlin, *Predatory Pricing*, *in* RESEARCH HANDBOOK ON THE ECONOMICS OF ANTITRUST LAW 144, 151–53 (Einer Elhauge ed., 2012) (explaining that formal models show that "[w]hether predation is a successful strategy depends very much on whether predator and prey believe it is a successful strategy").

132. *See, e.g.*, Bailey, *supra* note 81, at 5–7.

133. *See, e.g.*, *id.* *See generally* Daniel Kahneman et al., *Fairness as a Constraint on Profit Seeking: Entitlements in the Market*, 76 AM. ECON. REV. 728 (1986) (analyzing how standards of fairness can explain market anomalies).

reputation for offering low prices or to avoid negative reactions from consumers to prices the latter perceive as "unfair."[134]

A different behavioral antitrust inquiry concerns aftermarket power—both within and beyond monopolization claims—as illustrated by the divided Court in *Kodak*: the majority ruled that the manufacturer could have exercised aftermarket power in parts despite competition in the primary market for copiers. This holding would have been impossible if consumers were all perfectly rational, incorporating the expected costs of parts over the copier lifetime into the original purchase price of the machine.[135]

Importantly, the majority's conclusion did not require a finding that Kodak in fact exercised power in the parts aftermarket since the Court only affirmed the denial of summary judgment by the court of appeals.[136] Kodak's actual aftermarket power depended on the proportion of myopic consumers (who did not take future costs effectively into account) to their more sophisticated counterparts (who did account for these costs) as well as on the competitive conditions in the primary copier market. As the proportion of its sophisticated consumers increased, for instance, Kodak would have needed to dissipate more of its aftermarket profits to keep the copiers attractive to this group in the primary market.[137]

The ultimate welfare loss from the exercise of aftermarket power in this and similar situations therefore depends, first, on the relative proportions of sophisticated to myopic consumers and, second, on the intensity of primary market competition. However, in contrast to the prediction of rationality-based analyses, such as the one promoted by the *Kodak* dissent, a potentially significant loss to efficiency remains even when the primary market is fully competitive so long as the machines sold in the primary market are subsidized by the aftermarket, with an overconsumption in the former and underconsumption in the latter.[138] Hence even firms facing more competitive conditions—such as Kodak did in the copier market—may benefit from exploiting boundedly rational consumers.[139]

---

134. *See, e.g.*, Bailey, *supra* note 81, at 5–7.

135. *See supra* notes 66–74 and accompanying text (discussing the role of the rationality assumption in *Kodak*).

136. Eastman Kodak Co. v. Image Technical Servs., Inc., 504 U.S. 451, 485–86 (1992).

137. *See id.* (recognizing that it could be the case that the primary market at issue disciplined the aftermarket).

138. Bennett et al., *supra* note 7, at 135 n.25 (discussing the effects of aftermarket exploitation).

139. *See* OFFICE OF FAIR TRADING, *supra* note 10, §§ 1.1–.12 (finding that, in contrast to predictions of standard economic theory, the way prices are presented to consumers—or "framed"—affects consumer decision making and can cause consumer welfare losses); Oren Bar-Gill, *Competition and Consumer Protection: Behavioral Economics Account*, *in* SWEDISH COMPETITION AUTH., THE PROS AND CONS OF CONSUMER PROTECTION 12, 14, 25–27 (2012), *available at* http://www.konkurrensverket.se/upload/Filer/Trycksaker/Rapporter/Pros&Cons/

In the area of vertical restraints of trade, antitrust scholars have argued that product bundling and tying may exert more powerful effects on consumer behavior than traditional models acknowledge.[140] Behavioral analysts suggest, for example, that consumer inertia, the endowment effect, and the impact of defaults on consumer choice all indicate that consumers may find it difficult to switch even where the objective costs of switching from one product to another are small.[141] Dominant firms thus can use bundling, tying, and similar devices more effectively to foreclose competition than would be the case if consumers were strictly rational.[142] In the same vein, both theoretical arguments and experimental tests suggest that rebate schemes and other loyalty programs have stronger effects on the behavior of real consumers than traditional antitrust models expect them to have.[143]

Notably, the potential susceptibility of consumers to behavioral manipulation by firms will not always advantage monopolists or dominant firms. The stickiness of consumer behavior often redounds to these incumbents' benefit because new entrants and smaller competitors may find it more difficult to attract consumers based only on lower price or higher quality than standard models predict. Yet multiproduct firms with a smaller

---

rapport_pros_and_cons_consumer_protection.pdf (noting that firms may exploit the misperceptions of rational consumers despite the presence of competition); Bennett et al., *supra* note 7, at 121 ("[B]ehavioral economics suggests that even small switching costs [due to tying and bundling practices] can have significant effects on consumer behavior in the presence of consumer inertia, endowment effects, and default bias.").

140. *See, e.g.*, Bennett et al., *supra* note 7, at 121–22; Stucke, *Monopolization*, *supra* note 7, at 564–67; Petit & Neyrinck, *supra* note 7, at 7–11.

141. *See, e.g.*, Bennett et al., *supra* note 7, at 121; Stefano DellaVigna, *Psychology and Economics: Evidence from the Field*, 47 J. ECON. LITERATURE 315, 322–23 (2009) (explaining that transaction costs can be insufficient to account for default effects); Stucke, *Monopolization*, *supra* note 7, at 564–67 (discussing the role of default bias in recent cases); Petit & Neyrinck, *supra* note 7, at 9–10 (noting that EU competition law "accommodates such scenarios of 'psychological' bundling" and discussing Microsoft's pre-installation of Windows Media Player as an example of anticompetitive bundling without "coerc[ion] in the economic or technical sense" (emphasis omitted)).

142. Bennett et al., *supra* note 7, at 121; DellaVigna, *supra* note 141; Stucke, *Monopolization*, *supra* note 7, at 564–67; Petit & Neyrinck, *supra* note 7, at 9–10; *see also* Nicholas Economides & Ioannis Lianos, *The Elusive Antitrust Standard on Bundling in Europe and in the United States in the Aftermath of the* Microsoft *Cases*, 76 ANTITRUST L.J. 483, 544 (2009) (discussing efforts by firms to coerce intermediaries into rejecting competing bundles, thereby excluding rivals and limiting consumer choice).

143. *See* Martin Beckenkamp & Frank P. Maier-Rigaud, *An Experimental Investigation of Article 82 Rebate Schemes*, 2 COMPETITION L. REV. (SPECIAL ISSUE) 1 (2006) (presenting experimental evidence that rebate and discount schemes "exert a significant attraction" that enhances their potentially exclusionary effect beyond standard theoretical predictions); Alexander Morell et al., Sticky Rebates: Target Rebates Induce Non-Rational Loyalty in Consumers 22 (Feb. 2013) (unpublished manuscript), *available at* http://www.coll.mpg.de/pdf_dat/2009_23online.pdf (finding that "loyalty rebates induce a stickiness effect" in that they impede rational switching by consumers); *see also* Economides & Lianos, *supra* note 142 (discussing how firms can use practices such as rebates to influence consumer choices).

share in one market but sufficiently deep pockets otherwise may profitably expend resources on shaping consumer behavior and, consequently, exert greater competitive pressure on incumbents than extant models assume.[144]

In the case of vertical price restraints, both the historical evidence and the behavioral literature reveal that some manufacturers excessively impose RPM, when it is legal, on their retailers.[145] Manufacturers are prone to error with respect to vertical price restraints due to a confluence of behavioral phenomena. Judgmental biases—including anchoring, availability, and representativeness—lead them to overestimate the expected harms of retailer price cutting.[146] Loss aversion and fairness-driven behavior further make manufacturers averse to price cutting,[147] and they also find the direct price control offered by RPM an excessively attractive response to price cutting.[148] Moreover, RPM makes it particularly difficult for manufacturers to learn from experience whether it is in fact an efficient practice for their distribution system.[149] Notably, however, additional analysis reveals that even while boundedly rational RPM is costly for manufacturers and their retailers, the practice raises antitrust concerns only in those limited circumstances where it also harms the competitive process, such as when it is employed by firms with market power or is pervasive in an industry.[150]

Apart from this important lesson with respect to manufacturer behavior and RPM, the behavioral evidence is also informative regarding the impact of this practice on consumers. We have seen that economic scholars and the Court rejected manufacturers' loss-leader concerns because even deep discounts should not change rational perceptions of the quality of standard consumer goods.[151] However, behavioral marketing research long has identified a persistent positive relationship between price and perceptions of quality, in both the laboratory and the field, for a broad range of products.[152] Thus even while showing that manufacturers tend to use RPM excessively,

---

144. *But see* Bennett et al., *supra* note 7, at 119 (noting only the potential hindrance bounded rationality of consumers poses to dynamic competition but neglecting its potentially procompetitive effects).

145. Tor & Rinner, *supra* note 7, at 839–47.

146. *Id.* at 822–29. *But see infra* note 152 and accompanying text (discussing evidence for the effect of loss-leader practices on perceptions of the quality of manufacturers' products).

147. *Id.* at 829–33.

148. *See id.* at 833–39.

149. *Id.* at 842–47.

150. *See id.* at 839–54 (explaining that RPM is costly and unprofitable in many circumstances).

151. *Id.* at 813; *cf.* Marvel, *supra* note 76.

152. *See generally* Donald R. Lichtenstein et al., *Price Perceptions and Consumer Shopping Behavior: A Field Study*, 30 J. MARKETING RES. 234, 236 (1993) (describing "a positive association between price and perceived quality"); Valarie A. Zeithaml, *Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence*, J. MARKETING, July 1988, at 2, 11 ("Price reliance is a general tendency in some consumers to depend on price as a cue to quality.").

the empirical behavioral evidence at least partly confirms one reason for the longstanding resistance of these market participants to loss-leader practices, economists' disbelief notwithstanding.

The implications of loss-leader effects on consumers for antitrust doctrine, however, are not necessarily in line with the manufacturers' familiar argument. For example, although discounts that diminish perceptions of quality harm the manufacturers and reduce consumer welfare, they may generate efficiency gains if the retail prices favored by manufacturers send exaggerated quality signals that would not survive retail competition absent RPM.

With respect to merger policy, commentators draw on empirical evidence from the corporate finance literature as well as on some behavioral findings to note that many mergers prove inefficient rather than profit maximizing as the agencies commonly assume.[153] Empirical studies found, for one, that mergers often diminish the market value of the acquiring firm,[154] and behavioral research long has suggested that some excess merger activity is driven by the optimistic overconfidence of managers.[155] Related, a number of scholars contend that merger-specific efficiencies—which parties routinely proffer in accordance with the horizontal merger guidelines in support of transactions that raise competitive concerns[156]—not only are difficult to substantiate but often fail to materialize.[157]

Yet, even among those who note the prevalence of inefficient mergers, opinions diverge as to whether this systematic deviation from standard models matters for antitrust. Some argue that an accounting for the overall efficiency of proposed mergers is outside antitrust law's prohibition of only

153. *See* Horton, *supra* note 7, at 493 & n.118, 497; Reeves & Stucke, *supra* note 2, at 1561–62; Stucke, *At the Gate*, *supra* note 2, at 573–75; Stucke, *Reconsidering*, *supra* note 7, at 155–56; Spencer Weber Waller, *Corporate Governance and Competition Policy*, 18 GEO. MASON L. REV. 833, 873–81 (2011); *see also* Roberto A. Weber & Colin F. Camerer, *Cultural Conflict and Merger Failure: An Experimental Approach*, 49 MGMT. SCI. 400 (2003) (describing one particularly overlooked potential cause of merger inefficiencies or failures: culture conflict).

154. *See* Sara B. Moeller et al., *Wealth Destruction on a Massive Scale? A Study of Acquiring-Firm Returns in the Recent Merger Wave*, 60 J. FIN. 757 (2005); Dennis C. Mueller, *The Finance Literature on Mergers: A Critical Survey*, *in* COMPETITION, MONOPOLY AND CORPORATE GOVERNANCE: ESSAYS IN HONOUR OF KEITH COWLING 161, 178 (Michael Waterson ed., 2003); David J. Ravenscraft & F.M. Scherer, *Life After Takeover*, 36 J. INDUS. ECON. 147 (1987); F.M. Scherer, *Some Principles for Post-Chicago Antitrust Analysis*, 52 CASE W. RES. L. REV. 5, 17–18 (2001).

155. Richard Roll, *The Hubris Hypothesis of Corporate Takeovers*, 59 J. BUS. 197, 197–201 (1986). The agency problems generated by rational, self-interested managerial behavior can similarly lead to excess, inefficient merger activity.

156. *See* 2010 MERGER GUIDELINES, *supra* note 30, § 10 (observing that merger-generated efficiencies can help to enhance competition and endorsing such mergers).

157. *See* Horton, *supra* note 7, at 493–94; Oldale, *supra* note 7, at 143; Reeves & Stucke, *supra* note 2, at 1561–62; Stucke, *At the Gate*, *supra* note 2, at 573–75; Waller, *supra* note 153, at 875–76; Weber & Camerer, *supra* note 153, at 400–01.

those mergers that are likely "substantially to lessen competition."[158] They also aver that the agencies already are skeptical regarding claims of merger-specific efficiencies.[159] Others counter that the evidence of prevalent inefficient mergers justifies a closer scrutiny by the agencies of transactions with potentially anticompetitive effects.[160] After all, merger policy seeks to balance the uncertain prospects of over- and underenforcement—that is, the risk of blocking efficient mergers versus the risk of allowing the consummation of anticompetitive ones.[161] Hence, these commentators assert, the risks of overenforcement diminish, and a greater emphasis on preventing anticompetitive mergers is warranted, if inefficient mergers indeed are prevalent.[162]

The competition among new entrants into markets and the impact of entry on incumbents' market power offer a final illustration in an area with significant implications across antitrust law. Prospective entry plays an important role in merger assessments because it can counteract the anticompetitive effects of increased market power that might otherwise follow a merger.[163] More generally, effective entry can prevent even firms with large market shares from exerting market power,[164] an essential

158. *See* Clayton Act § 7, 15 U.S.C. § 18 (2012).

159. *See* Oldale, *supra* note 7, at 143 ("[B]ehavioral economics reinforces what competition authorities always suspected about claims that a merger will generate efficiencies: that these should be treated with a degree of healthy scepticism."); *see also* Reeves, *supra* note 7, at 8 (observing that agencies may scrutinize mergers in relatively minute detail to ensure that the result will be efficient); Werden et al., *supra* note 7, at 130 ("[T]he U.S. enforcement agencies have articulated a skeptical view of the power of entry to prevent anticompetitive effects from mergers."); Farrell & Shapiro Interview, *supra* note 10 (discussing agency skepticism toward party-supplied merger simulations). Parties who seek the approval of their proposed merger have a clear, rational interest in overstating the merger's efficiency benefits, for which reason the agencies are skeptical of such efficiency claims. *Cf.* Daniel A. Crane, *Rethinking Merger Efficiencies*, 110 MICH. L. REV. 347 (2011) (criticizing the hostile approach of antitrust agencies to efficiency claims).

160. *See, e.g.*, Reeves & Stucke, *supra* note 2, at 1563; Stucke, *At the Gate*, *supra* note 2, at 575; Waller, *supra* note 153, at 881; *see also* Horton, *supra* note 7, at 501–02.

161. William E. Kovacic, *Assessing the Quality of Competition Policy: The Case of Horizontal Merger Enforcement*, 5 COMPETITION POL'Y INT'L 129, 130 (2009).

162. Reeves & Stucke, *supra* note 2, at 1560–63; Stucke, *At the Gate*, *supra* note 2, at 573–75, 583; Stucke, *Reconsidering*, *supra* note 7, at 155–56; Waller, *supra* note 153, at 881.

163. *See* 2010 MERGER GUIDELINES, *supra* note 30, § 9 ("The prospect of entry into the relevant market will alleviate concerns about adverse competitive effects only if such entry will deter or counteract any competitive effects of concern so the merger will not substantially harm customers.").

164. *See* Ball Mem'l Hosp., Inc. v. Mut. Hosp. Ins., Inc., 784 F.2d 1325, 1335 (7th Cir. 1986) (rejecting the argument that market share indicates market power even in the absence of entry barriers and stating that "the lower the barriers to entry, and the shorter the lags of new entry, the less power existing firms have"); Will v. Comprehensive Accounting Corp., 776 F.2d 665, 672 n.3 (7th Cir. 1985) ("Unless barriers to entry prevent rivals from entering the market at the same cost of production, even a very large market share does not establish market power."); William M. Landes & Richard A. Posner, *Market Power in Antitrust Cases*, 94 HARV. L. REV. 937, 950

element of monopolization and attempted monopolization, tying, exclusive dealing, and other illegal practices.[165]

Traditional models assume that entry will only take place when it maximizes entrants' profits, thus requiring a positive risk-adjusted net present value.[166] The empirical evidence on entry paints a very different picture, however, showing abundant entry that appears unjustified based on entrants' objective prospects for survival and profitability.[167] Studies further reveal two additional puzzling entry phenomena. First, entry appears rather insensitive to some (though not all) of the main economic predictors of expected future profitability, including the expected intensity of competition, certain entry barriers, and more.[168] Second, start-up entrants not only fail more frequently than their diversifying counterparts—a pattern that alone might have reflected merely the greater riskiness of their ventures—but do so to such an extent that they obtain lower expected payoffs and thus exhibit inferior average performance altogether.[169]

A behavioral analysis of entrants' judgments reveals, however, that these three puzzling phenomena largely correspond to the patterns of the psychology of optimistic overconfidence.[170] New entrants typically make their personally significant judgments of entry's prospects under conditions of extreme uncertainty.[171] In such circumstances, overoptimism and a number of related phenomena lead real entrants, as a group, to overestimate their prospects upon entry.[172] These processes, moreover, both reduce entrants' sensitivity to market predictors of success[173] and exert a

---

(1981) (explaining the relationship between entry, market share, and market power and proposing that a significant amount of market share does not necessarily indicate market power).

165. *See* PHILLIP E. AREEDA ET AL., 2B ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION ¶ 420b & n.11 (3d ed. 2007) ("Entry conditions are therefore relevant to assessing the market power required by most antitrust rules" because "[m]arket power bears on the anticompetitive potential of challenged conduct").

166. Tor, *Entry*, *supra* note 2, at 489 & nn.15–17, 490.

167. *See, e.g.*, JOHN R. BALDWIN, THE DYNAMICS OF INDUSTRIAL COMPETITION: A NORTH AMERICAN PERSPECTIVE 359 (1995) (stating that empirical evidence supports the life-cycle model rather than entry only when profits are maximized); P.A. Geroski, *Some Data-Driven Reflections on the Entry Process*, *in* ENTRY AND MARKET CONTESTABILITY: AN INTERNATIONAL COMPARISON 282, 295 (P.A. Geroski & J. Schwalbach eds., 1991). *See generally* Tor, *Entry*, *supra* note 2, at 490–92 (discussing this puzzle).

168. Tor, *Entry*, *supra* note 2, at 492–94.

169. *Id.* at 494–96.

170. *See id.* at 504 ("[A] wealth of psychological data show that in circumstances of this kind . . . people tend to exhibit a significant bias. . . . [T]his bias results from a number of psychological processes that affect entrants' judgments of both the *probability* and *value* of their prospective ventures.").

171. *Id.* at 487, 527 n.192.

172. *See id.* at 505–14 (offering an in-depth analysis of the impact of these phenomena on entrants).

173. *Id.* at 514–20.

differential impact on start-ups versus diversifying entrants that makes the former more biased when judging their entry prospects.[174]

The behavioral forces that shape entrants' judgments generate a competitive landscape that differs significantly from that envisioned by traditional antitrust models. Because all entry is not the same, more biased and numerous start-ups fail at far greater proportions than their diversifying competitors but still are overrepresented among those few entrants who ultimately survive and prosper.[175] Furthermore, insofar as new entry is associated with innovation, particularly for start-ups, boundedly rational entry may be socially beneficial overall despite its costs for entrants.[176]

Where the impact of entry on incumbents is concerned, the behavioral analysis of entry suggests, for example, that while entry often is not exceptionally difficult, post-entry success and survival are unlikely for most entrants.[177] Start-ups, and small entrants generally, rarely pose a short-term competitive threat to incumbents, but some large diversifying entrants possibly do.[178] In the long run, however, the few successful, often biased, innovative entrants are an important source of competitive pressure on incumbents.[179] These outcomes have important implications for antitrust law and policy. For one, they support the law's hostility to unnecessary restrictions on new entry, given its important procompetitive benefits.[180]

At the same time, the behavioral analysis of entry indicates that the law should be wary of relying on low entry barriers alone to guarantee short-run competitive pressure on incumbents.[181] In the area of predatory pricing, for instance, we saw that *Brooke Group* requires plaintiffs to show that the alleged predator had a rational prospect of recoupment, and we also saw that such recoupment is considered unlikely when entry barriers are low.[182] Our analysis suggests, however, that courts should not rely on mere evidence of low entry barriers to conclude that recoupment is unlikely. After all, a high rate of *overconfident* entry may be accompanied by very limited market penetration that does little to prevent such recoupment. Instead, courts should focus on the likely and actual past success of entrants

174. *See infra* text accompanying notes 214–17 (discussing the role of moderating variables in shaping the competition among new entrants and its market effects).

175. Tor, *Entry*, *supra* note 2, at 531–33.

176. *Id.* at 537–40.

177. *Id.* at 490–92; *see also id.* at 531–43, 548–49 (discussing a number of additional significant consequences of boundedly rational entry for competition and antitrust law).

178. *Id.* at 494–96.

179. *Id.* at 537–43.

180. *Id.* at 549–50.

181. *Id.* at 550–52.

182. Brooke Grp. Ltd. v. Brown & Williamson Tobacco Corp., 509 U.S. 209, 224–26 (1993); Tor, *Predatory Pricing*, *supra* note 2, at 55.

in penetrating the market as better indicators of the short-term competitive constraint on incumbents' potential recoupment.[183]

All in all, the preceding examples clearly show that behavioral antitrust already provides a wide range of analyses across the field. These analyses vary with respect to the type of evidence they draw on, how they apply behavioral findings to antitrust-relevant market phenomena, and the lessons they offer antitrust doctrine and policy. Regrettably, moreover, some of the contributions to this new literature as well as many of its critiques manifest a common, fundamental methodological error in behavioral antitrust analysis to which we now turn.

D. The Fundamental Methodological Error

The apparent tension between antitrust law's extensive reliance on the rationality assumption on the one hand and the behavioral focus on deviations from strict rationality on the other may account for the heated debate now taking place between supporters and detractors of behavioral antitrust.[184] It may also explain, moreover, the delayed penetration of the behavioral approach into antitrust scholarship as compared to most other legal fields.[185]

Yet even as antitrust increasingly takes notice of behavioral insights, a more subtle, but no less significant, tension lies beneath the explicit contrast between the behavioral and traditional economic perspectives on antitrust rationality. The extensive use of neoclassical economics has inculcated in the antitrust community a reliance on simplifying assumptions as analytical tools where rationality is concerned and beyond. Hypothetical assumptions play an important role, for example, in the central antitrust concept of market definition, which assists in determinations of market power and the effects of potentially anticompetitive conduct throughout antitrust law.[186]

---

183. Tor, *Entry*, *supra* note 2, at 553–55. The author also notes that while courts often reject predatory pricing allegations summarily where entry barriers appear low, they sometimes examine factors that are associated with entrants survival and penetration rather than mere entry. *Id.* For instance, *Brooke Group* itself relied on evidence of rapid expansion in the relevant segment that was partly due to successful penetration, 509 U.S. at 233–34, and so did the First Circuit in *R.W. Int'l Corp. v. Welch Food, Inc.*, 13 F.3d 478, 488 (1st Cir. 1994), citing the Court's *Brooke Group* ruling.

184. *See supra* note 7 and accompanying text.

185. *See, e.g.*, Arnaudo, *supra* note 5 ("[W]hen considering the growing fortunes of [behavioral economics], the process towards a [behavioral antitrust] could have been expected to occur much faster . . . ." (footnote omitted)); Stucke, *At the Gate*, *supra* note 2, at 514 ("While tossed against the rocks elsewhere, within the quiet waters of antitrust these rational choice theories stand largely unchallenged.").

186. *See* 2010 MERGER GUIDELINES, *supra* note 30, § 4 (offering guidelines for defining the relevant market and detailing the assumptions used); *see also* Jonathan B. Baker, *Market Definition: An Analytical Overview*, 74 ANTITRUST L.J. 129 (2007) (discussing the significance of the market-definition process in determining anticompetitive effect and how that process should best be conducted); Louis Kaplow, *Why (Ever) Define Markets?*, 124 HARV. L. REV. 437 (2010)

The market definition process helps practitioners and antitrust economists predict and explain to clients how an enforcement agency will determine whether a proposed merger is likely substantially to lessen competition under Section 7 of the Clayton Act.[187] More generally, hypothetical assumptions provide antitrust with the benefits of increased tractability, predictability, and conceptual clarity.[188]

Commentators long familiar with the powerful simplifying assumptions of traditional antitrust law and economics quite naturally approach behavioral antitrust in the same way.[189] Whether asserting its virtues or criticizing its shortcomings, these commentators routinely speak of a behavioral approach that "attacks the rational profit-maximizer assumption head on by *assuming* that humans have cognitive limitations that prevent them from processing information perfectly and maximizing their utility,"[190] "replac[es] the assumption of rationality with one of 'bounded rationality,'"[191] or relies on an "irrationality hypothesis."[192]

Importantly, such statements reflect not merely casual, inaccurate usage, but rather a fundamental methodological error that permeates the recent behavioral antitrust discourse. When treating concrete, empirical behavioral findings as if they were broad, hypothetical propositions in the mold of the familiar rationality assumption, antitrust commentators

---

(noting the prominent role of the market-definition process in competition law cases and arguing that it should be abandoned).

187. *See* Baker, *supra* note 186, at 130–31; Gregory J. Werden, Why (Ever) Define Markets? An Answer to Professor Kaplow 1, 9–14 (Feb. 13, 2012) [hereinafter Werden, An Answer to Professor Kaplow] (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2004655. And while there may be disagreements—sometimes significant ones—among scholars or litigating parties on how to define a given product market, the concept itself is commonly understood and so are the more technical tests associated with it (such as those of the hypothetical monopolist and critical loss).

188. *See, e.g.*, Salinger, *supra* note 7, at 67 ("The rationality assumption plays so prominently in the literature because it is tractable . . . and yields some quite accurate predictions."); *see also* HOVENKAMP, ANTITRUST ENTERPRISE, *supra* note 3, at 7, 31–34 (discussing the Chicago School's "twin propositions that markets are relatively simple and tend naturally toward competitive outcomes"); Richard A. Posner, *The Chicago School of Antitrust Analysis*, 127 U. PA. L. REV. 925, 931 (1979) (describing the benefits of the Chicago School economics' "powerful simplifications"); *cf.* Werden et al., *supra* note 7, at 126 (concluding that "competition policy should continue to rely on neoclassical economic analysis based on the assumption of profit maximization").

189. A similar pattern can be found in other fields, particularly those where traditional economic analysis plays a central role. *See, e.g.*, H. Kent Baker & John R. Nofsinger, *Behavioral Finance: An Overview*, *in* BEHAVIORAL FINANCE: INVESTORS, CORPORATIONS, AND MARKETS 3, 3 (H. Kent Baker & John R. Nofsinger eds., 2010) [hereinafter BEHAVIORAL FINANCE] ("An underlying *assumption* of behavioral finance is that the information structure and the characteristics of market participants systematically influence individuals' investment decisions as well as market outcomes." (emphasis added)).

190. Reeves, *supra* note 7, at 2 (emphasis added).

191. Cooper & Kovacic, *supra* note 7, at 780.

192. Wright & Stone, *supra* note 7, at 1523.

misconceive the nature of the empirically based behavioral approach. This confusion of hypothesis for evidence is not always benign, at times leading otherwise sophisticated scholars to make three distinct classes of mistakes, each with its attendant erroneous applications and policy conclusions in behavioral antitrust.

II. The First Mistake: Assuming Constant and Uniform Bounded Rationality

Commentators make the first mistake when they erroneously equate behavioral antitrust with an *assumption of a constant and uniform set of deviations from strict rationality.* The tendency to assume constant bounded rationality leads antitrust scholars to pay little attention to the specific contours and boundaries of behavioral phenomena, while the assumed uniformity results in a failure to account for the heterogeneity of actors' behavior—both among similarly situated actors and for the same actor in different circumstances or with respect to different behavioral phenomena. This Part explains the significance of the limited constancy and uniformity of behavioral patterns and illustrates the problematic consequences of their neglect for behavioral analyses of antitrust.[193]

A. Variability, Not Constancy

In sharp contrast to the constancy of hypothetical strict rationality, the bounded rationality of real antitrust actors has specific empirical contours and boundaries. Different behavioral phenomena are manifested more strongly in some circumstances and more weakly in other situations, at times disappearing altogether.[194] Moreover, all behavioral phenomena are not created equal: some are more robust and pervasive while others exert significant impact on behavior only under limited conditions.[195] To determine whether and how these phenomena are likely to impact the behavior of antitrust actors one must therefore carefully attend to their boundaries and limits.

Most obviously, the proper application of behavioral phenomena requires their accurate understanding. Yet one finds within the extensive behaviorally oriented legal literature—mostly in other areas but now in antitrust as well—analyses that confuse different findings with one another,

---

193. This Part builds significantly on the more comprehensive review of the behavioral evidence showing variability and heterogeneity and its implications across the laws in Avishalom Tor, Law for a Behaviorally-Complex World (unpublished manuscript) (on file with author).

194. *See* Tor, *Behavioral Methodology*, *supra* note 1, at 292–96 (explaining the significance of boundary conditions for behavioral analyses of law).

195. *See id.* at 293 (recognizing that "processes of judgment and choice depend on the environment within which people operate"); *see also* Rachlinski, *supra* note 16 (comparing the broader set of circumstances under which overoptimism is manifested with the narrower boundaries of ambiguity aversion).

mix multiple phenomena together, or simply mischaracterize the empirical evidence.[196] Unsurprisingly, such mistakes occasionally lead legal researchers to erroneous conclusions.[197] Importantly, moreover, these confusions often follow a failure to engage the empirical behavioral evidence directly. Analysts instead rely on second- or third-hand accounts, primarily within the legal literature, of behavioral findings.[198]

Beyond such basic confusions, however, scholars who accurately understand behavioral findings may still fail to appreciate the significance of the contours of the empirical evidence for legal analysis. For example, one recent study of the historical effects of the decision in *United States v. Paramount Pictures, Inc.*[199] on the business model of the film industry argues in passing that, from a behavioral perspective, ambiguity aversion could explain the industry's reliance on relational—instead of formal—vertical contracting in situations characterized by extreme uncertainty.[200] The study's author seems to suggest that in such situations ambiguity-averse decision makers avoid formal contracting, preferring instead more open-ended relational contracts.[201] While intuitively plausible, this argument neglects to account for the contours of ambiguity aversion, which studies show is largely comparative rather than absolute: decision makers prefer a well-defined risk to an ambiguous one but routinely take ambiguous risks when the former option is unavailable.[202] In the film industry, however, ambiguity is so pervasive that an aversion to it is

196. *See, e.g.*, Wright & Stone, *supra* note 7, at 1530 (describing heuristics as "loose categories" and stating that Kahneman and Tversky's prospect theory—which is a theory of choice, not judgment—"grouped irrational behaviors together" within three categories of heuristics—which are judgment rather than choice phenomena—referring to framing effects as "biases," and more).

197. *See, e.g.*, *id.* at 1552 (concluding that the flaw of behaviorally informed antitrust is the uncertainty it introduces with regard to the predictive power of enforcement).

198. *See, e.g.*, Roger Van den Bergh, *Behavioral Antitrust: Not Ready for the Main Stage*, 9 J. COMPETITION L. & ECON. 203, 215–16 (2013) (citing Reeves & Stucke, *supra* note 2, to support assertions about behavioral biases). While this tendency may have resulted in part from legal scholars' unfamiliarity with behavioral research methods, it also reflects the common confusion between broad hypothetical assumptions (and other logical arguments) and concrete empirical evidence.

199. 334 U.S. 131 (1948).

200. Ryan M. Riegg, *Opportunism, Uncertainty, and Relational Contracting—Antitrust in the Film Industry*, 6 U. DENV. SPORTS & ENT. L.J. 107, 124–25 (2009).

201. *See id.*

202. *See* Craig R. Fox & Amos Tversky, *Ambiguity Aversion and Comparative Ignorance*, 110 Q.J. ECON. 585 (1995) (proposing the comparative ignorance hypothesis and finding that ambiguity aversion disappears when a person evaluates either a clear or vague prospect in isolation); Craig R. Fox & Martin Weber, *Ambiguity Aversion, Comparative Ignorance, and Decision Context*, 88 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 476 (2002) (expanding on the comparative ignorance hypothesis through four experiments). *But see* Clare Chua Chow & Rakesh K. Sarin, *Comparative Ignorance and the Ellsberg Paradox*, 22 J. RISK & UNCERTAINTY 129, 129, 138 (2001) (showing that ambiguity aversion sometimes remains in noncomparative settings, although it is significantly stronger in the comparative setting).

unlikely to play a significant role in the selection of vertical contracting practices, unlike concerns about risk, opportunistic behavior, and more that may well favor relational contracting in this setting.[203]

Another, more significant, illustration of the need to account for both the contours and boundaries of behavioral phenomena is some analysts' argument that market participants routinely avoid making rational entry attempts due to boundedly rational risk aversion.[204] Initially, the argument appears straightforward: once a profitable entry opportunity has been identified, potential entrants must decide whether to invest resources in the uncertain prospect of entry. Behavioral research shows that decision makers tend to be risk averse beyond the dictates of rationality when faced with prospects that are potentially beneficial vis-à-vis the status quo ("gains," in prospect theory parlance).[205] Therefore, so the argument goes, real entrants routinely avoid positive-net-present-value entry opportunities that would have attracted their hypothetical, strictly rational—and thus less risk averse—counterparts.[206]

Yet not only is the assertion of common risk aversion among potential entrants at odds with extensive empirical findings from industrial-organization research,[207] but a closer examination also reveals that it is not supported by the behavioral evidence either, for a number of reasons.[208] First, new entry entails not only the uncertain prospect of a gain compared to the status quo but also a significant possibility of a loss if entry fails. In other words, entrants face a mixed gain/loss gamble rather than one involving gains alone. Yet decision makers' reluctance to take such mixed gain/loss gambles primarily is a manifestation of loss aversion, not risk

---

203. Moreover, research shows that comparative ignorance is an important driver of this phenomenon. Individuals are less concerned about shared ambiguity but are averse to it when their counterparties have superior knowledge about the relevant decision. *See* Fox & Weber, *supra* note 202, at 476–77; *see also* Fox & Tversky, *supra* note 202, at 599 (finding that an "uncertain prospect becomes less attractive when people are made aware that the same prospect will also be evaluated by more knowledgeable individuals"). Yet as the study's author notes, citing a famous screenwriter, "nobody knows anything" about what makes a movie a hit or not. Riegg, *supra* note 200, at 129. Hence ignorance is largely shared, and ambiguity aversion is not a likely force with respect to the most significant factor in this contracting environment.

204. *See, e.g.*, Stucke, *At the Gate*, *supra* note 2, at 563–72; Stucke, *New Antitrust Realism*, *supra* note 7, at 6–11; Petit & Neyrinck, *supra* note 7, at 4–5. Potential entrants may be rationally risk averse in some situations, although traditional economic analysis assumes firms to be risk-neutral decision makers. *See supra* notes 104–07 and accompanying text.

205. Kahneman & Tversky, *Prospect Theory*, *supra* note 90, at 268–69.

206. *See, e.g.*, Stucke, *At the Gate*, *supra* note 2, at 569–72; Stucke, *New Antitrust Realism*, *supra* note 7, at 6–11; Petit & Neyrinck, *supra* note 7 at 4–5.

207. Tor, *Entry*, *supra* note 2, at 488–501.

208. Entrants may sometimes shun potentially profitable entry opportunities for a variety of traditional economic and behavioral reasons. Stucke, *New Antitrust Realism*, *supra* note 7, at 8–10. The present discussion only illustrates the limited explanatory power of framing and reference points in this case.

aversion.[209] Because losses are felt more strongly than comparable gains, individuals require an expected payoff that is significantly higher than what is needed merely to generate positive expected value to make up for their potential painful loss.[210] A reluctance to embark upon new entry that is born in loss aversion, however, not only differs from standard risk aversion in its psychological roots but also has different behavioral contours.[211]

Most notably, potential entrants are likely to exhibit risk seeking instead of risk aversion because they usually do not consider the prospects of entry in the abstract but rather contemplate a specific venture. Hence they tend to compare the various possible outcomes of entry to the successful outcome they hope to achieve, perceiving those outcomes that fall short of their aspiration as undesirable ones, which generates loss aversion.[212] Such loss-averse entrants, however, will embark on far riskier ventures than their risk-averse counterparts would be willing to undertake.[213]

The role of framing and reference points in shaping entrants' risk preferences thus illustrates why antitrust analysts should consider the specific contours and boundaries of the behavioral phenomena they apply. Upon closer scrutiny, the same effects of framing that initially seem to

---

209. The seminal article that introduced prospect theory uses a simple mixed gamble example to illustrate loss aversion. *See* Kahneman & Tversky, *Prospect Theory*, *supra* note 90, at 263–64, 279; *see also* Sabrina M. Tom et al., *The Neural Basis of Loss Aversion in Decision-Making Under Risk*, 315 SCI. 515, 515–18 (2007) (describing neural imaging evidence for loss aversion in mixed gambles); Amos Tversky & Daniel Kahneman, *Advances in Prospect Theory: Cumulative Representation of Uncertainty*, 5 J. RISK & UNCERTAINTY 297 (1992) (offering evidence of loss aversion in mixed gambles).

210. *See* Kahneman & Tversky, *Prospect Theory*, *supra* note 90, at 285.

211. For one, loss-averse entrants will be more reluctant to enter than entrants who are merely risk averse. *Cf.* Matthew Rabin, Comment, *Risk Aversion and Expected-Utility Theory: A Calibration Theorem*, 68 ECONOMETRICA 1281, 1288 (2000) (showing that standard risk-aversion cannot explain commonly observed, more extreme instances of risk-averse behavior, which is driven by loss-aversion). The former entrants also will respond more to changes in the *magnitude* of the potential loss they would face if their entry attempt were to fail compared to changes in its *probability* instead of simply adjusting the net present value of entry to their degree of risk aversion as merely risk-averse entrants would do. *See, e.g.*, George Wu & Alex B. Markle, *An Empirical Test of Gain-Loss Separability in Prospect Theory*, 54 MGMT. SCI. 1322, 1332 (2008) (showing that when choosing between mixed gambles, individuals are less sensitive to differences in the probabilities of potential outcomes).

212. *See* Johannes Abeler et al., *Reference Points and Effort Provision*, 101 AM. ECON. REV. 470, 487 (2011) (showing, experimentally, how expectations impact real effort provision and contrasting risk-averse behaviors with loss-averse behavior); Chip Heath et al., *Goals as Reference Points*, 38 COGNITIVE PSYCHOL. 79, 93 (1999) (presenting evidence that goals both function as reference points and exhibit the properties of loss aversion and diminishing sensitivity).

213. The present analysis focuses only on the risk attitudes as an illustration, while in fact entrants are likely to be risk seeking due to the contribution of judgmental biases. *See* Tor, *Entry*, *supra* note 2, at 503–31 (describing the relevant evidence at length).

make potential entrants risk averse in fact may facilitate loss-averse, risk-seeking entry.

Moreover, the case of entry highlights the importance of accounting not only for the basic contours and boundaries of behavioral phenomena but also for the key variables that moderate their effects on market participants.[214] We saw that a behavioral analysis of entrants' judgments of the prospects of entry helps explain a series of otherwise puzzling empirical findings regarding patterns of new entry into markets.[215] We further saw that the variables that moderate optimistic overconfidence help explain the inferior average performance of start-up entrants compared to their diversifying counterparts.[216] Two such variables in particular—the intensity of preferences and the ambiguity of the decision environment—systematically lead start-up entrants to exhibit more biased judgments of their prospects than those manifested by diversifying entrants.[217]

While these findings bear important implications for the competition among entrants, for entry's effects on incumbent firms in the market, and for various antitrust rules, they also reveal the necessity for behavioral antitrust scholars to consider the effect of moderating variables on those market behaviors they study.[218] Without attending to the effects of preference intensity and ambiguity on the competition among entrants, behaviorally informed analysts might erroneously expect excess entry to be more effective than it is in disciplining incumbents[219] or mistakenly believe that the cohort of *ex post* successful entrants resembles the pool of those attempting entry *ex ante*.

B. Heterogeneity, Not Uniformity

Much like they neglect the variability of empirical behavioral phenomena, antitrust commentators frequently fail to appreciate the

---

214. A moderator variable affects the direction and/or strength of the relationship between two other variables. Reuben M. Baron & David A. Kenny, *The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations*, 51 J. PERSONALITY & SOC. PSYCHOL. 1173, 1174 (1986).

215. *See* Tor, *Entry*, *supra* note 2, at 503–31.

216. *See id.* at 487, 520–31.

217. *See id.*; *see also* Arnold C. Cooper et al., *Entrepreneurs' Perceived Chances for Success*, 3 J. BUS. VENTURING 97, 103 (1988) (finding that "entrepreneurs' perceptions of their own odds for success display a noteworthy degree of optimism"); Ken G. Smith et al., *Decision Making Behavior in Smaller Entrepreneurial and Larger Professionally Managed Firms*, 3 J. BUS. VENTURING 223, 223 (1988) (finding that entrepreneurs are less likely to follow a formal rational, decision process than established firms).

218. *See* Tor, *Entry*, *supra* note 2, at 520–31.

219. *See* Wright & Stone, *supra* note 7, at 1541 ("[T]he existence of irrationally optimistic potential entrants policing for the existence of supracompetitive profits, and even entering in their absence from time to time, reduces the incentive to engage in all sorts of anticompetitive behavior.").

heterogeneity of human behavior.[220] Instead, they assume population-level uniformity, both among different actors who are similarly situated and for the same actor across different circumstances and different behavioral phenomena.[221]

Yet in reality human judgment and decision behavior is highly heterogeneous. Different antitrust actors will manifest different deviations from strict rationality depending on factors such as cognitive ability,[222] thinking style,[223] risk-taking propensity,[224] personality traits,[225] and more.[226] Notwithstanding this evidence for systematic individual differences in specific behavioral phenomena, however, the correlation *within* individuals among different deviations from rationality generally is small and so is the proportion of the overall variance in behavior that systematic individual differences account for.[227] Moreover, people exhibit particular behavioral

220. *But see* CHRISTOPH ENGEL, GENERATING PREDICTABILITY: INSTITUTIONAL ANALYSIS AND DESIGN 1–10 (2005) (describing the heterogeneity of human judgment and decision behavior as a challenge for the predictability needed for human interaction generally and policy and institutional design more specifically).

221. *See, e.g.*, Stucke, *Reconsidering*, *supra* note 7, at 121–22; Wright & Stone, *supra* note 7, at 1537.

222. *See, e.g.*, Edward T. Cokely & Colleen M. Kelley, *Cognitive Abilities and Superior Decision Making Under Risk: A Protocol Analysis and Process Model Evaluation*, 4 JUDGMENT & DECISION MAKING 20 (2009) (finding that individual differences in cognitive abilities and related skills can systematically predict normatively superior and logically consistent judgments and decision making); Keith E. Stanovich & Richard F. West, *Individual Differences in Framing and Conjunction Effects*, 4 THINKING & REASONING 289 (1998) (discussing the implications of the finding that subjects with higher cognitive abilities were disproportionately likely to avoid potential framing and conjunctive fallacies).

223. *See, e.g.*, Richard F. West et al., *Heuristics and Biases as Measures of Critical Thinking: Associations with Cognitive Ability and Thinking Dispositions*, 100 J. EDUC. PSYCHOL. 930, 930 (2008) (discovering that "[m]easures of thinking dispositions" including "actively open-minded thinking and need for cognition" actually predicted "variance in . . . classes of critical thinking skills after general cognitive ability had been controlled").

224. *See, e.g.*, Kevin T. Mahoney et al., *Individual Differences in a Within-Subjects Risky-Choice Framing Study*, 51 PERSONALITY & INDIVIDUAL DIFFERENCES 248 (2011) (utilizing risk style and thinking style to predict individual differences in response to framing problems).

225. *See, e.g.*, Marco Lauriola & Irwin P. Levin, *Personality Traits and Risky Decision-Making in a Controlled Experimental Task: An Exploratory Study*, 31 PERSONALITY & INDIVIDUAL DIFFERENCES 215 (2001) (exploring the "big five" personality traits and their respective correlations to risky decision making); Irwin P. Levin et al., *A New Look at Framing Effects: Distributions of Effect Sizes, Individual Differences, and Independence of Types of Effects*, 88 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 411, 425 (2002) (finding personality traits "predictive of the magnitude of framing effects" and concluding that the experimental design provided evidence that individual differences in framing effects are linked to stable personality characteristics, such as conscientiousness and agreeableness).

226. *See* Ann-Renée Blais et al., *Individual Differences in Decision Processing and Confidence Judgments in Comparative Judgment Tasks: The Role of Cognitive Styles*, 38 PERSONALITY & INDIVIDUAL DIFFERENCES 1701 (2005) (finding that stable individual differences in decision time, accuracy, and response confidence emerged across all comparative judgment tasks, although the basis for these differences remained elusive).

227. *See* Kirstin C. Appelt et al., *The Decision Making Individual Differences Inventory and Guidelines for the Study of Individual Differences in Judgment and Decision-Making Research*, 6

phenomena to different degrees at different times in different contexts.[228] Hence some antitrust actors will better approximate the assumptions of rationality-based models, or deviate from them, on some occasions, while other actors will do so in other situations or with respect to different phenomena.[229] Indeed, those robust, systematic, and predictable deviations from rationality that are documented at the *population* level do not reflect individual-level uniformity but rather are the aggregation of significant individual-level heterogeneity in judgment and decision behavior.

Antitrust analyses that disregard the heterogeneity of market behavior may misconstrue the reasons for and the consequences of competitive and anticompetitive practices alike. The behavioral analysis of RPM revealed, for example, that some manufacturers excessively impose this vertical restraint in their distribution systems when it is legal, to their own detriment and at a cost to some of their retailers, and may only learn of their costly mistake after extended periods of time.[230] Commentators neglecting the heterogeneity of manufacturer behavior mistakenly suggested that this behavioral finding reveals an additional anticompetitive harm of RPM beyond those identified by traditional rationality-based analyses, a harm that could support a return to the now-discarded rule of per se illegality.[231]

JUDGMENT & DECISION MAKING 252, 253, 257 (2011) (noting inconsistent results for individual difference measures in judgment and decision making and attributing them in part to the greater impact of situational variables, "which can overwhelm any impact of individual differences"); Wim De Neys & Jean-François Bonnefon, *The 'Whys' and 'Whens' of Individual Differences in Thinking Biases*, 17 TRENDS COGNITIVE SCI. 172, 172 (2013) (discussing different approaches to research in this area and explaining that the "accounts we survey . . . [do] not entail that different reasoners cannot be biased for different reasons or that the same reasoner is always biased for the same reasons. Obviously, the locus of individual differences *need not be fixed and can be contingent on specific task, context, person*, or developmental factors" (emphasis added)).

228. *See, e.g.*, N.S. Fagley & Paul M. Miller, *Framing Effects and Arenas of Choice: Your Money or Your Life?*, 71 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 355, 355, 370 (1997) (noting that although "[t]here was a significant sex by frame interaction such that only women exhibited framing effects on choice," the authors found that "[subjects] made more risky choices when outcomes involved human lives than when they involved money, regardless of [frame]"); Rui Mata et al., *Age Differences in Risky Choice: A Meta-Analysis*, 1235 ANNALS N.Y. ACAD. SCI. 18 (2011) (using a systematic literature review to discover that age-related differences varied considerably based on the task at hand); *cf.* Gregory Mitchell, *Why Law and Economics' Perfect Rationality Should Not Be Traded for Behavioral Law and Economics' Equal Incompetence*, 91 GEO. L.J. 67, 98–99 (2002) (reviewing some empirical evidence, with an emphasis on the impact of changes in affect, and suggesting that "the propensity to act rationally varies not only across individuals but also within individuals over time"). *See generally* Appelt et al., *supra* note 227, at 257 (advocating a "person-by-decision-and/or-situation interaction approach that examines how individual differences interact with other individual differences, with decision features, and with situational factors to influence behavior in a given context").

229. Moreover, behavior that deviates further from the assumptions of rationality is not always associated with inferior performance in the market. *See infra* section III(A)(1).

230. Tor & Rinner, *supra* note 7, at 839–42.

231. *See, e.g.*, Ginsburg & Moore, *supra* note 7, at 98 (suggesting that the evidence of boundedly rational RPM "is of greater relevance to a legislature considering whether to make

Others offered to designate RPM, in light of the behavioral evidence, a presumptively illegal practice that courts could dispose of with a "quick look."[232] Yet the significant heterogeneity of manufacturer behavior, where boundedly rational uses of RPM coexist with other rationally procompetitive and rationally anticompetitive instances of the practice, makes per se illegality inappropriate here.[233] Behavioral antitrust in fact supports the *Leegin* Court's overruling of earlier precedents and embrace of a rule of reason approach, even while it highlights the need for a structured rule of reason for RPM that also accounts for behavioral regularities as opposed to the alternative of an open-ended, unstructured rule of reason.[234]

More recently, some analysts have begun considering explicitly the implications of systematic differences between classes of antitrust actors.[235] By routinely falling prey to the fundamental methodological error, however, these commentators still tend to reach erroneous conclusions. For instance, one behavioral antitrust scholar sought to determine how the law should respond to the interaction among firms, consumers, and the government, assuming that each of these different classes of actors is either uniformly "rational" or uniformly "boundedly rational."[236] Intersecting these two alternative assumptions with respect to the three sets of actors, this scholar reached a series of conclusions regarding the consequences of, say, a "boundedly rational" government responding to "rational" firms' exploitation of "boundedly rational" consumers[237] or of another, strange, hypothetical world in which consumers are "rational" but firms are "boundedly rational."[238]

Because of this assumed uniformity, many otherwise interesting and potentially valuable observations that are made throughout this scholar's analysis—whether with respect to consumers, firms, or the government's role—are simultaneously too broad and too narrow. Take for instance the argument that a "rational" government that seeks to respond to the possible

---

resale price maintenance illegal *per se*" even while noting that courts making decisions in specific cases must determine whether a particular instance of RPM is anticompetitive).

232. Reeves & Stucke, *supra* note 2, at 1582 n.341.

233. Tor & Rinner, *supra* note 7, at 854–55.

234. *See id.* at 855–63 (arguing that the presence of excessive or inefficient use of RPM should be part of the rule of reason inquiry after *Leegin* since market forces are slow to eliminate such use and finding the factors enumerated in *Leegin* to be relevant to this inquiry); *see also infra* subpart V(B).

235. See *infra* Part III for an examination at length of the distinction that scholars increasingly make between firms and consumers following the intuition that firms are sophisticated organizations that benefit from experience and expertise, with advantages consumers usually lack that allow exploitation.

236. Stucke, *Reconsidering*, *supra* note 7, at 121–22. Strikingly, Stucke explicitly acknowledges variability and heterogeneity even while failing to realize that treating bounded rationality as a broad, universal assumption could lead to mistaken conclusions. *See id.* at 122.

237. *Id.* at 144–53.

238. *Id.* at 154–62.

exploitation of "boundedly rational" consumers by "rational" firms must be concerned with factors such as the identification of consumer preferences, the impact of defaults set by the government, or the negative effects of intervention on consumer autonomy, while balancing these against the need to protect consumers from "corporate autocracy."[239] Some of these concerns merit serious consideration in antitrust and regulatory policy, but they do not always apply when "rational" firms and government face "boundedly rational" consumers. Nor are the enumerated concerns limited to this specific hypothetical juxtaposition of a differing but uniform rationality of the three classes of antitrust actors. "Boundedly rational" firms with superior resources and information sometimes may also exploit consumers, whether or not the latter are "rational." A "boundedly rational" government that sets defaults still can shape consumer behavior or encroach on consumer autonomy, and so on. Instead, a clearer and more fruitful approach would recognize the inevitable bounded rationality of all classes of actors. It would then seek to account for the variability and heterogeneity of behavior both among the different classes and within each class—consumers, firms, and government actors—and develop policy prescriptions based on the empirical evidence most relevant to the question at hand.[240]

Similarly, erroneous conclusions plague some analyses that aim to criticize the behavioral approach based on sweeping assumptions of rationality or "irrationality." To illustrate, one pair of commentators recently argued that the "behavioralist model myopically focuses on the implications of irrationality on certain *specific* market participants, usually incumbent firms or cartel members, while ignoring or assuming away the broader implications of applying an identical cognitive bias to others."[241] The criticism of some behavioral antitrust applications, which the present Article considers at length, is appropriate. But when mistakenly asserting that valid legal analysis must assume that all behavioral phenomena apply identically to all market participants all the time, this commentary offers yet another clear example of the fundamental methodological error.

Ironically, the same two authors level a further charge of naïveté that behavioral scholars allegedly manifest when they "impute a given cognitive bias to *only* a monopolist or to *only* entrants, but not to both, or to other firms at large. There is simply no basis in the behavioral economics literature for this assumption . . . ."[242] By now it should be clear that this charge itself is naïve, not only failing to recognize the extensive empirical

239. *See id.* at 139–44.

240. A full analysis of these questions would also address the significant role of institutions in shaping the behavior of the different classes of actors, as Part III, below, illustrates at length.

241. Wright & Stone, *supra* note 7, at 1535.

242. *Id.* at 1535–36.

evidence for the heterogeneity of judgment and decision behavior at the individual level but also revealing a misunderstanding of the inherently variable nature of behavioral phenomena even at the population level, when whole classes of actors are concerned.

To use these analysts' own illustration, incumbents and entrants may differ in important respects, both between them and within each group of antitrust actors. We already saw, for instance, that new entrants exhibit greater optimistic overconfidence than their diversifying competitors.[243] Importantly, the systematic difference between the two categories of entrants—an example of the variability of overconfidence—does not result from an assumption that one class of entrants fails to manifest a phenomenon that another exhibits. Instead, the factors that are empirically shown to moderate this judgment bias—including the degree of ambiguity and the intensity of preferences—cause a systematic divergence between the two categories of entrants.[244] Similarly, one should expect the class of potential new entrants itself to be heterogeneous, with entrants revealing different degrees of optimistic overconfidence. Since the decision to attempt entry involves self-selection, however, those more biased potential entrants will be overrepresented among the actors who end up attempting entry.[245] Both variability and heterogeneity thus suggest that new entrants will be particularly biased, as a group, generating a prediction that clearly differs from either traditional antitrust models that assume perfect entrant rationality or analyses that mistakenly assume universal and uniform bounded rationality.

Finally, the same two authors who seek to criticize behavioral antitrust advance their main critique by proposing a "behavioral irrelevance theorem" that they "believe provides a more realistic account of firm-level irrationality as it relates to antitrust policy."[246] In another illustration of the fundamental methodological error, the proposed theorem is based on a model that imagines incumbents and entrants, respectively, as either "rational" or "irrational" and proceeds to outline the implications of the four resulting combinations of entrant–incumbent interaction.[247] Thus even while presenting their approach as more sophisticated, these scholars repeatedly and naïvely assume a single, uniform, all-encompassing "irrationality," ignoring the evidence of variability and heterogeneity, with

243. Tor, *Entry*, *supra* note 2, at 520–31.

244. *Id.*

245. *Cf. id.* at 563–64 (examining the factors that lead to an overrepresentation of more biased entrants when considering the variability of entry judgments).

246. Wright & Stone, *supra* note 7, at 1527.

247. *Id.* at 1536–48.

respect to both behavioral phenomena generally and competition among entrants and between them and incumbents more specifically.[248]

III. The Second Mistake: Assuming (Away) Institutional Effects

Real antitrust actors do not operate in an abstract, context-free environment. Both consumers and producers make their judgments and decisions in market settings, where the former seek to satisfy their wants while the latter try to succeed as businesses by the means they believe most effective for accomplishing their goal. Moreover, the producers antitrust law is concerned about typically are large business associations, whose significant actions in the market are determined by often complex interactions among multiple individuals within the organization. The legal decision makers who shape antitrust law and policy—from judges and juries in antitrust courts to enforcement officials in regulatory agencies—similarly operate within, and are affected and constrained by, institutional frameworks.

Yet commentators frequently take one of two extreme approaches, either ignoring the effects of antitrust institutions altogether or assuming that these institutions perfectly align the behavior of antitrust actors with rationality-based models. Behavioral proponents who assume away institutional effects usually tend not to explain why these effects are unimportant or irrelevant to the antitrust questions they examine.[249] Nor do behavioral opponents, who routinely assume that antitrust institutions—particularly markets and firms[250]—guarantee rational behavior, tend to pause to examine the specific conditions that determine whether and how these institutional effects take place.[251]

---

248. *See, e.g.*, Tor, *Entry*, *supra* note 2, at 565 ("A related important lesson . . . is that the legal analyst should strive to develop an accurate understanding of those variables that determine whether and how biased different actors are likely to be."). Notably, the analysis offered by the two authors is flawed even based on its erroneous assumptions, arguing, for example, that rational entrants make irrational incumbent behavior irrelevant. *See* Wright & Stone, *supra* note 7, at 1541. As explained above, "irrational" predation can be successful (and even comprise a rational strategy).

249. *See, e.g.*, Stucke, *Monopolization*, *supra* note 7, at 552–53 (discussing learning by firms and consumers in the same breath, without considering the institutional differences between the two types of market participants).

250. *But see* Cooper & Kovacic, *supra* note 7, at 782 (examining, theoretically, how the behavior of enforcement agencies may be shaped by some behavioral forces).

251. *See, e.g.*, Werden et al., *supra* note 7, at 128 (arguing that evidence of individual decision behavior need not carry to firms and that "[m]oreover, what really matters in competition policy is not so much the behavior of firms as the performance of markets, which need not be significantly impaired by firm decision making subject to behavioral biases"). This common error is puzzling given the incorporation of behavioral insights into mainstream economics in recent decades. Behavioral antitrust opponents today echo earlier arguments made by scholars outside antitrust. *See, e.g.*, Jennifer Arlen, *Comment: The Future of Behavioral Economic Analysis of Law*, 52 VAND. L. REV. 1765, 1770 (1998) (suggesting that the effects of the heuristics and biases offered by behavioral analyses might be weaker than generally assumed and making no distinction

In reality, however, the effects of institutions on antitrust actors are pervasive yet variable. Consumers behave differently in market and nonmarket environments, while producers' incentives and competitive pressures vary depending on the specific market settings, organizational environments, the type of business conduct involved, and more. Judges, juries, and regulators similarly are likely to exhibit varying degrees of rationality depending on the tasks and institutional contexts they face.

The following subparts therefore illustrate the importance of institutional effects in behavioral antitrust analysis by examining how markets and firms shape the behavior of antitrust actors. These illustrations will reveal that the neglect of either the significance of institutional effects or their limits can lead to erroneous antitrust conclusions.

A. Markets

Markets are perhaps the most significant antitrust institution given the primary concern of the field with protecting the competitive process—that is, the competition among producers to supply consumer demand.[252] From a behavioral perspective, markets play an additional, complex role, however, sometimes aligning consumer and producer behavior with the normative standards of rationality while at other times failing to do so or even facilitating deviations from these standards.[253]

1. Demand-Side Rationality.—For consumers, markets supply not only goods and services but also the information that can help them form more rational beliefs and make more rational decisions.[254] When markets offer better and more readily available information, consumers' judgments and decisions may be more accurate and better aligned with their

---

between antitrust actors in making this suggestion). For one rare and insightful, if brief, discussion of the important-yet-limited role of institutions in promoting rationality in legal settings generally, see Jeffrey J. Rachlinski, *The Psychological Foundations of Behavioral Law and Economics*, 2011 U. ILL. L. REV. 1675, 1689–90.

252. *See, e.g.*, Pac. Bell Tel. Co. v. Linkline Commc'ns, Inc., 555 U.S. 438, 447–49 (2009) (recognizing that antitrust law is limited to protecting competition in the marketplace); Leegin Creative Leather Prods., Inc. v. PSKS, Inc., 551 U.S. 877, 885–87 (2007) (discussing both the rule of reason and per se restrictions as mechanisms that allow courts to police the competitive landscape); Bell Atl. Corp. v. Twombly, 550 U.S. 544, 553–54 (2007) (highlighting the role of antitrust law in facilitating competition by policing only certain types of conduct).

253. Tor, *Behavioral Methodology*, *supra* note 1, at 310–13; *cf.* Bar-Gill, *supra* note 139, at 15 (noting, with respect to the interaction between consumers and producers, that "[t]he behavioral economics model . . . is context dependent" such that while his analysis is "often stated in general terms, implementation must be market specific" and "[t]he severity of the behavioral market failure, and the ability of competition to mitigate the welfare costs of the behavioral market failure, will vary from market to market").

254. For one classical economic treatment of the role and importance of information for consumer decisions, see George J. Stigler, *The Economics of Information*, 69 J. POL. ECON. 213 (1961), in which the author discusses the costs of information search and their implications for advertising, the role of specialized intermediaries, and more.

preferences.[255] The available evidence on consumer behavior, however, paints a more complex picture. For one, the products and services that consumers must choose among will not always justify a commitment of significant time or cognitive or financial resources to make optimal judgments and decisions, leading consumers rationally to ignore relevant information.[256]

However, producers who expect to benefit from consumers' educated choices may respond by providing relevant information to consumers via advertising campaigns, marketing, and similar efforts.[257] Such responses not only tap the superior information that producers already possess about their products and services but also offer significant economies of scale, given the low cost of offering the same (or similar) information to additional consumers.[258] Nevertheless, insofar as numerous competing producers offer such information, extolling the superiority of their wares, consumers still must determine which products and services best match their preferences.

In some cases, the opportunity profitably to provide consumers with unbiased information and advice will attract an additional set of market participants—namely, information intermediaries—to fulfill this

255. On the conditions necessary for such improvements, see *infra* notes 299–303 and accompanying text.

256. *Cf.* Melvin Aron Eisenberg, *The Limits of Cognition and the Limits of Contract*, 47 STAN. L. REV. 211 (1995) (discussing rational ignorance regarding contractual terms as a form of bounded rationality); Richard A. Posner, *The Law and Economics of Contract Interpretation*, 83 TEXAS L. REV. 1581, 1585–86 (2005) (noting that the expected benefit of a "good" form contract to the consumer is slight and therefore is unlikely to figure in the decision whether to buy the seller's product).

257. *See, e.g.*, Justin P. Johnson & David P. Myatt, *On the Simple Economics of Advertising, Marketing, and Product Design*, 96 AM. ECON. REV. 756, 756 (2006) (recognizing that advertising and marketing activity can shift the demand curve in a positive direction for producers' firms if consumers respond favorably to a product); Phillip Nelson, *Advertising as Information*, 82 J. POL. ECON. 729, 743–44 (1974) (highlighting the importance of the consumer in producer advertising decisions); Stigler, *supra* note 254, at 220–24 (noting that advertisements serve in part to identify the sellers and prices of goods and also suggesting that the value of information increases as well-informed customers seek additional information more extensively).

258. *See, e.g.*, Stigler, *supra* note 254, at 220 ("A small $5 advertisement in a metropolitan newspaper reaches (in the sense of being read) perhaps 25,000 readers, or fifty readers per penny, and, even if only a tiny fraction are potential buyers (or sellers), the economy they achieve in search . . . may be overwhelming."). Recent scholarship further suggests that producers may possess better information regarding consumers' behavior than do consumers themselves. *See* Oren Bar-Gill & Franco Ferrari, *Informing Consumers About Themselves*, 3 ERASMUS L. REV. 93 (2010) (arguing the significance of product-use information and the need to regulate its disclosure); Oren Bar-Gill & Oliver Board, *Product-Use Information and the Limits of Voluntary Disclosure*, 14 AM. L. & ECON. REV. 235 (2012) (same). *But see* Emir Kamenica et al., *Helping Consumers Know Themselves*, 101 AM. ECON. REV. 417 (2011) (showing that requiring firms to inform consumers about themselves decreases consumer expenditure at given prices but can increase equilibrium prices, offsetting the direct benefit of this information).

function.[259] These specialized service providers, ranging from long-standing outlets aimed at the general public, such as *Consumer Reports*, to more recent internet databases and services, to personalized consultants and advisors, can help improve the quality of consumers' judgments and decisions.

Yet, despite the increasing abundance of information—and occasionally because of it—many consumers still commonly and routinely make product and service choices that are suboptimal for them.[260] Even when competition is present, producers in some markets prefer to offer only partial or opaque information to limit the ability of consumers to evaluate their products.[261] Specifically, producers can benefit by designing products that lead more naive consumers to make inferior, costly decisions—as in the case of some credit card plans—that both increase producers' profits and subsidize the superior products chosen by more sophisticated consumers, helping attract the latter as well.[262] In other instances, sellers develop products that are more complex than necessary to satisfy consumer

259. *See* FRANK ROSE, THE ECONOMICS, CONCEPT, AND DESIGN OF INFORMATION INTERMEDIARIES: A THEORETIC APPROACH 36–37 (1999) (explaining that buyers of goods demand information to make more informed purchases and this demand has given rise to a market for information to be purchased); DANIEL F. SPULBER, MARKET MICROSTRUCTURE: INTERMEDIARIES AND THE THEORY OF THE FIRM xxiii (1999) (observing that customers can be asymmetrically informed about product quality and that intermediaries help to fill this gap); Thomas F. Cosimano, *Intermediation*, 63 ECONOMICA 131 (1996) (offering a model showing the conditions for beneficial intermediation and the costs its presence imposes on sellers who do not use it); Daniel F. Spulber, *Market Microstructure and Intermediation*, J. ECON. PERSP., Summer 1996, at 135 (elucidating the role of the intermediary in the market system); *see also* Stigler, *supra* note 254, at 216–17 (discussing the role of distributors as information intermediaries).

260. *See* Oren Bar-Gill & Elizabeth Warren, *Making Credit Safer*, 157 U. PA. L. REV. 1, 26–55 (2008) (arguing that consumers make systematic mistakes in their choice and use of credit products due to imperfect information and irrationality); Oren Bar-Gill, *Seduction by Plastic*, 98 NW. U. L. REV. 1373, 1395–1408 (2004) (contending that "imperfect self-control" can explain consumers' systematic underestimation of their future borrowing); Brian Bucks & Karen Pence, *Do Homeowners Know Their House Values and Mortgage Terms?* 1, 26–27 (Fed. Reserve Bd. Fin. & Econ. Discussion Series, Working Paper No. 2006-03, 2006), *available at* http://www.federalreserve.gov/Pubs/feds/2006/200603/200603pap.pdf (reporting that borrowers with a limited understanding of mortgage terms are at risk of taking out more costly mortgages than they qualify for).

261. *See, e.g.*, OREN BAR-GILL, SEDUCTION BY CONTRACT: LAW, ECONOMICS, AND PSYCHOLOGY IN CONSUMER MARKETS 80 (2012) ("Increased complexity may be attractive to issuers, as it allows them to hide the true cost of the credit card in a multidimensional pricing maze.").

262. Xavier Gabaix & David Laibson, *Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets*, 121 Q.J. ECON. 505, 519–20 (2006). *See generally* BAR-GILL, *supra* note 261, at 8 (arguing that sellers must be aware of consumer biases and perceptions in order to remain in the market).

demand—such as where certain cellular service plans are concerned[263]—making it exceedingly difficult to compare among competing offerings.[264]

All in all, while current-day markets typically provide consumers with abundant information that can facilitate better judgments and decisions, consumers still face significant challenges. Where the interests of producers and consumers diverge substantially, the latter frequently are at a fundamental disadvantage compared to the former—who have the experience, opportunity, and resources needed to exploit consumers. Nonetheless, the basic observation of consumer disadvantage that permeates other areas of the law—most notably consumer protection and some regulatory regimes—until recently has largely been absent from antitrust discourse.[265] The hypothetical consumer in traditional antitrust models is not just strictly rational, but usually also immune to the institutional constraints that impact real consumers in market settings.[266]

Yet should the evidence of consumers' bounded rationality enter the antitrust calculus? After all, Nobel Prize-winning economist Milton Friedman argued that markets in the aggregate perform "as if" they were comprised of rational participants because the deviations of irrational actors

---

263. Bar-Gill, *supra* note 139, at 35–41.

264. *See, e.g.*, Bruce I. Carlin, *Strategic Price Complexity in Retail Financial Markets*, 91 J. FIN. ECON. 278 (2009) (presenting a model showing that as competition intensifies firms add complexity to their price structures to increase market power by preventing some consumers from becoming knowledgeable about prices); Chris M. Wilson, *Ordered Search and Equilibrium Obfuscation*, 28 INT'L J. INDUS. ORG. 496 (2010) (demonstrating "the incentives for an oligopolist to obfuscate by deliberately increasing the cost with which consumers can locate its product and price"); Glenn Ellison & Alexander Wolitzky, *A Search Cost Model of Obfuscation* (Nat'l Bureau of Econ. Research, Working Paper No. 15237, 2009), *available at* http://www.nber.org/papers/w15237 (offering search models that show obfuscation can arise even with rational consumers who bear search costs). *But see* Alexia Gaudeul & Robert Sugden, *Spurious Complexity and Common Standards in Markets for Consumer Goods*, 79 ECONOMICA 209 (2012) (modeling a countervailing force of consumer preference for simple choice, which can reduce complexity); Eugenio J. Miravete, *Competition and the Use of Foggy Pricing*, 5 AM. ECON. J. MICROECONOMICS 194 (2013) (concluding that the transition from monopoly to competition in the early U.S. cellphone industry did not generally foster the use of tariff options aimed at profiting from consumer mistakes and offering alternative accounts for the observed "foggy" pricing).

265. Recently, economists considering some aspects of behavioral antitrust have begun considering the possibility and implications of consumer manipulation for antitrust. *See* Bennett et al., *supra* note 7, at 121; Huffman, *Neo-Chicago with Behavioral Antitrust*, *supra* note 7, at 131–35; Salinger, *supra* note 7, at 81–82; Maurice E. Stucke, *Behavioral Exploitation and its Implications on Competition and Consumer Protection Policies*, *in* THE PROS AND CONS OF CONSUMER PROTECTION, *supra* note 139, at 77; Petit & Neyrinck, *supra* note 7, at 9–11.

266. *See* Stucke, *Reconsidering*, *supra* note 7, at 122–23. One exception is the practice of merger enforcement, where the antitrust agencies routinely consider case-specific evidence, including evidence of consumer behavior with respect to the relevant products, when such data is available. *See infra* notes 270–74 and accompanying text (discussing the approach of the agencies to consumer behavior in merger investigations).

collectively cancel each other out.[267] This argument, however, fails to account for *systematic* deviations from rationality that bias market behavior in predictable and consistent directions and therefore do not cancel out in the aggregate.[268]

Gary Becker, another Nobel Prize-winning economist, made a different argument, showing that one can derive the main implication of traditional economic models of consumer behavior—namely, the negatively sloping demand curve that associates higher prices with lower demand—without assuming rational behavior.[269] Becker's argument suggests that consumers' systematic deviations from strict rationality should still generate recognizable markets, with negatively sloping demand curves, as we routinely observe in fact. But this insight is not particularly helpful for antitrust law and enforcement policy, which rely on assumptions of consumer rationality well beyond setting up negatively sloping demand curves.

We have already seen, in fact, that consumer rationality impacts antitrust doctrine in a number of areas, from the debate over aftermarket power in *Kodak*, to the analysis of bundling and tying, RPM, and even the efficacy of new entry. Systematic bias on the part of consumers may be troublesome for other key aspects of merger enforcement as well. The agencies and merging parties routinely estimate the unilateral effects of mergers based on models in which firms price to maximize profits in the face of aggregate consumer demand.[270] Hence, merger predictions that fail to account for systematic biases in consumer demand—whereby consumers, for instance, over- or underreact to changes in the relative prices of products in a given market—may result in erroneous predictions of merger outcomes.[271]

Some economists argue that there is little reason for alarm because merger assessments already account for any systematic consumer bias by drawing on data regarding consumers' actual choices in the relevant product

---

267. MILTON FRIEDMAN, *The Methodology of Positive Economics*, *in* ESSAYS IN POSITIVE ECONOMICS 3, 21–22 (1953) (emphasis omitted); *see also* Allan Gibbard & Hal R. Varian, *Economic Models*, 75 J. PHIL. 664, 669–73 (1978) (discussing the concepts of approximation and fit in microeconomic models that are based on false assumptions); Richard A. Posner, *Rational Choice, Behavioral Economics, and the Law*, 50 STAN. L. REV. 1551, 1556 (1998) (echoing Friedman's argument in his critique of behavioral law and economics).

268. Tor, *Entry*, *supra* note 2, at 563.

269. Gary S. Becker, *Irrational Behavior and Economic Theory*, 70 J. POL. ECON. 1, 4–9 (1962) (showing that this basic feature of markets results not from the assumed rationality of market participants, but more directly from the effect of a change in price on opportunities).

270. *See* 2010 MERGER GUIDELINES, *supra* note 30, §§ 6.1–.3; DAVIS & GARCÉS, *supra* note 80; Werden & Froeb, *supra* note 80.

271. *Cf.* Bennett et al., *supra* note 7, at 119 (noting that "[p]assive" consumers can reduce both the overall price elasticity of a product or its cross-elasticity with other products).

market.[272] This argument, however, does not apply to most merger investigations, only to those limited situations where extensive, quantitative scanner or similar data is readily available, such as in consumer goods markets.[273] It also glosses over the role of consumer rationality even in some of those merger simulation methods that are commonly used when sufficiently detailed consumer-level data is available.[274]

Yet even when aggregate data from real consumer transactions enables reasonable predictions of merger effects, it does not resolve the more fundamental challenge of systematic consumer bias in merger assessments. When consumer choices are partly driven by systematic errors of judgment, choices in the market may fail to reflect consumers' true preferences.[275] To illustrate, when consumers underestimate the risks associated with a given product, they demand greater quantities of that product than they would have absent their judgment error. A somewhat different problem occurs when consumers systematically deviate from rational choice precepts.[276] Their behavior then may manifest real preferences, yet these may not always be those "true" preferences these consumers would have exhibited had they not been biased.[277] Hence it is clear, even without resolving these significant questions, that the empirical behavioral evidence is potentially relevant to merger enforcement.[278]

Indeed the law should not ignore the complex relationship between markets, competition, and consumers' deviations from strict rationality,

---

272. *See, e.g.*, Wright & Stone, *supra* note 7, at 1523.

273. *See, e.g.*, DAVIS & GARCÉS, *supra* note 80, § 9.3.2 (explaining the limited availability of such data).

274. *See* sources cited *supra* note 81.

275. Stucke, *Reconsidering*, *supra* note 7, at 139–40; *see also* Werden et al., *supra* note 7, at 127 (implying that behavioral evidence could be used to challenge the assumption that consumers maximize utility via choice).

276. For the distinction between errors of judgment and errors of choice and their potentially distinct normative ramifications, see Tor, *Behavioral Methodology*, *supra* note 1, at 244–72.

277. *See id.* at 318 & n.344 (explaining that "true" preferences refer to those that individuals would have had if they had possessed an accurate estimate of the risks and benefits of products).

278. This observation raises a separate question of some importance concerning the goals of antitrust law. Modern antitrust law promotes efficiency by protecting competition based on the assumption that consumers maximize utility through their choices in the market. Therefore, evidence showing that consumer choice sometimes fails to maximize utility might raise questions regarding the validity of the competition–welfare link at the foundation of antitrust law, as scholars supporting and criticizing behavioral antitrust already have noted. *See* Alexander Morrell, *Behavioral Antitrust and Merger Control: Comment*, 167 J. INST. THEORETICAL ECON. 143, 146–47 (2011) (criticizing the argument made by Werden et al. that the behavioral evidence should be ignored, regardless of its validity, if it makes current practices in welfare economics obsolete). Note, however, that the behavioral evidence does not negate the importance of competition for consumer welfare. For one, the relevant inquiry is not whether consumer choice always maximizes welfare but whether competition tends significantly to improve consumer welfare compared to what prevails when competition diminishes. Yet a full consideration of this issue is beyond the scope of the present analysis.

given the many roles of hypothetical consumer rationality in both antitrust doctrine and enforcement practices.

2. Supply-Side Rationality.—Beyond providing producers with incentives and opportunities to react to the bounded rationality of consumers—whether by seeking to correct it or trying to exploit it—markets also help align producers' own behavior with rationality through a number of mechanisms.[279] Some of these—such as the consequences of simple aggregation or the inevitably constrained resources of market participants—operate at the macro-level, facilitating "as if" rational outcomes for the market as a whole, irrespective of the actual rationality of specific market participants. Other, micro-level mechanisms—including profit motivation and learning from experience—directly facilitate more rational behavior on the part of individual producers. A final pair of mechanisms—product-market competition and arbitrage—operates at the market level yet impacts micro-level behavior by weeding out boundedly rational producers.[280] Importantly, though powerful and significant, the various mechanisms of market rationality are imperfect, at times failing to ensure producer rationality and occasionally even facilitating systematic deviations from it.[281]

Economists have long argued that markets overall may comport with the predictions of strictly rational models even while individual firms deviate from them. Milton Friedman explained that the aggregation of firm behavior in the market means that random errors will cancel out in the aggregate.[282] Nonetheless, as we saw already with respect to consumer behavior, systematic deviations from rationality may not cancel out, instead

279. The notion that nonmarket behavioral phenomena may not appear in market settings is common not just to opponents, but even to some proponents of the behavioral approach outside antitrust law. *See, e.g.*, Arlen, *supra* note 251, at 1781–82 (suggesting that behavioral findings from nonmarket settings may not necessarily generalize to market settings); Jolls et al., *supra* note 1, at 1473 (finding it necessary to state that "law is a domain where behavioral analysis would appear to be particularly promising in light of the fact that nonmarket behavior is frequently involved"); Thomas S. Ulen, *The Growing Pains of Behavioral Law and Economics*, 51 VAND. L. REV. 1747, 1748–49, 1758–60 (1998) (arguing that bounded rationality may be of limited importance for the analysis of market behaviors because of competitive discipline).

280. To distinguish among the different mechanisms and their effects on rationality, the present subpart treats the firm as a single decision maker, while the next subpart considers in detail the various intra-organizational mechanisms that operate within the producer firm.

281. For the distinction between market-level rationality—namely, the compatibility of aggregate market outcomes with models based on assumptions of strict rationality—and micro-level, or individual, rationality that pertains to specific firms, see Becker, *supra* note 269, who introduces this distinction and uses it to explain seemingly contradictory findings at the two levels.

282. *See* FRIEDMAN, *supra* note 267, at 21–23.

generating broader market patterns that differ from predictions based on hypothetical rationality.[283]

Similarly, Becker extended his argument regarding irrational consumer behavior, *ceteris paribus*, to producers.[284] He showed that even firms that do not maximize profits must respond systematically and predictably to changes in their production opportunity set: as the price of inputs or the competitive conditions in the market change, even firms acting randomly or those guided by inertia respond accordingly.[285] For instance, Becker showed that the basic economic finding that a competitive market that becomes monopolized (or cartelized) will tend to lower output holds when firms are irrational.[286]

However, the generality of this result—which applies not just to business firms but to all decision makers with resource constraints[287]—also spells its limited significance for antitrust purposes. Becker indeed showed that rationality on the part of the individual decision-making unit is not required for aggregate market responses to move in the same *direction* as predicted by traditional models. Yet antitrust law treats differently market behaviors with the same general propensity—such as an increase in price or a reduction in output—depending on the *magnitude* of change. Mergers among competitors are legal unless they are likely substantially to lessen competition,[288] monopolization and attempted monopolization both apply only to firms above a certain market power threshold,[289] and exclusive dealing, tying arrangements, and some other restraints of trade similarly are prohibited for some firms yet permitted for others depending, *inter alia*, on their degree of market power.[290] In each of these areas of antitrust doctrine,

---

283. *See* Tor, *Entry*, *supra* note 2, at 563 (discussing this caveat with respect to Friedman's argument).

284. Becker, *supra* note 269, at 9–12 (showing how the argument extends to firms that use decision rules other than profit maximization).

285. *Id.*

286. *Id.* at 11 (explaining that "a change from competition to monopoly shifts the production opportunity set toward lower outputs, which in turn encourages irrational firms to lower their outputs").

287. *See id.* at 12.

288. *See* Clayton Act § 7, 15 U.S.C. § 18 (2012); 2010 MERGER GUIDELINES, *supra* note 30, *passim*.

289. *See* United States v. Grinnell Corp., 384 U.S. 563, 570–71 (1966) (recognizing that a plaintiff must demonstrate that the defendant possessed a certain level of "monopoly power" in order to prevail in a monopoly claim); Ball Mem'l Hosp., Inc. v. Mut. Hosp. Ins., Inc., 784 F.2d 1325, 1334–35 (7th Cir. 1986) ("Unless the defendants possess market power, it is unnecessary to ask whether their conduct may be beneficial to consumers. Firms without power bear no burden of justification.").

290. *See* Aspen Skiing Co. v. Aspen Highlands Skiing Corp., 472 U.S. 585, 601–02 (1985) (noting that the right to select customers and associates is qualified where a firm seeks to create or maintain a monopoly); *see also* Avishalom Tor, *Unilateral, Anticompetitive Acquisitions of Dominance or Market Power*, 76 ANTITRUST L.J. 847 (2010) (expanding upon the special responsibilities of firms with monopoly power).

different markets that move in the same general direction generate very different legal results depending on the respective magnitude of change in market power. The nature of these changes, however, depends in part on the nature of producers' rationality in a given market setting.[291]

Even when it evaluates market-wide outcomes, moreover, antitrust law ultimately is concerned with the conduct of specific firms. Yet the conduct of a given producer and its competitive effects also may depend on the degree to which the that firm, its competitors, suppliers, and so on adhere to the precepts of rational profit maximization. To illustrate, the same allegedly predatory conduct that would not occur in a world populated only by perfectly rational firms—say, because market conditions make recoupment of the costs invested in predation unlikely—can both take place and generate significant competitive harm where a real monopolist is (or is perceived to be) irrationally aggressive in the face of new entry.[292]

All in all, the rationality of market participants—as distinct from aggregate market outcomes—can be material for antitrust analysis, meriting a careful evaluation of the micro-level mechanisms of supply-side market rationality. Perhaps the most fundamental micro-level difference between market and nonmarket behavior is that the former primarily aims at earning profits. Whatever other motivations may contribute to their activities, profits are the *raison d'être* of business firms. We thus expect profit-seeking suppliers to be more rational, avoiding errors that decision makers exhibit in nonmarket settings.[293]

Traditionally in economics, the notion that monetary incentives matter and that larger monetary incentives lead to greater effort and better performance is near axiomatic.[294] Yet the empirical evidence suggests this is not always the case: though financial incentives can increase effort, this greater effort generates only limited improvements in the rationality of people's intuitive judgment and decision making.[295] In fact, sometimes

---

291. *Cf.* Thomas Russell & Richard Thaler, *The Relevance of Quasi Rationality in Competitive Markets*, 75 AM. ECON. REV. 1071, 1071 (1985) (showing, in an early model, that in the presence of systematic deviations from rationality the standard "rational" outcome at the market level holds only under very specific conditions but not as a general case).

292. Leslie, *supra* note 3, at 298–300; Tor, *Predatory Pricing*, *supra* note 2, at 55–57.

293. Or that consumers exhibit, even in market settings, given their very different incentives.

294. *See, e.g.*, Uri Gneezy et al., *When and Why Incentives (Don't) Work to Modify Behavior*, J. ECON. PERSP., Fall 2011, at 191 (discussing the conditions under which extrinsic—particularly monetary—incentives work and do not work).

295. Colin F. Camerer & Robin M. Hogarth, *The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework*, 19 J. RISK & UNCERTAINTY 7, 7 (1999); Ondrej Rydval & Andreas Ortmann, *How Financial Incentives and Cognitive Abilities Affect Task Performance in Laboratory Settings: An Illustration*, 85 ECON. LETTERS 315 (2004) (showing that cognitive abilities are at least twice as important as financial incentives for performance). This is not to say that financial incentives do not matter. *See, e.g.*, Vernon L. Smith, *Method in Experiment: Rhetoric and Reality*, 5 EXPERIMENTAL ECON. 91, 101–02 (2002) (highlighting the efficacy of financial incentives); Vernon L. Smith & James M. Walker,

increased monetary incentives even diminish the rationality of performance.[296] Furthermore, producers' increased competitive efforts at times may be directed at goals such as increased market share or relative position in the market rather than pure profit maximization.[297]

For profit motivation to improve the performance of boundedly rational producers they must learn to correct their deviations from rationality. Effective learning requires market participants to identify their judgment and decision errors, to associate these errors with specific negative consequences, and, finally, to replace their deviations with more rational judgments and decisions.[298] However, in the typical antitrust settings, such learning can be exceedingly difficult. Most judgments and decisions in product markets are made under uncertainty: outcomes are multiply determined and delayed; feedback is limited and noisy; and there is no reliable information about the counterfactual outcomes that would have occurred had a different choice been made.[299]

Over time and with experience producers nevertheless can improve their performance even without "true" learning. They may imitate

---

Monetary Rewards and Decision Cost in Experimental Economics, 31 ECON. INQUIRY 245, 259–60 (1993) (concluding, based on a review of experimental economics studies, that increased financial incentives reduce the variances in participants' performance and also tend to improve its quality).

296. *See, e.g.*, Dan Ariely et al., *Large Stakes and Big Mistakes*, 76 REV. ECON. STUD. 451, 451–52 (2009) (comparing performance on identical tasks with varying monetary incentives); Uri Gneezy & Aldo Rustichini, *Pay Enough or Don't Pay at All*, 115 Q.J. ECON. 791, 791 (2000) (providing evidence that low financial incentives lead to worse performance than no incentives by crowding out alternative motivations but that high incentives improve performance); Dan N. Stone & David A. Ziebart, *A Model of Financial Incentive Effects in Decision Making*, 61 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 250, 258 (1995) (providing evidence that financial incentives can improve performance via increased motivation and diminish it by generating negative affect).

297. *See* Armstrong & Huck, *supra* note 7, at 13–17 (discussing how firms will sometimes choose to imitate similar firms rather than attempt to calculate their own optimal strategies); Stephen N. Garcia, Avishalom Tor & Richard Gonzalez, *Ranks and Rivals: A Theory of Competition*, 32 PERSONALITY SOC. PSYCHOL. BULL. 970 (2006); Gneezy et al., *supra* note 294; Goppelsroeder, *supra* note 111.

298. *See, e.g.*, Hillel J. Einhorn, *Learning from Experience and Suboptimal Rules in Decision Making*, *in* COGNITIVE PROCESSES IN CHOICE AND DECISION BEHAVIOR 1 (Thomas S. Wallsten ed., 1980) (emphasizing the importance of unambiguous feedback for learning); Richard E. Nisbett et al., *Improving Inductive Inference*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 445, 445–46 (Daniel Kahneman et al. eds., 1982) (noting that decision makers need to know that an error has occurred, how it has occurred, and how to improve the decision process); Tversky & Kahneman, *Rational Choice*, *supra* note 90, at S274–75 (explaining that effective learning only takes place when "accurate and immediate feedback about the relation between the situational conditions and the appropriate response" is available).

299. On the importance of effective feedback, see Nigel Harvey & Ilan Fischer, *Development of Experience-Based Judgment and Decision Making: The Role of Outcome Feedback*, *in* THE ROUTINES OF DECISION MAKING 119, 119 (Tilmann Betsch & Susanne Haberstroh eds., 2005) (noting that "when feedback is present . . . significant improvements are likely to occur. . . . [But] when people perform the task repeatedly without finding out anything about the results of their efforts, improvements are usually negligible").

successful competitors, follow established industry norms,[300] or seek the advice of service providers with expertise in improving business outcomes.[301] Sometimes such efforts will align the producers' conduct with strict rationality, but at other times they will fail to do so. Imitation may be directed at the wrong elements of competitors' conduct, industry norms may be neither rational nor efficient, and to seek and invest in outside advice—not to mention follow it successfully—one must first recognize the suboptimal behavior.

If the challenges involved in learning from experience in product markets were not enough, many particularly significant antitrust-relevant judgments and decisions are infrequent, sometimes unique. Entry into new markets, mergers and acquisitions, the development of new business strategies and vertical arrangements, and so on all offer producers exceedingly limited learning opportunities.

At the micro level, then, though limited in their efficacy, both profit seeking and learning can improve the rationality of individual producers. But should the competitive process itself not suffice to align producer behavior with rationality-based models simply by weeding out those less capable competitors who fail to learn? Alchian's familiar argument states that underperforming producers in competitive markets will be less profitable than their competitors and ultimately will not survive.[302] Based on this logic, commentators frequently assume that competition will weed out boundedly rational decision makers who must deplete their resources by making inefficient decisions while their rational competitors enjoy higher profits.[303]

A more careful consideration shows, however, that rationality-inducing competition is limited in antitrust-relevant environments for two sets of reasons, one relating to the nature of behavioral deviations from rationality in markets, the other having to do with the subject matter of

---

300. *See* Armstrong & Huck, *supra* note 7, at 13–18 (discussing findings suggesting that "imitation enables firms to make use of other firms' private information and optimizing behavior, and to enjoy the benefits of conformity (as imitating firms are likely to do as well as the average of their peers)"); *see also* Goppelsroeder, *supra* note 111, at 26 (noting that firms entering mature, cartelized industries might "simply join[] the cartel, [emulating successful competitors by] internalizing the collusive agreement so that tacit collusion can be sustained").

301. *See* Bailey, *supra* note 81, at 6 (discussing arguments for continuing to treat firms as rational, profit-maximizing entities, one of which is that "firms may have access to a wide array of consultants and advisors who can assist in information processing and making optimal pricing decisions").

302. Armen A. Alchian, *Uncertainty, Evolution, and Economic Theory*, 58 J. POL. ECON. 211, 213 (1950); *see also* Becker, *supra* note 269, at 9–12 (noting that Alchian's argument is a specific iteration of the broader argument that markets behave rationally irrespective of the rationality of participants when resources are constrained).

303. *Cf.* Ernst Fehr & Jean-Robert Tyran, *Individual Irrationality and Aggregate Outcomes*, J. ECON. PERSP., Fall 2005, at 43, 44 (describing the common argument "that rational agents will drive the irrational agents from the market because the former make higher profits").

antitrust law: first, while competition may weed out those who consistently underperform, deviations from rationality are variable and heterogeneous.[304] When decision makers exhibit different biases to different degrees at different times, however, even those who ultimately outperform their competitors may still differ significantly from the hypothetical rational actor.[305]

Furthermore, even effective competitive discipline penalizes only behaviors that reduce profitability. Deviations from strict rationality that benefit market participants, on the other hand, are facilitated rather than hindered by competition. For example, competitive selection rewards with higher returns some biased decision makers who take risks that their rational competitors avoid, so these particular competitors will outperform their rational peers even while the majority of their boundedly rational counterparts fail.[306] Similarly, producers who trust their peers, adhere to social norms, or exhibit other nonstandard social preferences may obtain higher profits through oligopolistic coordination or cartelization in market conditions that would prevent strictly rational competitors from doing so.[307]

Second, antitrust law largely focuses on those least competitive markets, which inevitably exert more limited disciplinary pressure on market participants. For instance, a monopolist or oligopolists in markets with significant entry barriers may dissipate some of their supracompetitive profits by operating less efficiently.[308] Hence, systematic deviations from rationality, even when unprofitable, may survive in noncompetitive markets much like other inefficiencies.

Another disciplinary force besides product market competition is arbitrage by rational actors who identify, exploit, and, thus, erode the profit opportunities generated by the errors of boundedly rational decision makers.[309] For this to happen, however, there must exist a sufficiently large

---

304. *See supra* Part II.

305. *Cf.* Fehr & Tyran, *supra* note 303, at 54 (showing how individual irrationality—even assuming its constancy—may translate to different aggregate market performance depending on whether deviations from rationality are strategic substitutes or strategic complements).

306. *See* Tor, *Entry*, *supra* note 2, at 504–11 (describing this type of competitive selection process in the competition between more and less biased entrants into manufacturing industries); *cf.* J. Bradford De Long et al., *The Survival of Noise Traders in Financial Markets*, 64 J. BUS. 1 (1991) (showing how overoptimistic traders in financial markets—as a group—may in fact earn higher returns on average and thus exhibit long-run survival). For related intra-firm processes that sometimes select for managerial deviations from strict rationality, see *infra* subpart III(B).

307. *See, e.g.*, Armstrong & Huck, *supra* note 7, at 21–24; Leslie, *supra* note 3, at 280–83.

308. *See generally* HARVEY LEIBENSTEIN, BEYOND ECONOMIC MAN: A NEW FOUNDATION FOR MICROECONOMICS (1976) (discussing the theory of and evidence for "x-inefficiency"—that is, a nonallocative efficiency loss—where firms enjoy some degree of sheltering from competitive pressures).

309. *See* Andrei Shleifer & Robert W. Vishny, *The Limits of Arbitrage*, 52 J. FIN. 35, 35 (1997) (noting that a function of arbitrage "is to bring prices to fundamental values and to keep markets efficient"). *See generally* ANDREI SHLEIFER, INEFFICIENT MARKETS: AN INTRODUCTION

group of arbitrageurs who can both identify the opportunity and bear the risk and costs involved with selling to or buying from the boundedly rational actors; it also requires the ready availability of substitutes for the products overpriced or underpriced by boundedly rational actors.[310]

Yet these conditions are uncommon even in sophisticated financial markets,[311] not to mention real product markets. For example, rational arbitrageurs would find it difficult to engage in product market activities that profitably exploit a given manufacturer's excessive use of RPM that causes the ultimate overpricing of some products at retail. In fact, significant arbitrage is impractical in most product markets even where products are underpriced rather than overpriced so that no short selling is required. To illustrate, a monopolist engaging in below-cost predatory pricing would have offered arbitrageurs a profit opportunity, had they been able to purchase very large quantities of the heavily discounted product and then resell the product at higher prices on a later date. In reality, however, arbitrage is impractical here given the risks involved, the costs of buying sufficient quantities, stocking and reselling, and so on.

All in all, the myriad mechanisms of market rationality clearly constrain deviations from strict rationality, partly confirming the common intuition that producers are more likely to behave rationally than consumers. At the same time, however, the rationality-inducing effects of aggregation and resource constraints, of profit seeking and learning, and of competitive discipline and arbitrage are more limited than many analysts recognize, particularly in those market settings that antitrust law and policy are most concerned with.

B. Managers and Firms

While markets—particularly when they are competitive—can promote rationality, producers' judgment and decision behavior is also shaped by intra-firm institutions. Because producers are business organizations rather than mere individuals, they can: recruit experienced, highly capable agents to manage them; draw on organizational routines to guide managers' behavior; use contractual arrangements to align these agents' motivation

---

TO BEHAVIORAL FINANCE (2000) [hereinafter SHLEIFER, INEFFICIENT MARKETS] (providing a readable and comprehensive review of the early behavioral finance literature).

310. *See* SHLEIFER, INEFFICIENT MARKETS, *supra* note 309, at 4 (noting that arbitrage is most effective when "substitute securities are readily available"); Barberis & Thaler, *supra* note 84, at 5–7 (noting the importance of a close substitute for the mispriced security in minimizing risk for the arbitrageur); Denis Gromb & Dimitri Vayanos, *Limits of Arbitrage*, 2 ANN. REV. FIN. ECON. 251 (2010) (surveying the theoretical literature and offering a simple model that incorporates costs and constraints of arbitrage including risk; short-selling costs; leverage and margin constraints; and constraints on equity capital).

311. This observation is illustrated by the famous collapse of a multi-billion-dollar hedge fund whose trading strategy was based on arbitrage. *See* ROGER LOWENSTEIN, WHEN GENIUS FAILED: THE RISE AND FALL OF LONG-TERM CAPITAL MANAGEMENT (2000).

with the interests of the firm; and make group decisions by corporate boards of directors that can direct, monitor, and discipline managers.[312]

As in the case of markets, however, the empirical evidence on managerial and firm behavior—both generally and with respect to antitrust-relevant tasks in particular—reveals a complex picture. Managers are sophisticated and experienced professional actors, but still human. Notably, managers are selected and shaped by institutional forces to manifest greater rationality in some respects but systematic bias in other respects, as amply illustrated by the empirical and theoretical literature in behavioral corporate finance.[313] In the same vein, corporate governance research demonstrates the limited ability of key intra-firm mechanisms—from contractual arrangements to boards—to guarantee desirable behavior by corporate decision makers.[314]

1. Managers.—Business managers may be more rational in their judgment and decision behavior than other individuals because of their experience and expertise. Research shows that experts in some fields outperform individuals who do not have domain-specific expertise.[315] However, the evidence also reveals that where the rationality of judgment

312. *Cf.* Donald C. Langevoort, *Behavioral Approaches to Corporate Law*, *in* RESEARCH HANDBOOK ON THE ECONOMICS OF CORPORATE LAW 442 (Claire A. Hill & Brett H. McDonnell eds., 2012) (discussing some basic challenges of the behavioral analysis of legal questions pertaining to firms and managers in the context of corporate governance and the securities laws).

313. *See infra* section III(B)(1).

314. This literature is vast; some recent findings are reviewed. *See, e.g.*, Lucian A. Bebchuk & Michael S. Weisbach, *The State of Corporate Governance Research*, 23 REV. FIN. STUD. 939 (2010). *See generally* KRAAKMAN ET AL., *supra* note 116 (providing basic insights into corporate governance structure); Luigi Zingales, *Corporate Governance*, *in* 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 497 (Peter Newman ed., 1998) (offering a clear analysis of the key issues concerning the various constraints and institutions employed within firms to incentivize and monitor managerial behavior).

315. Much of this research developed under "Naturalistic Decision Making" (or NDM)—an approach that focuses on how people make decisions in real-world settings—and while significant, its scope is limited compared to the broader judgment and decision-making literature. NDM defines experts based on the subjective perceptions of people in the field. *See* James Shanteau, *Competence in Experts: The Role of Task Characteristics*, 53 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 252, 255 (1992) (defining experts as "those who have been recognized within their profession as having the necessary skills and abilities to perform at the highest level"). *See generally* Gary Klein, *Naturalistic Decision Making*, 50 HUM. FACTORS 456, 457 (2008) (noting that NDM "shifted our conception of human decision making from a domain-independent general approach to a knowledge-based approach exemplified by decision makers who had *substantial experience*" (emphasis added)); Rebecca Pliske & Gary Klein, *The Naturalistic Decision-Making Perspective*, *in* EMERGING PERSPECTIVES ON JUDGMENT AND DECISION RESEARCH 559, 559 (Sandra L. Schneider & James Shanteau eds., 2003) (citing a definition of NDM as the study of "how people use their experience to make decisions in field settings"). *But see id.* at 577–80 (discussing common criticism of the weakness of the methods used by NDM researchers and questioning the inferences they make about the efficacy of expert decision making in the fields they study).

and decision behavior specifically is concerned, experts often make mistakes that resemble those of other individuals.[316]

The main factors that determine experts' performance—besides the extent of their experience and subject matter expertise, of course—are the nature of the task and the decision environment.[317] Apparently, the learning processes that help experts develop more rational behavior resemble those that individuals use for learning more generally. In domains where feedback is clear and readily available—such as in the area of weather forecasting—experts can perform well even in the face of uncertainty, and they continuously improve their performance.[318] Yet in many other domains—particularly where feedback is limited and ambiguous—experts often fail to exhibit more rational behavior. As Kahneman and Klein recently noted, there are two preconditions for "the process of skill acquisition that supports the intuitive judgments and preferences of genuine experts," namely, "high-validity environments and an adequate opportunity to learn [in] them."[319]

Skilled intuition can only develop in environments that offer valid—causal and statistical—cues to the nature of the situation with sufficient regularity.[320] Importantly, validity should not be confused with certainty; some uncertain environments provide decision makers with significant statistical cues that can assist in acquiring domain-specific expertise. In games of chance like bridge or poker, for instance, experts can identify superior bets that improve their overall performance without guaranteeing the success of every given choice.[321]

---

316. Colin F. Camerer & Eric J. Johnson, *The Process–Performance Paradox in Expert Judgment: How Can Experts Know So Much and Predict So Badly?*, *in* RESEARCH ON JUDGMENT AND DECISION MAKING, *supra* note 88, at 342, 342–43 (noting that behavioral studies "suggest that a wide range of experts . . . are not much better predictors than [those with lesser expertise]" and seeking to reconcile this with the view of cognitive scientists, who argue that "expertise is a rare skill that develops only after much instruction, practice, and experience"); *see also* Daniel Kahneman & Gary Klein, *Conditions for Intuitive Expertise: A Failure to Disagree*, 64 AM. PSYCHOLOGIST 515, 515 (2009) (mentioning "the commonplace that expert intuition is sometimes remarkably accurate and sometimes off the mark").

317. *See* Kahneman & Klein, *supra* note 316, at 520–23.

318. Allan H. Murphy & Robert L. Winkler, *Probability Forecasting in Meteorology*, 79 J. AM. STATISTICAL ASS'N 489, 493 fig.1, 494 fig.2 (1984) (reviewing evidence for good calibration of weather forecasters probability judgments and their improvements over time in real-world and experimental settings alike).

319. Kahneman & Klein, *supra* note 316, at 519; *see also* ROBIN M. HOGARTH, EDUCATING INTUITION 90 (2001) (arguing that because "[i]ntuitions are acquired through experience" and "the validity of a person's intuition depends on the kind of learning structure in which that intuition was acquired," the concept of learning structure not only provides a framework for assessing the validity of intuition but also suggests the kinds of environments that would develop good intuitions).

320. Kahneman & Klein, *supra* note 316, at 520.

321. *Id.*

In the same vein, individuals may fail to develop reliable skilled intuitions even in environments that in principle offer high-validity cues. For one, where the decision task involved is uncommon, there may not be sufficient opportunities to learn its rules. In other cases, people may hold subjectively compelling intuitions that lead them to overemphasize some environmental cues or ignore others despite ample opportunities to acquire skill in judgment.[322]

Unsurprisingly, therefore, numerous studies reveal experts making some systematic judgment and decision errors, even while these more experienced, sophisticated actors outperform nonexperts in some settings. In fact, some of the earliest studies of intuitive judgment biases used experts in statistics as participants.[323] Other experimental studies and field evidence show biases in the clinical judgments of doctors, psychiatrists, and other health professionals.[324]

The evidence also reveals systematic errors by professionals with expertise in tasks that require complex judgments and decisions in business and finance.[325] For instance, studies found anchoring effects among veteran accountants and real estate brokers,[326] desirability bias among investment

---

322. *Id.* at 521–22 (discussing these limitations and providing some illustrations).

323. This was the case, for example, with some of the famous early studies of heuristics and biases that Amos Tversky and Daniel Kahneman conducted. *See* Daniel Kahneman & Amos Tversky, *On the Psychology of Prediction*, 80 PSYCHOL. REV. 237, 238 (1973) (describing studies showing judgment errors using, *inter alia*, a large sample of graduate students in psychology at three major U.S. universities); Amos Tversky & Daniel Kahneman, *Belief in the Law of Small Numbers*, 76 PSYCHOL. BULL. 105, 105 (1971) (using as experimental subjects the experts participating in meetings of the Mathematical Psychology Group and the American Psychological Association).

324. *See, e.g.*, Hal R. Arkes et al., *Hindsight Bias Among Physicians Weighing the Likelihood of Diagnoses*, 66 J. APPLIED PSYCHOL. 252 (1981) (physicians exhibiting the hindsight bias); Loren J. Chapman & Jean P. Chapman, *Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs*, 74 J. ABNORMAL PSYCHOL. 271 (1969) (erroneous intuitive beliefs in psychotherapists); Barbara J. McNeil et al., *On the Elicitation of Preferences for Alternative Therapies*, 306 NEW ENG. J. MED. 1259, 1261–62 (1982) (physicians showing framing effects); Roy M. Poses & Michele Anthony, *Availability, Wishful Thinking, and Physicians' Diagnostic Judgments for Patients with Suspected Bacteremia*, 11 MED. DECISION MAKING 159, 159 (1991) (physicians influenced by "the availability heuristic and by wishful thinking, a form of the value bias"). *See generally* Katherine H. Hall, *Reviewing Intuitive Decision-Making and Uncertainty: The Implications for Medical Education*, 36 MED. EDUC. 216 (2002) (providing a detailed description of associated heuristics and biases in clinical decision making); Natalia Karelaia & Robin M. Hogarth, *Determinants of Linear Judgment: A Meta-Analysis of Lens Model Studies*, 134 PSYCHOL. BULL. 404 (2008).

325. *Cf.* Derek J. Koehler et al., *The Calibration of Expert Judgment: Heuristics and Biases Beyond the Laboratory*, *in* HEURISTICS AND BIASES, *supra* note 85, at 686, 710 (concluding that expert judgment was miscalibrated in line with the qualitative predictions of the heuristics and biases approach in all of the five domains surveyed but that the magnitude of bias was greater in areas such as medical, business, and sports judgments, where experts had less training and technical assistance in statistical modeling).

326. *See* Edward J. Joyce & Gary C. Biddle, *Anchoring and Adjustment in Probabilistic Inference in Auditing*, 19 J. ACCT. RES. 120 (1981) (accountants); Gregory B. Northcraft &

managers,[327] subadditive judgments by options traders,[328] and framing effects among financial planners.[329]

Still, one might hope for a better alignment with rational models on the part of top corporate managers due to selection effects.[330] The managers whose behavior is most relevant for antitrust purposes are not only expert, experienced business decision makers; they also belong to a smaller, more select group that reaches elevated positions on the corporate ladder. These managers may differ from other professionals both in their stronger drive to succeed in the business world and in consistently outperforming their competitors in the intra-firm tournament for top management positions. Hence these more accomplished, better-performing managers might also be more rational than their typical competitors.[331]

Although selection effects can promote more rational behavior among senior corporate managers, however, both theory and evidence suggest these processes are of limited efficacy. In some respects, the limited efficacy of intra-firm competitive selection echoes the limits of marketplace competition discussed above at length.[332] Yet certain mechanisms of market rationality are even more constrained or altogether irrelevant where managerial rationality is concerned: managerial behavior, by definition, is a matter of individual, not aggregate, macro-level rationality. And managerial rationality is an even less likely target for arbitrage than firm-level conduct in product markets.[333]

Margaret A. Neale, *Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions*, 39 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 84 (1987) (real estate brokers).

327. *See* Robert A. Olsen, *Desirability Bias Among Professional Investment Managers: Some Evidence from Experts*, 10 J. BEHAV. DECISION MAKING 65, 66–70 (1997).

328. *See* Craig R. Fox, Brett A. Rogers & Amos Tversky, *Options Traders Exhibit Subadditive Decision Weights*, 13 J. RISK & UNCERTAINTY 5, 16 (1996).

329. Michael J. Roszkowski & Glenn E. Snelbecker, *Effects of "Framing" on Measures of Risk Tolerance: Financial Planners Are Not Immune*, 19 J. BEHAV. ECON. 237 (1990).

330. *See* Werden et al., *supra* note 7, at 134 ("[F]irms do not randomly select individuals from the general population to make their important decisions, but rather hire and promote employees on the basis of their skills."); *cf.* Wright & Stone, *supra* note 7, at 1525 ("[T]he firm is not merely a heterogeneous hodgepodge of individuals, but an institution constructed to lower transaction costs relative to making use of the price system . . . .").

331. *Cf.* Werden et al., *supra* note 7, at 134 ("Even if most individuals make badly biased decisions in the face of risk and uncertainty, Wall Street analysts do not because they are selected for their understanding of probability theory."); Langevoort, *supra* note 312, at 442–45 (making a similar observation in the corporate governance context).

332. *See supra* subpart III(A).

333. *See* Baker & Nofsinger, *supra* note 189, at 8 (suggesting that intra-firm behavior will be further removed from strict rationality than market behavior given the greater limits of arbitrage in this setting and the concentration of decision-making power in a few hands).

Moreover, managerial tournaments at best reward those performance elements that most closely correlate with the firm's long-run profitability.[334] Because rationality only partly correlates with the firm's success, even effective competitive selection within the firm will promote managers who exhibit some systematic deviations from rationality. For instance, successful managers may benefit from a reputation for consistency and commitment, which can lead them to take into account sunk costs that rational actors are supposed disregard.[335]

Similarly, managerial tournaments may promote overconfidence—a term denoting a cluster of loosely related deviations from rational judgment[336] that has received much scholarly attention in recent years.[337]

---

334. In reality, intra-firm promotions also depend on additional factors that may be only tenuously related to the firm's long-run profitability, from the manager's ability to avoid association with failure through various social and cultural factors. Insofar as they exert significant influence on the outcome of managerial tournaments, therefore, these factors further attenuate the rationality-promoting potential of intra-firm competition.

335. *See* Dawes, *supra* note 87, at 500–02 (describing the attention to sunk costs as an ambiguous anomaly on this and related grounds); Langevoort, *supra* note 312, at 444 (noting that attention to sunk costs may be beneficial for managers); *see also* Barry M. Staw, *The Escalation of Commitment: An Update and Appraisal*, *in* ORGANIZATIONAL DECISION MAKING 191, 210 (Zur Shapira ed., 1997) (asserting that the debate over the rationality of escalation decisions in organizational settings should be replaced with an effort to account for the behavioral evidence in this domain).

336. As a term of art, overconfidence refers both to the miscalibration of probability estimates—when decision makers are too confident in the accuracy of their judgments—as well as optimistic overconfidence in one's relative or absolute performance, outcomes, and so on. *See* Markus Glaser & Martin Weber, *Overconfidence*, *in* BEHAVIORAL FINANCE, *supra* note 189, at 241, 242 (stating, albeit with some imprecision, that "[t]he two main facets of overconfidence are miscalibration and the better-than-average effect"). For an introduction to research on miscalibration, see Dale Griffin & Lyle Brenner, *Perspectives on Probability Judgment Calibration*, *in* BLACKWELL HANDBOOK OF JUDGMENT AND DECISION MAKING 177 (Derek J. Koehler & Nigel Harvey eds., 2004), which reviews some key findings and approaches in this area, and Sarah Lichtenstein et al., *Calibration of Probabilities: The State of the Art to 1980*, *in* JUDGMENT UNDER UNCERTAINTY, *supra* note 298, at 306, which provides an earlier, paradigm-setting contribution. For a summary of key findings on optimistic overconfidence, see Tor, *Entry*, *supra* note 2, at 504–14. *See also* HEURISTICS AND BIASES, *supra* note 85, at 230–378 (reviewing and discussing phenomena relating to both types of overconfidence).

337. *See, e.g.*, Itzhak Ben-David et al., *Managerial Miscalibration*, 128 Q.J. ECON. 1547, 1547 (2013) (finding that "executives are severely miscalibrated" and that "realized market returns are within the executives' 80% confidence intervals only 36% of the time"); Wen-I Chuang & Bong-Soo Lee, *An Empirical Evaluation of the Overconfidence Hypothesis*, 30 J. BANKING & FIN. 2489 (2006) (finding empirical evidence that overconfident investors overreact to private information and underreact to public information and that market gains increase investors' overconfidence so they trade more aggressively in subsequent periods); Stephen P. Ferris et al., *CEO Overconfidence and International Merger and Acquisition Activity*, 48 J. FIN. & QUANTITATIVE ANALYSIS 137, 137 (2013) (finding that "[o]verconfidence helps to explain the number of offers made by a CEO, the frequencies of nondiversifying and diversifying acquisitions, and the use of cash to finance a merger deal"); Simon Gervais et al., *Overconfidence, Compensation Contracts, and Capital Budgeting*, 66 J. FIN. 1735, 1761 (2011) [hereinafter Gervais et al., *Capital Budgeting*] (studying "the interaction of managerial overconfidence and compensation in the context of a firm's investment policy"); Anand M. Goel & Anjan V. Thakor, *Overconfidence, CEO Selection, and Corporate Governance*, 63 J. FIN. 2737 (2008) (showing that

For example, overconfident managers may persevere in difficult situations, exhibit greater ambition and confidence in their performance, and disproportionately attribute their successes to their own prowess over luck, all of which may make them more attractive to the firm than their unbiased peers and, consequently, more likely to be selected for top positions.[338]

Of course, overconfidence is not all-around beneficial for either the managers or their firms. On the one hand, a recent empirical study shows that firms in innovative industries with overconfident CEOs invest more in innovation, obtain more patents and patent citations, and achieve greater innovative success for given research and development expenditures (but also have more volatile returns).[339] Some models further show that

---

an overconfident manager "has a higher likelihood than a rational manager of being deliberately promoted to CEO under value-maximizing corporate governance" (emphasis omitted)); Dirk Hackbarth, *Managerial Traits and Capital Structure Decisions*, 43 J. FIN. & QUANTITATIVE ANALYSIS 843, 843 (2008) ("Optimistic and/or overconfident managers choose higher debt levels and issue new debt more often . . . ."); Ulrike Malmendier & Geoffrey Tate, *CEO Overconfidence and Corporate Investment*, 60 J. FIN. 2661, 2661 (2005) [hereinafter Malmendier & Tate, *CEO Overconfidence*] ("Overconfident managers overestimate the returns to their investment projects and view external funds as unduly costly."); Ulrike Malmendier & Geoffrey Tate, *Does Overconfidence Affect Corporate Investment? CEO Overconfidence Measures Revisited*, 11 EUR. FIN. MGMT. 649 (2005) [hereinafter Malmendier & Tate, *Corporate Investment*] (linking CEO overconfidence with firm investment decisions); Jijun Niu, *The Effect of CEO Overconfidence on Bank Risk Taking*, 30 ECON. BULL. 3288 (2010) (finding that banks managed by overconfident CEOs take more risks than their peer institutions); Alexander Puetz & Stefan Ruenzi, *Overconfidence Among Professional Investors: Evidence from Mutual Fund Managers*, 38 J. BUS. FIN. & ACCT. 684 (2011) (finding that equity mutual fund managers trade more after good personal performance and that market performance has no significant impact); Catherine M. Schrand & Sarah L.C. Zechman, *Executive Overconfidence and the Slippery Slope to Financial Misreporting*, 53 J. ACCT. & ECON. 311, 311 (2012) (concluding that "[o]verconfident executives are more likely to exhibit an optimistic bias and thus are more likely to start down a slippery slope of growing intentional misstatements"); Anwer S. Ahmed & Scott Duellman, *Managerial Overconfidence and Accounting Conservatism* (Mays Bus. Sch., Tex. A&M Univ., Research Paper No. 2012-77, 2012), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2097895 (finding a negative correlation between CEO overconfidence and accounting conservatism and noting that external monitoring does not appear to mitigate this effect); Alberto Galasso & Timothy S. Simcoe, *CEO Overconfidence and Innovation* (Nat'l Bureau of Econ. Research, Working Paper No. 16041, 2010), *available at* http://www.nber.org/papers/w16041.pdf (reporting that overly optimistic and overconfident CEOs are more likely to pursue innovation); Ulrike Malmendier et al., *Corporate Financial Policies with Overconfident Managers* (Nat'l Bureau of Econ. Research, Working Paper No. 13570, 2007) [hereinafter Malmendier et al., *Overconfident Managers*], *available at* http://www.nber.org/papers/w13570 (linking individual managerial characteristics, particularly overconfidence, to firm financing decisions).

338. *See* Goel & Thakor, *supra* note 337 (showing that overconfident managers are more likely to win intra-firm promotions); Langevoort, *supra* note 312, at 444 (noting that overconfidence can be a useful and highly adaptive trait in a business environment).

339. *See* David Hirshleifer et al., *Are Overconfident CEOs Better Innovators?* 67 J. FIN. 1457 (2012); *see also* Galasso & Simcoe, *supra* note 337 (finding a robust positive association between CEO overconfidence and citation-weighted patent counts, with a larger effect in more competitive industries, in a sample of large public firms from 1980 to 1994).

overconfidence or mild optimism can better align managerial behavior with shareholder interests.[340]

On the other hand, behavioral corporate finance research also reveals that banks with overconfident CEOs take greater risks than their peer institutions,[341] and top-performing mutual fund managers tend to trade more following their success—to a degree not explained by other factors—and exhibit worse performance when they do so.[342] Studies further show that managerial overconfidence distorts both investment and financing decisions at the firm level;[343] helps explain the volume, type, and financing of mergers and acquisitions activity;[344] and is even linked to aggressive accounting and an increased likelihood of financial misreporting.[345]

All in all, the evidence makes clear that managerial overconfidence and certain other deviations from strict rationality can survive—sometimes even thrive on—intra-firm selection processes. Most importantly, both theory and the rapidly accumulating evidence also show that behavioral phenomena like managerial overconfidence indeed exert significant, measurable effects on firm-level conduct in the market.

2. Firms.—Beyond their potentially superior individual rationality, managers also operate within large, complex business organizations that should be capable of generating better outcomes than individuals do, for a number of reasons. First, when firms have the time and means to learn from experience and repeated feedback, they can develop "organizational repairs"—that is, internal procedures and rules that aim to overcome systematic individual shortcomings.[346] The management literature provides anecdotal illustrations, for example, of organizations using maxims intended to remind employees not to make biased attributions; utilizing strategies aimed at collecting sufficient, relevant information; and developing methods for evaluating their information and hypotheses more objectively.[347]

---

340. *See, e.g.*, Gervais et al., *Capital Budgeting*, *supra* note 337.

341. *See* Niu, *supra* note 337.

342. *See* Puetz & Ruenzi, *supra* note 337.

343. *See, e.g.*, Malmendier & Tate, *CEO Overconfidence*, *supra* note 337; Malmendier & Tate, *Corporate Investment*, *supra* note 337; Malmendier et al., *Overconfident Managers*, *supra* note 337.

344. *See* Ferris et al., *supra* note 337.

345. *See* Schrand & Zechman, *supra* note 337.

346. *See generally* Chip Heath et al., *Cognitive Repairs: How Organizational Practices Can Compensate for Individual Shortcomings*, 20 RES. ORGANIZATIONAL BEHAV. 1, 4–12 (1998) (discussing various common judgment and decision errors and suggesting ways organizations may attempt to correct them and also providing anecdotal evidence for such cognitive repairs).

347. *See id.*

Nevertheless, organizational repairs have limited success and largely are unpredictable, tending to be most efficacious when based on bottom-up learning within the firm in a specific domain.[348] These characteristics, however, do not apply to most of the significant antitrust-related tasks managers face, which concern judgments and decisions regarding the firm's overall pricing or distribution strategy, strategic alliances with actual or potential competitors, mergers and acquisitions, and so on. The judgments and choices required in such cases are made infrequently, at the highest management levels, and usually offer only limited and noisy feedback, all of which make systemic organizational repairs unlikely.[349]

Second, managers may better approximate rational action simply because they function as agents of the firm.[350] There is some evidence that agents—who operate on behalf of others—act more rationally than individuals acting on their own behalf. For example, the endowment effect—wherein individuals value entitlements they possess more highly than identical ones they do not hold[351]—was not manifested by experimental participants taking the role of agents and transacting on behalf of their principals.[352] In the same vein, the behavioral evidence suggests

348. *Cf. id.* at 12–16 (discussing various classifications of repairs along different dimensions and their likely efficacy).

349. *See id.* at 12–15 (discussing methods of social feedback). *See generally supra* notes 298–99 and accompanying text (discussing factors that make learning difficult in many situations).

350. The agency relationship between managers and firms also generates some disadvantages, most notably due to the potential divergence of the parties' self-interest, which is of lesser concern here. For further background on managerial incentives and agency costs, see generally FRANK H. EASTERBROOK & DANIEL R. FISCHEL, THE ECONOMIC STRUCTURE OF CORPORATE LAW 90–108 (1991); Eugene F. Fama, *Agency Problems and the Theory of the Firm*, 88 J. POL. ECON. 288 (1980); Michael C. Jensen & William H. Meckling, *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, 3 J. FIN. ECON. 305 (1976); and Oliver E. Williamson, *Managerial Discretion and Business Behavior*, 53 AM. ECON. REV. 1032 (1963). For a review of more recent corporate governance research, see Bebchuk & Weisbach, *supra* note 314.

351. *See* Daniel Kahneman et al., *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, J. ECON. PERSP., Winter 1991, at 193, 194–97; Richard Thaler, *Toward a Positive Theory of Consumer Choice*, 1 J. ECON. BEHAV. & ORG. 39, 43–47 (1980) (describing the endowment effect and providing examples of its application in business strategies and everyday life); *see also* Russell Korobkin, *The Endowment Effect and Legal Analysis*, 97 NW. U. L. REV. 1227 (2003) (applying the endowment effect to several areas of law).

352. Jennifer Arlen et al., *Endowment Effects Within Corporate Agency Relationships*, 31 J. LEGAL STUD. 1, 31–32 (2002) (finding that experimental participants acting as agents did not exhibit a significant endowment effect because they framed entitlements in terms of exchange value). Another study found a significant decrease in fairness concerns when participants in a bargaining transaction acted as agents owing a duty—such as that of corporate managers—to maximize the return to the principal. *See* Kent Greenfield & Peter C. Kostant, *An Experimental Test of Fairness Under Agency and Profit-Maximization Constraints (With Notes on Implications for Corporate Governance)*, 71 GEO. WASH. L. REV. 983, 1000, 1003–04 (2003).

that egocentric biases are less likely to impact judgments made on behalf of others insofar as the agent has not adopted the principal's perspective.[353]

The better alignment of agent judgment and choice with rational action, however, would be of limited assistance to managers in overcoming those judgment and decision errors they manifest with respect to major corporate decisions. For one, even the limited evidence of agents' increased rationality pertains only to a few of those behavioral phenomena that can impact antitrust-relevant behavior. Furthermore, agents' rationality advantages over principals do not apply to most antitrust-relevant managerial tasks. The experimental elimination of agents' endowment effect, for example, was driven by participants' framing of the entitlements they traded based on the exchange value of those entitlements.[354] The impact of loss aversion on key antitrust-relevant decisions, on the other hand, concerns the managers' own strategic decisions about the overall course of the firm rather than about entitlements such as goods held by the firm for routine transactions.[355] Agents' advantage regarding egocentric biases similarly is unlikely to pertain to judgments of their own managerial ability and expertise. More generally, the greater rationality of agents is less applicable to managers' judgments and decisions concerning their own abilities, plans, and performance.[356]

---

353. For instance, much of the evidence of optimistic bias comes from studies that compare participants' beliefs about their own prospects with their beliefs about the prospects of others. Hence the evidence that shows a systematic bias with respect to beliefs about oneself implies an unbiased (or at least a less optimistic) view of the prospects of third parties. *See, e.g.*, Tor, *Entry*, *supra* note 2, at 505–08 (citing some of the key studies showing overoptimistic judgments). Similarly, the evidence regarding the moderating role of preference intensity on the processes of optimistic overconfidence, *id.* at 520–22 (discussing this evidence), also implies that agents may be less biased when making judgments on behalf of their principals.

354. *See, e.g.*, Daniel Kahneman et al., *Experimental Tests of the Endowment Effect and the Coase Theorem*, 98 J. POL. ECON. 1325, 1325 (1990) (finding no endowment effect when using induced-value tokens); Amos Tversky & Daniel Kahneman, *Loss Aversion in Riskless Choice: A Reference-Dependent Model*, 106 Q.J. ECON. 1039, 1055 (1991) ("[R]eluctance to sell is surely absent in routine commercial transactions, in which goods held for sale have the status of tokens for money."). *But see* Ian Bateman et al., *A Test of the Theory of Reference-Dependent Preferences*, 112 Q.J. ECON. 479 (1997) (finding some loss aversion for monetary payoffs); Ian Bateman et al., *Testing Competing Models of Loss Aversion: An Adversarial Collaboration*, 89 J. PUB. ECON. 1561, 1576–77 (2005) (same). *See generally* Nathan Novemsky & Daniel Kahneman, *The Boundaries of Loss Aversion*, 42 J. MARKETING RES. 119 (2005) (exploring the boundary conditions of loss aversion).

355. Tor & Rinner, *supra* note 7, at 829–30 (applying this distinction to managers' decisions to employ RPM in their distribution systems); *see also* Eric van Dijk & Daan van Knippenberg, *Buying and Selling Exchange Goods: Loss Aversion and the Endowment Effect*, 17 J. ECON. PSYCHOL. 517, 517 (1996) (finding that experimental market participants exhibited loss aversion for exchange goods when traders were uncertain about future exchange prices).

356. *Cf.* Tor, *Entry*, *supra* note 2, at 535–36 (arguing that the advantage of financiers over new entrants in making decisions regarding new ventures diminishes when they adopt the entrants' perspective). *See generally* MAX H. BAZERMAN & DON A. MOORE, JUDGMENT IN MANAGERIAL DECISION MAKING (7th ed. 2009) (reviewing and applying individual-level phenomena to managerial decision making); LEE ROY BEACH & TERRY CONNOLLY, THE

Third, corporate managers nonetheless may exhibit superior performance because they often do not make significant judgments and decisions alone but in a small group of top managers or the corporate board of directors, with the benefits of multiple viewpoints, cumulative experience, and deliberation.[357]

Despite its intuitive appeal, however, the empirical evidence does not support the claim that boards (or top management groups) will reliably avoid those systematic decision errors that plague individual managers. Instead, the evidence shows small groups outperform individual rationality in some cases but at other times exhibit similar or even more extreme judgmental biases and decision errors, with their ultimate performance largely dependent on case-specific variables.[358]

PSYCHOLOGY OF DECISION MAKING: PEOPLE IN ORGANIZATIONS (2d ed. 2005) (same); Leigh Thompson & Jo-Ellen Pozner, *Organizational Behavior*, *in* SOCIAL PSYCHOLOGY: HANDBOOK OF BASIC PRINCIPLES 913, 914 (Arie W. Kruglanski & E. Tory Higgins eds., 2d ed. 2007) (reviewing research on individual decision making in organizations and stating that "[t]he fundamental theme is that organizational decision makers . . . are hopelessly victimized by their own nonrational thought processes").

357. *See* Stephen M. Bainbridge, *Why a Board? Group Decisionmaking in Corporate Governance*, 55 VAND. L. REV. 1, 19–31 (2002) (arguing that boundedly rational managers function optimally on a board with diverse viewpoints); Cass R. Sunstein, *Group Judgments: Statistical Means, Deliberation, and Information Markets*, 80 N.Y.U. L. REV. 962, 978–79 (2005) (discussing the intuitive appeal of reliance on deliberating groups to make better judgments and decisions than individuals and their extensive use in various domains, including corporate boards). *But see* Donald C. Langevoort, *The Human Nature of Corporate Boards: Law, Norms, and the Unintended Consequences of Independence and Accountability*, 89 GEO. L.J. 797, 800 (2001) (asserting that for a board to function optimally, some board members may need to function as mediators to mitigate the group's inevitable polarization); Donald C. Langevoort, Essay, *Resetting the Corporate Thermostat: Lessons from the Recent Financial Scandals About Self-Deception, Deceiving Others and the Design of Internal Controls*, 93 GEO. L.J. 285, 315–16 (2004) (advocating enhanced internal-reporting controls that increase as the risk of self-serving managerial behavior increases).

358. *See* Daniel Gigone & Reid Hastie, *Proper Analysis of the Accuracy of Group Judgments*, 121 PSYCHOL. BULL. 149 (1997) (reviewing the literature and concluding that groups excel as judges only under limited conditions and tend to perform at the level of their average members when performing tasks whose solutions are not easily demonstrable); Gayle W. Hill, *Group Versus Individual Performance: Are* N + *1 Heads Better Than One?*, 91 PSYCHOL. BULL. 517, 535 (1982) (providing an extensive literature review finding across a variety of tasks that "group performance was generally qualitatively and quantitatively superior to the performance of the average individual" but that it was "often inferior to that of the best individual in a statistical aggregate and often inferior to the potential suggested in a statistical pooling model"); Norbert L. Kerr et al., *Bias in Judgment: Comparing Individuals and Groups*, 103 PSYCHOL. REV. 687 (1996) (reviewing the empirical literature on the relative susceptibility of individuals and groups to systematic judgmental biases and finding that there is no clear or general pattern); Norbert L. Kerr & R. Scott Tindale, *Group Performance and Decision Making*, 55 ANN. REV. PSYCHOL. 623 (2004) (reviewing some of the main findings in this area); John M. Levine & Richard L. Moreland, *Small Groups*, *in* 2 THE HANDBOOK OF SOCIAL PSYCHOLOGY, *supra* note 87, at 415, 438–39 (same). *But see* R. Scott Tindale et al., *Group Decision Making*, *in* THE SAGE HANDBOOK OF SOCIAL PSYCHOLOGY 381 (Michael A. Hogg & Joel Cooper eds., 2003) (reviewing a number of research strands in group research and arguing that they show the general superiority of groups, despite some unique biases and problems in their decision making). For

Moreover, beyond their limited capacity to ameliorate individuals' errors, some common characteristics of group decision making—most notably deliberation—can generate additional, group-level biases.[359] Groups, for instance, may exhibit "groupthink," promoting an erroneous consensus that does not reflect the information held by individual group members.[360] Their deliberations, instead of leading to a superior integration

---

examples of specific studies comparing individuals and groups, see Linda Argote et al., *The Base-Rate Fallacy: Contrasting Processes and Outcomes of Group and Individual Judgment*, 46 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 296 (1990), finding that group discussion amplifies judgment by representativeness when individuating information is informative but also increases the normatively appropriate impact of base rates when information is not representative; Roger Buehler et al., *Collaborative Planning and Prediction: Does Group Discussion Affect Optimistic Biases in Time Estimation?*, 97 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 47 (2005), detailing studies showing an optimistic bias for both individual and group predictions, with the latter being more optimistic than those generated individually; Chip Heath & Rich Gonzalez, *Interaction with Others Increases Decision Confidence But Not Decision Quality: Evidence Against Information Collection Views of Interactive Decision Making*, 61 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 305 (1995), arguing that interaction with others in the decision-making process increases the decider's confidence but not his decision quality; L. Robin Keller et al., *An Examination of Ambiguity Aversion: Are Two Heads Better than One?*, 2 JUDGMENT & DECISION MAKING 390 (2007), finding that a majority of dyads exhibited greater ambiguity aversion than two individual subjects' average; Richard F. Martell & Mae R. Borg, *A Comparison of the Behavioral Rating Accuracy of Groups and Individuals*, 78 J. APPLIED PSYCHOL. 43 (1993), finding that groups' delayed ratings of the behavior of others were more accurate than those of individuals but demonstrated greater response bias; Paul W. Paese et al., *Framing Effects and Choice Shifts in Group Decision Making*, 56 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 149, 160–63 (1993), finding, *inter alia*, that groups can increase or decrease individual framing effects depending on how decisions are presented; and Glen Whyte, *Escalating Commitment in Individual and Group Decision Making: A Prospect Theory Approach*, 54 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 430 (1993), finding that groups exhibit more extreme escalation of commitment.

359. *See* Joyce Berg et al., *The Individual Versus the Aggregate*, *in* JUDGMENT AND DECISION-MAKING RESEARCH IN ACCOUNTING AND AUDITING 102 (Robert H. Ashton & Alison Hubbard Ashton eds., 1995) (developing a framework for comparing individual performance to group settings and providing evidence of, among other things, group-level biases and individual biases that extend to groups, concluding that the impact of aggregation on individual-level biases varies widely); Sunstein, *supra* note 357, at 984–86 (discussing the role of information and social influence in contributing to the failure of deliberation in groups to consistently produce rational outcomes). *See generally* BLACKWELL HANDBOOK OF SOCIAL PSYCHOLOGY: GROUP PROCESSES (Michael A. Hogg & R. Scott Tindale eds., 2001) (providing a collection of articles reviewing group processes that introduce additional complexity and phenomena beyond those found in individuals).

360. *See, e.g.*, IRVING L. JANIS, GROUPTHINK (2d ed., rev. 1983) (providing the original development of the concept and its applications); Robert S. Baron, *So Right It's Wrong: Groupthink and the Ubiquitous Nature of Polarized Group Decision Making*, 37 ADVANCES EXPERIMENTAL SOC. PSYCHOL. 219 (2005) (reviewing thirty years of research and debate over Janis's groupthink model and concluding, *inter alia*, that groupthink-like phenomena are common in mundane, temporary, and even minimal groups, though not universally part of group decision making); James K. Esser, *Alive and Well After 25 Years: A Review of Groupthink Research*, 73 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 116 (1998) (providing a summary of empirical research on groupthink theory). *See generally* Robert J. MacCoun, *Comparing Micro and Macro Rationality*, *in* JUDGMENTS, DECISIONS, AND PUBLIC POLICY 116, 121–26 (Rajeev

of group members' information and perspectives, can cause group polarization so that the resulting collective view of the group is more extreme than the individual members' predeliberation tendencies.[361] Hence, while senior management's collective judgment and decision making can and will sometimes generate superior performance, there is little reason to believe they will approximate the predictions of rational models across the board.

Finally, corporate governance scholarship suggests that in reality corporate boards possess limited efficacy and often are dominated by CEOs and senior management.[362] Even with the gradual shift in recent years towards increased board power at the expense of management, corporate boards are unlikely to shape most senior management's significant, antitrust-relevant judgments and decisions.[363]

In sum, while the market behavior of firms and of the mangers who make judgments and decisions on their behalf may approximate rational models in some cases, the empirical behavioral evidence reveals a more complex reality. Behavioral antitrust scholars who assume that firms replicate all (and only) individual-level deviations from strict rationality often will be wrong and so will those analysts who make the mirroring assumption that intra-firm processes guarantee strictly rational conduct.

Gowda & Jeffrey C. Fox eds., 2002) (reviewing variables that sometimes cause groups to exhibit less accurate judgments than individuals).

361. *See, e.g.*, Daniel J. Isenberg, *Group Polarization: A Critical Review and Meta-Analysis*, 50 J. PERSONALITY & SOC. PSYCHOL. 1141 (1986) (examining polarization studies focusing on the two central accounts for this effect: social comparison and persuasive argumentation processes); *see also* Sunstein, *supra* note 357, at 984–1006 (reviewing biases that may be generated by group deliberation and dividing their underlying mechanisms into informational influences and social pressures).

362. *See, e.g.*, LUCIAN BEBCHUK & JESSE FRIED, PAY WITHOUT PERFORMANCE: THE UNFULFILLED PROMISE OF EXECUTIVE COMPENSATION (2004) (exploring the ways in which managerial power and influence have shaped the executive-compensation landscape); Iman Anabtawi, *Explaining Pay Without Performance: The Tournament Alternative*, 54 EMORY L.J. 1557, 1574–79 (2005) (analyzing the relationship between managerial power and boards' not entering into efficient contracts with managers); Sanjai Bhagat & Bernard Black, *The Non-Correlation Between Board Independence and Long-Term Firm Performance*, 27 J. CORP. L. 231, 263 (2002) (reporting data suggesting that greater board independence does not lead to greater firm performance). *See generally* Renée B. Adams et al., *The Role of Boards of Directors in Corporate Governance: A Conceptual Framework and Survey*, 48 J. ECON. LITERATURE 58 (2010) (reviewing theoretical models and empirical findings concerning the behavior and effectiveness of corporate boards of directors); Benjamin E. Hermalin & Michael S. Weisbach, *Boards of Directors as an Endogenously Determined Institution: A Survey of the Economic Literature*, FED. RES. BANK N.Y. ECON. POL'Y REV., Apr. 2003, at 7 (providing an earlier and an updated review of the relevant economic literature on board composition and actions and their effects on firm performance).

363. *See, e.g.*, Tor & Rinner, *supra* note 7, at 851–52 (discussing the limited efficacy of boards in the RPM context).

IV. The Third Mistake: Confusing Bounded Rationality with Anticompetitiveness

Unlike the first two categories of mistakes—which primarily cause errors in scholars' understanding of antitrust actors' behavior—the third and last category of mistakes leads to errors in the normative evaluation of deviations from standard assumptions of rationality. Many commentators mistakenly equate deviations from these assumptions with privately or socially suboptimal behavior and even with anticompetitive outcomes that necessarily would justify antitrust scrutiny.[364] This mistaken chain of inference leads analysts to embrace or reject the behavioral approach based on their respective antitrust policy predispositions rather than the merits of the evidence. In reality, however, many systematic deviations from strict rationality are of no antitrust concern: some are purely procompetitive or at least procompetitive on balance, and even some socially undesirable consequences of bounded rationality do not generate sufficient competitive harm to merit antitrust intervention.

A. Procompetitive Deviations

Some deviations from the assumptions of rationality clearly are procompetitive. Standard models in antitrust law and economics assume, for instance, that producers determine their behavior in the market based solely on the expected value of the different options available to them.[365] Such producers, for instance, form and maintain cartels whenever the expected economic benefits of cartelization outweigh its expected economic costs.[366] Cartels can be extremely profitable, and the probability of their detection still is quite low, although the advent of successful leniency

364. *See, e.g.*, Cooper & Kovacic, *supra* note 7 at 800 (noting that "[m]uch work in the nascent field of behavioral antitrust prescribes expanded use of competition law to correct consumer harm that arises from biased firm behavior"); *Ginsburg* & Moore, *supra* note 7, at 96–98 (arguing that "BE could serve only to broaden, rather than to narrow[,] the meaning of the term 'unfair'"); Huffman, *Neo-Chicago with Behavioral Antitrust*, *supra* note 7, at 106 (expressing the view that while "[l]ike foregoing antitrust economics movements, Behavioral Antitrust is on its face result-neutral, . . . as it has been discussed to date, it has a political slant. Until very recently, all of the writing advocating Behavioral Antitrust favored increased antitrust enforcement.).

365. More precisely, although it is not material for the purposes of the present example, standard models assume that firms behave as to maximize risk-adjusted net present value. *See, e.g.*, Tor, *Entry*, *supra* note 2, at 489–90 (discussing the net present value concept and its application to antitrust economics). For other general applications, see AREEDA & HOVENKAMP, *supra* note 3; RICHARD A. BREALEY ET AL., PRINCIPLES OF CORPORATE FINANCE 23, 101–05, 127–45, 801–05 (10th ed. 2011); JEAN TIROLE, THE THEORY OF INDUSTRIAL ORGANIZATION 34–35 (1988).

366. *See generally* John M. Connor, *Global Cartels Redux: The Lysine Antitrust Litigation*, *in* THE ANTITRUST REVOLUTION: ECONOMICS, COMPETITION, AND POLICY 300, 305–06 (John E. Kwoka, Jr. & Lawrence J. White eds., 5th ed. 2009) (discussing the economics and law of cartels).

programs has increased it dramatically.[367] So, with strictly rational managers, cartels would have been not only common—as they appear to be[368]—but ubiquitous.

Yet if some real managers are law abiding, or at least attribute some value to legal compliance beyond accounting for expected sanctions alone, the real incidence of cartelization is lower than it would have been in a world populated with strictly rational actors. Whether they are more law-abiding because of moral considerations, due to social norms, or for fear of the extralegal costs associated with criminal conviction, real-world, boundedly rational managers thus may act more procompetitively than standard antitrust models assume.

In principle, a similar outcome—of procompetitive deviations from the assumption of rationality—should occur whenever managers place some positive value on compliance with the antitrust laws beyond what the expected legal sanction merits. Such monopolists, for example, may avoid some profitable predatory actions toward weaker competitors. But the forces that contribute to legal compliance beyond that predicted by standard models are weaker in most areas of antitrust law beyond simple horizontal collusion, given the current dearth of bright-line rules that would make clear what conduct is illegal.[369] When very little conduct clearly is illegal, neither moral intuitions nor social norms of legal compliance are likely to

---

367. *See, e.g.*, Christopher R. Leslie, *Trust, Distrust, and Antitrust*, 82 TEXAS L. REV. 515, 642–43 (2004) (discussing the success of the Department of Justice's leniency program); Nathan H. Miller, *Strategic Leniency and Cartel Enforcement*, 99 AM. ECON. REV. 750 (2009) (developing a model of cartel behavior that helps overcome the difficulty that active cartels are never observed in the data, testing it empirically, and finding evidence consistent with enhanced detection and deterrence following the introduction of the Department of Justice leniency program); Gordon J. Klein, *Cartel Destabilization and Leniency Programs: Empirical Evidence* (Ctr. for European Econ. Research, Discussion Paper No. 10-107, 2010), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1854426## (offering econometric estimations based on OECD data for 23 countries and a period of 20 years that show positive and significant effects of leniency programs on the competition intensity as measured by price–cost margins).

368. *See* Stucke, *At the Gate*, *supra* note 2, at 565–68 (noting the susceptibility of certain industries, even those with low entry barriers, to price fixing).

369. The lack of bright line rules is apparent, for instance, in monopolization law. *See, e.g.*, Daniel A. Crane, *Rules Versus Standards in Antitrust Adjudication*, 64 WASH. & LEE L. REV. 49, 66 (2007) ("Monopolization law has always been more flexible and fact-sensitive. . . . Section 2 [of the Sherman Act] . . . contains no clear target [for adjudication] because all of a firm's amorphous conduct may be relevant to answering the question whether it unlawfully monopolized."); Tor, *supra* note 290, at 859 (explaining that "the common-law, open-ended monopolization standard of Section 2 incorporates no explicit rule-like elements"); *see also* William F. Adkinson, Jr. et al., *Enforcement of Section 2 of the Sherman Act: Theory and Practice* 16 (FTC & Dep't of Justice Working Paper on Section 2 Hearings, 2008), *available at* http://www.ftc.gov/os/sectiontwohearings/docs/section2overview.pdf ("Section 2's brief language offers little guidance in identifying prohibited conduct. Rather than defining its central concept—'monopolize'—the statute leaves that task to the courts." (footnote omitted)).

exert a pro-compliance influence.[370] Similarly, the lack of criminal liability in practice for most categories of antitrust violations removes managerial concerns about those extralegal sanctions that follow criminal convictions.[371]

Other nonstandard managerial preferences beyond valuing legal compliance can also generate procompetitive behavior. Oligopolistic coordination, for instance, is a common practice that causes significant competitive harm—much like the effect of explicit cartels—but is not prohibited by the antitrust laws.[372] Both theoretical models and experimental evidence suggest, however, that oligopolists that seek to protect and advance their market share, rather than simply to maximize profits, may find it harder to coordinate their behavior.[373] Hence the common preference for a superior relative position sometimes will generate more competitive markets.[374]

More commonly, however, even while not purely procompetitive, deviations from the assumptions of rationality can still be procompetitive on balance. Managerial overconfidence is a case in point: we saw that firms may select overconfident managers for a variety of reasons, and these managers impact firm-level behavior.[375] At times, managerial overconfidence leads to inefficient firm-level outcomes, as when it distorts investment or financing decisions.[376] Other effects appear more positive, such as where firms with overconfident managers generate more innovative activity.[377] Insofar as this increased innovative activity facilitates dynamic competition in the market more broadly, its procompetitive benefits may well outweigh the static efficiency losses generated by overconfidence-

370. *See, e.g.*, Yuval Feldman & Alon Harel, *Social Norms, Self-Interest and Ambiguity of Legal Norms: An Experimental Analysis of the Rule vs. Standard Dilemma*, 4 REV. L. & ECON. 81 (2008) (offering experimental evidence that social norms of noncompliance, but not those of compliance, exert a significant effect on the level of compliance when more ambiguous legal standards (rather than bright-line rules) are concerned).

371. *Cf.* V.S. Khanna, *Corporate Criminal Liability: What Purpose Does It Serve?*, 109 HARV. L. REV. 1477, 1492 (1996) (arguing that corporate criminal liability has "more severe and, arguably, unique sanctions (such as stigma)[] and a greater message-sending role than corporate civil liability"); Daniel S. Nagin, *Criminal Deterrence Research at the Outset of the Twenty-First Century*, 23 CRIME & JUST. 1, 19–22 (1998) (discussing evidence for the link between formal and informal criminal sanctions and referring to stigmatization as "the foundation of the deterrent effect"). *See generally* Uri Gneezy & Aldo Rustichini, *Incentives, Punishment, and Behavior*, *in* ADVANCES IN BEHAVIORAL ECONOMICS 572 (Colin F. Camerer et al. eds., 2004).

372. *See* POSNER, *supra* note 3, at 136–69; Louis Kaplow, *On the Meaning of Horizontal Agreements in Competition Law*, 99 CALIF. L. REV. 683 (2011).

373. *See supra* notes 117–21 and accompanying text.

374. *See supra* notes 117–21 and accompanying text.

375. *See supra* notes 334–38, 372 and accompanying text.

376. *See* Armstrong & Huck, *supra* note 7, at 26; Leslie, *supra* note 3, at 275–77; Tor, *Entry*, *supra* note 2, at 505–08.

377. Tor, *Entry*, *supra* note 2, at 540.

driven behavior.[378] Moreover, although overconfidence can lead to excessive managerial risk taking, it may in fact bring managers' behavior closer to—rather than farther away from—rationality when managers are too risk averse for other reasons.[379]

Optimistic overconfidence may also have procompetitive-on-balance consequences in other market settings. The phenomenon of excess entry into manufacturing industries, for instance, is partly driven by the psychology of overconfidence.[380] New entry is inefficient for those entrants who are making boundedly rational entry attempts with a negative net present value.[381] Moreover, excess entry typically is not procompetitive in the sense of exerting greater pressure on incumbents, notwithstanding the intuitive appeal of this proposition.[382] In the short term, overconfident entrants diminish the likelihood that other entrants will survive and prosper, with limited impact on incumbents.[383] But even while generating static efficiency losses, overconfident entry may be beneficial on balance: From a dynamic perspective, entrants' efforts can help other market participants, including future entrants, to identify and develop new products and services or to exploit potential efficiencies.[384] And while the ultimate balance of benefits and costs is not easily quantified, the beneficial, dynamic spillover effects of excess entry could well outweigh its static costs.[385]

---

378. *Id.* at 542.

379. *See, e.g.*, Gervais et al., *Capital Budgeting*, *supra* note 337, at 1749–50 (arguing that overconfidence can balance out managers' risk aversion); Richard H. Thaler, *Mental Accounting Matters*, 12 J. BEHAV. DECISION MAKING 183, 200–01 (1999) (detailing managers' propensity to suffer from "myopic loss aversion"); Tor, *Entry*, *supra* note 2, at 523 (discussing the incentives for managers to be overconfident but also pointing out that managers suffer professional and reputational effects if they fail); *see also* Daniel Kahneman & Dan Lovallo, *Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking*, 39 MGMT. SCI. 17 (1993) (discussing factors that contribute to managers' overly cautious conduct yet overly optimistic attitudes).

380. *See* Tor, *Entry*, *supra* note 2, at 490–91, 504–08 (describing the high rate of negative net present value entry in the manufacturing industry and the phenomenon of overconfidence); Colin Camerer & Dan Lovallo, *Overconfidence and Excess Entry: An Experimental Approach*, 89 AM. ECON. REV. 306 (1999) (presenting experimental evidence of overconfident entry).

381. *See* Tor, *Entry*, *supra* note 2, at 489–90.

382. *See* Werden et al., *supra* note 7, at 130 (noting that evidence of non-profit-maximizing entry "could suggest that entry is a more potent competitive force than the profit-maximization assumption suggests, but proponents of behavioral antitrust argue that non-profit-maximizing entry almost certainly is unsuccessful," thus finding this behavioral research to be of no direct relevance); Wright & Stone, *supra* note 7, at 385.

383. Tor, *Entry*, *supra* note 2, at 491–92, 522–24 (finding that most new entrants simply replace preceding entrants, while larger, diversifying entrants tend to be less overconfident and to exert more significant pressure on incumbents).

384. *See id.* at 540–43 (describing the consequences of innovative entry for the market).

385. *See id.* at 545 (explaining that there are social costs and social benefits associated with the spillover effect).

B. *Inefficient, Competitively Neutral Deviations*

Some deviations from strict rationality generate clear efficiency losses but do not raise antitrust concern. For instance, when RPM is legal, some manufacturers use it excessively, even for significant periods of time.[386] Boundedly rational RPM is inefficient and harmful for both the manufacturers exercising it and some of their retailers.[387] Nevertheless, the behavioral analysis of the practice also shows it is unlikely to generate competitive harm unless it is pervasive in the market or exercised by firms with substantial market power.[388]

The impact of considerations of fairness offers another example of inefficient, yet not anticompetitive, deviations from the assumptions of rationality. Standard economic theory expects producers fully to exploit their market power, raising product prices or limiting the quantities they produce, to maximize profits.[389] Yet both anecdotal evidence and controlled experimental tests suggest that firms do not always fully exploit their market power when it would have been rational to do so in a world populated solely by rational actors.[390] For example, firms enjoying short-term market power due to some external shock—such as when hurricane damage causes a dramatic increase in the demand for certain products—often avoid raising prices to market-clearing levels.[391] Producers may not exploit their power fully because they believe that charging higher prices is unfair or due to concerns about long-term reputational harm when fairness-minded consumers react negatively to these price hikes.[392] Either way, such fairness-minded behavior not only deviates from the standard prediction in

386. *See generally* Tor & Rinner, *supra* note 7, at 819–39 (describing why boundedly rational manufacturers have tended to rely upon RPM).

387. *See id.* at 839–42 (describing studies showing that RPM is inefficient and has become less popular).

388. *See id.* at 857 (suggesting that firms lacking market power are less capable of causing competitive harm).

389. THE ANTITRUST REVOLUTION, *supra* note 366, at 8–10. On occasion, rational firms will not exploit their market power fully, such as when they want to make potential new entry less attractive or wish to avoid scrutiny by the enforcement agencies. *See, e.g.*, Paul Milgrom & John Roberts, *Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis*, 50 ECONOMETRICA 443 (1982) (reviewing the early literature on limit pricing and providing a model showing that the practice is credible when potential entrants have incomplete information about incumbents' cost); *see also* PHILLIP AREEDA ET AL., ANTITRUST ANALYSIS: PROBLEMS, TEXT, AND CASES 488 (6th ed. 2004) (discussing unexploited market power).

390. *See* Kahneman et al., *supra* note 133 (offering a series of experimental demonstrations of how fairness impacts market choices, including the avoidance of fully exploiting short-term market power); *see also* Raymond F. Gorman & James B. Kehr, *Fairness as a Constraint on Profit Seeking: Comment*, 82 AM. ECON. REV. 355 (1992) (offering follow-up evidence with respect to business managers).

391. *See* Kahneman et al., *supra* note 133, at 738.

392. Bailey, *supra* note 81, at 6–7; *see also* Reeves & Stucke, *supra* note 2, at 1579–80 (suggesting that Merck did not sell a patented drug at monopoly prices due to potential adverse reputational effects).

antitrust economics but also causes a misallocation of social resources. In the face of a shortage, the market serves to match the limited available products or services with those consumers who place the highest value on them, as manifested by the prices they are willing to pay.[393] When producers avoid raising prices, however, they effectively allocate their goods through an inefficient queue system, on a first-come basis.

The illustrations offered here thus suffice to show how boundedly rational decision behavior by producers and consumers can generate efficiency losses that still fall short of raising antitrust concerns.

C. *Normative Bias?*

In principle, the behavioral approach is normatively neutral, an empirically driven effort to offer antitrust law a better understanding of market behavior.[394] Yet in practice, behavioral antitrust analyses currently more often promote a greater role for antitrust law, rather than a more limited one, due to the combined effects of the fundamental methodological error and the current, consistently prodefendant use of rationality assumptions in antitrust doctrine due to the Court's concerns regarding the costs and effects of antitrust litigation.[395] This state of affairs is not inevitable, however, and a well-developed behavioral approach could offer an important set of tools for antitrust scholars irrespective of their policy predispositions.[396]

The fundamental methodological error contributes both directly and indirectly to the tendency of behavioral antitrust to support a more active role for antitrust law. Directly, because proponents who treat behavioral phenomena as broad assumptions instead of concrete evidence tend to overstate the anticompetitive harm of deviations from assumptions of rationality. In fact, however, we saw that boundedly rational behavior can be purely procompetitive or procompetitive on balance, and not even all of its inefficient manifestations call for antitrust intervention.[397] Indirectly, since proponents who neglect the behavioral analysis of antitrust

393. This is true, of course, only under the common, if oft-criticized, economic approach that equates willingness to pay with utility, an approach that still underpins antitrust law's focus on protecting the competitive process.

394. Tor, *Behavioral Methodology*, *supra* note 1, at 314.

395. *See, e.g.*, Brooke Grp. Ltd. v. Brown & Williamson Tobacco Corp., 509 U.S. 209, 224 (1993) (holding that a "reasonable expectation of recover[y]" must be present to sustain antitrust liability for predatory pricing because otherwise the investment would be irrational).

396. *Cf.* Tor, *Behavioral Methodology*, *supra* note 1, at 314–17 (explaining that the normatively neutral behavioral toolbox can serve different normative goals from efficiency to justice or fairness).

397. *See supra* subparts IV(A)–(B).

institutions naturally tend to focus on market participants' limitations,[398] implicitly assuming that the enforcement agencies or the courts can and should respond to these limitations.[399] This is not to say, of course, that the agencies or the courts can never offer effective responses to anticompetitive, boundedly rational behavior, only that the limits of these institutions also require consideration, as already explained.[400]

Nevertheless, the current tendency of much behavioral antitrust scholarship towards supporting more assertive doctrines is also a testament to the present state of antitrust law. The courts rely on assumptions of rationality to constrain antitrust claims in many areas, including monopolization, horizontal conspiracies, vertical price restraints, and more.[401] Against this backdrop, the exposure by the behavioral approach of situations where rationality assumptions miss the mark naturally tends to challenge those rationality-based doctrines that limit the reach of the antitrust laws.

The empirical behavioral evidence, however, should also be used to caution against antitrust intervention in appropriate cases. For one, Part III already noted that closer attention to the complex relationship between antitrust institutions and rationality exposes, *inter alia*, some limits of courts and enforcement agents that might otherwise go unnoticed.[402]

Furthermore, while behavioral findings generally will not immunize specific defendants against antitrust liability, because they do not guarantee that a given firm necessarily will act in a particular way, they can be marshaled by defendants as well as by plaintiffs. This can be illustrated by the Court's ruling that Kodak could have exercised power in the aftermarket for the sale of copier parts despite competition in the primary market for copiers.[403] The behavioral approach recognizes the possibility of a sufficient proportion of boundedly rational consumers that would have provided Kodak with aftermarket power and justified the denial of summary judgment.[404] At trial, however, Kodak could have argued that its aftermarket tying was procompetitive on balance. Kodak might have

---

398. *See supra* Part III (noting also that behavioral opponents often make a mirror-image mistake: assuming that behavioral phenomena are of no antitrust concern because of the institutional corrections provided by markets and firms).

399. *See, e.g.*, Reeves & Stucke, *supra* note 2, at 1577–81 (surveying areas of antitrust policy that could be more strongly enforced using behavioral antitrust).

400. *See supra* notes 249–51 and accompanying text (briefly discussing the limits of antitrust enforcement institutions); *see also* Tor, *Entry*, *supra* note 2, at 546–47 (noting the insurmountable challenges that would face regulators who wished to identify and prevent overconfident entry).

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

402. *See* Cooper & Kovacic, *supra* note 7 (examining the implications of some potential behavioral factors for agency decision making); Tor, *Entry*, *supra* note 2, at 546–47 (illustrating the limits enforcers face when regulating negative-expected-value entry).

403. *See* Eastman Kodak Co. v. Image Technical Servs., Inc., 504 U.S. 451, 471 (1992).

404. *See supra* notes 66–74 and accompanying text (discussing the case in more detail).

needed to preserve its brand reputation with real-world, boundedly rational consumers who—unlike their hypothetical, rational counterparts—might have misattributed to the manufacturer problems with copiers that were serviced with non-Kodak parts.[405] Assuming such evidence were available, combined with the finding of a competitive primary market for copiers, a behaviorally informed evaluation of *Kodak* might have favored the defendant rather than the plaintiffs.

V. Two Essential Lessons

The fundamental methodological error often leads astray scholars who try their hands at behavioral antitrust. Proponents and opponents who mistakenly treat concrete behavioral phenomena as broad hypothetical assumptions can fail to recognize the impact of variability and heterogeneity on market behavior or the complex effects of institutions on rationality in antitrust settings. They also tend to equate bounded rationality with anticompetitiveness although that frequently is not the case. Yet the alternative—namely a more careful, empirically driven behavioral analysis of antitrust law—is fraught with significant challenges. Some of these challenges are outside the scope of the present analysis because they are common to applications of behavioral findings to the law more generally,[406] but this Part charts the essential lessons behavioral antitrust already can offer for both doctrine and enforcement policy.

A. Lesson One: The Value of Case-Specific Evidence in Antitrust Adjudication and Enforcement

Behavioral antitrust highlights the essential role of case-specific evidence in antitrust adjudication, in merger enforcement actions, and more generally in helping the courts and agencies account for boundedly rational market behavior that is systematic and predictable overall, yet variable, heterogeneous, and differently shaped by antitrust institutions in specific cases.

1. Antitrust Adjudication.—Plaintiffs should not be barred from introducing case-specific evidence in appropriate cases in areas such as conspiracy or monopolization.[407] Courts rely on the rationality assumption

405. I thank Steve Salop for making this observation at our 60th ABA Section of Antitrust Law Spring Meeting panel on behavioral antitrust.

406. *See generally* Tor, *Behavioral Methodology*, *supra* note 1, at 291–314 (discussing gaps between the focus of behavioral decision researchers and the needs of legal scholarship).

407. *See* Leslie, *supra* note 3, at 341 (highlighting the need for a more fact-specific inquiry in antitrust cases); *see also* Reeves, *supra* note 7, at 8–9 (encouraging the agencies to delve into behavioral findings before the federal courts do so); Hal Singer & Andrew Card, *Lessons from Kahneman's* Thinking, Fast and Slow*: Does Behavioral Economics Have a Role in Antitrust*

to grant summary judgment for defendants unless plaintiffs can show that the alleged conduct could be rational given market conditions. *Matsushita* summarily rejected allegations of a predatory horizontal conspiracy under Section 1 once the Court determined that a conspiracy would have required irrational behavior by the alleged conspirators.[408] Later on, *Brooke Group* applied similar reasoning to allegations of predatory pricing under Section 2, instituting the recoupment requirement.[409] More recently, the same rationale was applied in *Weyerhaeuser* to reject predatory bidding allegations.[410]

Yet the empirical evidence shows that horizontal conspiracies routinely take place where they could not have been sustained if market participants were strictly rational.[411] Behavioral and experimental findings similarly show that dominant firms or monopolists may act in a predatory manner even when doing so would appear irrational under standard assumptions.[412] Such conduct, in fact, can even be rational from the predator's perspective when actual and potential competitors know that market participants may be boundedly rational.[413]

Horizontal conspiracies that are sustained by boundedly rational behavior are no less anticompetitive than strictly rational conspiracies, however. Boundedly rational predatory pricing similarly may harm competition even while benefiting consumers in the short term by offering them lower prices.[414] But the Court's approach dismisses cases of both horizontal conspiracy and predation at the summary judgment stage, before plaintiffs have the opportunity to present actual, case-specific evidence of the alleged anticompetitive conduct.[415]

Importantly, the evidence showing that boundedly rational conduct not only exists in the market but sometimes generates anticompetitive effects does not imply that all theories of conspiracy or predation should suffice for

---

Analysis?, ANTITRUST SOURCE, Aug. 2012, art. 5 at 1, 8 (book review) (discussing the agencies' recent applications of behavioral economics to antitrust issues).

408. Matsushita Elec. Indus. Co. v. Zenith Radio Corp., 475 U.S. 574, 588–90, 595–97 (1986).

409. Brooke Grp. Ltd. v. Brown & Williamson Tobacco Corp., 509 U.S. 209, 224–25 (1993).

410. Weyerhaeuser Co. v. Ross-Simmons Hardwood Lumber Co., 549 U.S. 312, 318–20, 325–26 (2007).

411. *See supra* notes 111–13 and accompanying text.

412. *See supra* notes 128–31 and accompanying text.

413. *See supra* note 131 and accompanying text.

414. AREEDA & HOVENKAMP, *supra* note 36, ¶ 723a; Frank H. Easterbrook, *Predatory Strategies and Counterstrategies*, 48 U. CHI. L. REV. 263, 266–67 (1981) (explaining that predatory pricing benefits consumers in the short term and may be difficult to distinguish from desirable, aggressive competition).

415. Notably, by lowering the expected sanction for boundedly rational conduct, the Court's restrictive approach also makes such collusive or predatory conduct more attractive to its perpetrators compared to "rational" collusion or predation.

antitrust plaintiffs routinely to proceed beyond summary judgment. Courts and scholars have rightly emphasized the significant social costs of antitrust litigation,[416] which is not only costly but in some cases also risks chilling the same aggressive competition in the marketplace that the antitrust laws seek to foster.[417]

Instead, the behavioral lesson is that the law should ignore neither the potentially significant costs of boundedly rational and rationally anticompetitive conduct nor the costs of baseless antitrust litigation or erroneous judicial decisions. Behavioral antitrust militates for balancing those risks and costs of over- and underenforcement, sometimes referred to as type I versus type II errors in antitrust law.[418] However, unlike familiar calls for curbing antitrust complaints to avoid risking overenforcement,[419] behavioral antitrust also recognizes the risks and costs of underenforcement in some real-world markets, where such a risk would not have existed in a hypothetical world populated solely by perfectly rational market participants.

Future behavioral antitrust research will need to flesh out in greater detail and precision the balance of over- and underenforcement costs in key antitrust areas. Besides evaluating the familiar effects of the antitrust laws, such analyses will have to factor in the likelihood of both procompetitive

---

416. *See* Bell Atl. Corp. v. Twombly, 550 U.S. 544, 557–58 (2007) (citing Asahi Glass Co. v. Pentech Pharms., Inc., 289 F. Supp. 2d 986, 995 (N.D. Ill. 2003) (Posner, J., sitting by designation)) (describing the costly nature of discovery in antitrust suits as a justification for increased scrutiny at the motion to dismiss stage); Matsushita Elec. Indus. Co. v. Zenith Radio Corp., 475 U.S. 574, 594 (1986) ("[C]utting prices in order to increase business often is the very essence of competition. Thus, mistaken inferences in cases such as this one are especially costly, because they chill the very conduct the antitrust laws are designed to protect."); Frank H. Easterbrook, *The Limits of Antitrust*, 63 TEXAS L. REV. 1, 4 (1984) ("Antitrust is costly. The judges act with imperfect information about the effects of the practices at stake. The costs of action and information are the limits of antitrust.").

417. *See* Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 414 (2004) (emphasizing the slight benefit of antitrust enforcement as compared with the social cost of a "false positive[]"); *Matsushita*, 475 U.S. at 594 (stating that the mistaken enforcement of antitrust laws can reduce the very competition they are designed to foster); Easterbrook, *supra* note 416, at 4–5 (contending that strict antitrust enforcement has proven very costly by reducing the strategies available to firms).

418. *See, e.g.*, Paul L. Joskow, *Transaction Cost Economics, Antitrust Rules, and Remedies*, 18 J.L. ECON. & ORG. 95, 97–100 (2002) (discussing the balancing of type I and type II errors in antitrust analysis). *See generally* Bruce D. Abramson, *Analyzing Antitrust Analysis: The Roles of Fact and Economic Theory in Summary Judgment Adjudication*, 69 ANTITRUST L.J. 303 (2001) (drawing on this framework to consider the relationship between case-specific evidence and economic theory at the summary judgment stage).

419. *See, e.g.*, Easterbrook, *supra* note 416, at 34–39 (arguing that certain types of antitrust actions either should not be brought at all or are inherently suspect); Joshua D. Wright, *Overshot the Mark? A Simple Explanation of the Chicago School's Influence on Antitrust*, 5 COMPETITION POL'Y INT'L 1, 11–12 (2009) (arguing that error costs are an important component of the Chicago School's approach to antitrust and explaining that the overenforcement concern "begins with the presumption that the costs of false convictions in the antitrust context are likely to be significantly larger than the costs of false acquittals").

and anticompetitive conduct that behavioral forces may enable or even fuel, at least where such conduct is capable of generating substantial benefit or harm to competition.

2. Merger Enforcement Practices.—Even the antitrust agencies, which already seek and rely on case-specific evidence in enforcement actions, should reevaluate and adjust some of their merger practices in light of behavioral antitrust. Specifically, the behavioral approach at times can help identify and assess case-specific merger evidence.[420] The outcomes of merger investigations depend significantly on both the type of evidence the agencies choose to examine and their interpretation of the evidence they collect. At present, however, the process of evidence generation and interpretation is based in part—both implicitly and explicitly—on assumptions of rationality that occasionally may lead to erroneous merger enforcement outcomes.

Merger investigations frequently use a "hypothetical monopolist" test to delineate the boundaries of the relevant product market, which determine both the merging parties' market shares and other concentration measures that help predict the likely effects of the merger.[421] As the *2010 Horizontal Merger Guidelines* explain, "[m]arket definition focuses solely on demand substitution factors, i.e., on customers' ability and willingness to substitute away from one product to another in response to a price increase or a corresponding non-price change such as a reduction in product quality or service."[422] The hypothetical monopolist test focuses on demand substitution, asking how consumers of the product in question would react to a small but significant and nontransitory increase in price (or, a

420. *See, e.g.*, Bennett et al., *supra* note 7, at 124–25 (noting that while deviations from rationality might not significantly affect merger analysis, "there may still be some subtle implications of supply-side biases for mergers which merit further consideration"); Werden et al., *supra* note 7, at 137 (arguing that assessments of mergers should continue to rely solely on neoclassical economics but that these analyses might sometimes be enriched by integrating behavioral insights into demand models).

421. 2010 MERGER GUIDELINES, *supra* note 30, § 4.1. Note that although the structure of these recent guidelines and their emphasis on direct evidence of competitive effects indicates that market definition, and therefore the hypothetical monopolist test, are less central to merger investigations than they previously were, they still play an important role in practice. *See* Directorate for Fin. & Enter. Affairs Competition Comm., OECD, *Roundtable on Market Definition*, at 11–14, OECD Doc. DAF/COMP(2012)19 (Oct. 11, 2012), *available at* http://www.oecd.org/daf/competition/Marketdefinition2012.pdf (discussing the central role of market definition in competition analysis but noting that in reaction to its specific drawbacks, jurisdictions are implementing new tools and approaches to overcome its limitations in particular cases); *see also* Baker, *supra* note 186, at 129–31 (surveying the reasons for defining markets in antitrust law). For opposing viewpoints regarding the appropriate role of market definition, see generally Kaplow, *supra* note 186, criticizing, severely, the dominant market definition/market share paradigm, and Werden, An Answer to Professor Kaplow, *supra* note 187, arguing that market definition still fulfills a central role in merger enforcement.

422. 2010 MERGER GUIDELINES, *supra* note 30, § 4.

SSNIP).[423] The test "requires that a product market contain enough substitute products so that it could be subject to post-merger exercise of market power significantly exceeding that existing absent the merger."[424]

Customer surveys are one common method of data collection for purposes of predicting consumer demand substitution away from the hypothetical monopolist.[425] Yet scholars have noted that the surveys designed by the agencies may not take into account the possible impact of framing effects on consumers[426]: consumers exhibiting framing effects would react more strongly to price increases for the focal product they are surveyed about than to price reductions in substitute products, notwithstanding the analytical equivalence of the two possibilities.[427] In the presence of framing effects, therefore, the survey might overestimate consumers' willingness to switch away from the focal product, suggesting overbroad product markets that underestimate the market share and potentially the power of the focal product.[428] A behaviorally informed approach to customer surveys would consider instead the possible effects of framing on survey outcomes in order to reduce the likelihood of bias in determinations of product market boundaries.

The potential impact of new entry again illustrates how behavioral antitrust directs merger evaluation to pay particular attention to case-specific evidence. Entry plays a role in the *Guidelines*' merger analysis at two distinct stages: First, "rapid entrants"—"[f]irms that are not current producers in a relevant market, but that would very likely provide rapid supply responses with direct competitive impact in the event of a SSNIP, without incurring significant sunk costs"—are included as current market participants.[429] The behavioral approach supports the *Guidelines*' inclusion of rapid entrants in the market since firms that can enter without incurring

423. *Id.* § 4.1.1.

424. *Id.*

425. *See* Stephen Hurley, *The Use of Surveys in Merger and Competition Analysis*, 7 J. COMPETITION L. & ECON. 45 (2011) (discussing the potential and limitations of surveys in merger enforcement); Graeme Reynolds & Chris Walters, *The Use of Customer Surveys for Market Definition and the Competitive Assessment of Horizontal Mergers*, 4 J. COMPETITION L. & ECON. 411 (2008) (discussing the prevalent use of customer surveys in the U.K.'s merger enforcement process); Darren S. Tucker et al., *The Customer is Sometimes Right: The Role of Customer Views in Merger Investigations*, 3 J. COMPETITION L. & ECON. 551, 576 (2007) (arguing that customers can provide important information regarding several merger issues including, *inter alia*, demand substitution). Note that the reliance on surveys would be more likely where more direct evidence of consumer behavior, such as point-of-sale scanner data for consumer goods, is not available.

426. This is apart from other shortcomings of surveys relying on consumers' predictions of their likely reactions to hypothetical changes in the market.

427. *See* Reeves & Stucke, *supra* note 2, at 1533–35 (discussing framing as an explanation for this phenomenon).

428. *See* Baker, *supra* note 186, at 148–66 (explaining how broadly defined markets underestimate participants' market shares and vice versa when markets are defined too narrowly).

429. 2010 MERGER GUIDELINES, *supra* note 30, § 5.1.

significant sunk costs typically already are operating in related markets, and their identification necessarily will be based on case-specific evidence. Moreover, the behavioral analysis of entry also shows why such entrants tend to be more successful and to provide more effective competitive discipline of incumbents.[430]

At the second stage, on the other hand, the *Guidelines* consider potential future entry into the market as a factor that may alleviate concerns regarding the adverse competitive effects of an otherwise harmful merger.[431] In this respect, merger evaluations ask whether "entry would be timely, likely, and sufficient in its magnitude, character, and scope to deter or counteract the competitive effects of concern."[432] The behavioral approach reveals, however, that rationality-based models understate the likelihood of entry and sometimes overstate its competitive effects.[433] Start-up entry, in particular, is more likely in many markets than traditional models predict but fails at very high rates, while diversifying entry is less common but tends to fare somewhat better over time.[434] When seeking to predict the occurrence and impact of potential future entry on the adverse competitive effects of mergers, however, the *Guidelines* do not distinguish explicitly between entry per se and successful entry more specifically.[435]

In principle, the *Guidelines* leave the agencies room to focus on entry that is more likely to be effective, through the requirement that the prospective entry also be "sufficient."[436] The sufficiency requirement, however, focuses on specific circumstances that limit the efficacy of successful entry—such as greater product differentiation, entrants' limited capabilities, or reputational barriers to expansion—rather than those making entry less likely to succeed altogether.[437] Yet the *Guidelines* also mention in passing the possibility that entry will not be sufficient if it occurs at such a smaller scale, compared to the merging parties, so that the potential

430. *See* Tor, *Entry*, *supra* note 2, at 494–96 (reviewing evidence for the superior performance of diversifying entrants compared to start-ups); *id.* at 520–31 (explaining the behavioral factors that make diversifying entrants less biased regarding their entry prospects than start-ups are).

431. 2010 MERGER GUIDELINES, *supra* note 30, § 9.

432. *Id.*

433. *See* Tor, *Entry*, *supra* note 2, at 548–49.

434. *See supra* notes 166–83 and accompanying text.

435. The *Guidelines* allow for circumstances in which "[e]ntry may . . . be insufficient due to constraints that limit entrants' competitive effectiveness, such as limitations on the capabilities of the firms best placed to enter or reputational barriers to rapid expansion by new entrants." 2010 MERGER GUIDELINES, *supra* note 30, § 9.3. But this allowance applies only in limited circumstances.

436. *Id.* § 9.

437. *Id.* § 9.3.

entrants will be at a significant competitive disadvantage.[438] Without speaking directly to it, this aspect of sufficiency could help account for the systematically different impact of start-up entrants and diversifying entrants, which tend to enter at a smaller versus larger scale, respectively.[439]

The *Guidelines* also offer a potential role for case-specific evidence here, noting that "[t]he Agencies consider the actual history of entry into the relevant market and give substantial weight to this evidence."[440] The *Guidelines* even explain that the "[l]ack of successful and effective entry in the face of non-transitory increases in the margins earned on products in the relevant market tends to suggest that successful entry is slow or difficult."[441] Nonetheless, the *Guidelines* do fall short of explaining how market-specific evidence could be used to determine whether entry that is otherwise potentially timely, likely, and sufficient will also be successful and effective. This seeming shortcoming may prove problematic only in a limited number of cases, however, since the agencies in practice are quite skeptical of entry as a counterweight to the adverse competitive effects of mergers.[442]

Finally, besides pointing to the essential role of case-specific evidence in achieving more accurate market definitions or assessments of entry effects, behavioral antitrust also highlights some otherwise unrecognized difficulties in the agencies' interpretation of quantitative market data in merger analysis.[443] Some commentators argue that merger enforcement

---

438. *Id.* ("Entry by a single firm that will replicate at least the scale and strength of one of the merging firms is sufficient. Entry by one or more firms operating at a smaller scale may be sufficient if such firms are not at a significant competitive disadvantage.").

439. Tor, *Entry*, *supra* note 2, at 495–96 (noting that start-up entrants start out well behind diversifying entrants and their market share only decreases over time).

440. 2010 MERGER GUIDELINES, *supra* note 30, § 9.

441. *Id.*; *see also* U.S. DEP'T OF JUSTICE & FTC, COMMENTARY ON THE HORIZONTAL MERGER GUIDELINES 38 (2006) [hereinafter COMMENTARY], *available at* http://www.justice.gov/atr/public/guidelines/215247.htm (noting, in a published commentary on the preceding merger guidelines, that "[i]f a merger does attract entry, that entry still may be insufficient to deter or fully counteract the merger's anticompetitive effect, or the entrant may take so long to achieve market significance that the merger nevertheless produces sustained anticompetitive effects").

442. *See* Werden et al., *supra* note 7, at 130 (discussing COMMENTARY, *supra* note 441, as well as a theoretical model that suggests the limited efficacy of entry in disciplining post-merger unilateral effects).

443. In practice the more sophisticated empirical methods that are an increasingly important component of the agencies' merger review process play a much smaller role in the courts. *E.g.*, Malcolm B. Coate & Jeffrey H. Fischer, *Why Can't We All Just Get Along? Structural Modeling and Natural Experiments in Merger Analysis*, 8 EUR. COMPETITION J. 41, 53–70 (2012) (reviewing these advanced methods and noting their limited role in litigated merger cases). On the other hand, the agencies' greater acceptance of these approaches is manifested by the greater emphasis in Section 2 of the *2010 Guidelines* on direct and quantitative evidence of merger effects compared to the centrality of indirect inferences from market definition, market share, and concentration measures in the preceding *1992 Horizontal Merger Guidelines*. *Compare* 2010

need not be concerned with behavioral findings because it is based on real-world data on demand in the market.[444] This argument seems to refer in particular to simulation models that predict merger outcomes using an economic model of demand, supply, and competition in the market,[445] though it is relevant to other empirical methods that predict merger effects based in part on demand estimation.[446] As explained earlier, simulation and other structural models primarily are relevant for merger categories in which sufficient quantitative data is available—typically from point-of-sale scanners or similar sources—such as those taking place between suppliers of consumer goods.[447] Yet in addition to the limitations of this methodology that economists already have noted,[448] behavioral antitrust suggests a further need for caution in interpreting the outcomes of simulation models to guide merger evaluations. Even where a model correctly predicts price effects, for instance, it does not resolve the more fundamental challenge of systematic consumer bias. Consumer demand that is shaped in part by systematic bias, however, may not reflect consumers' true preferences accurately and therefore may offer a biased estimate of merger welfare effects.[449]

MERGER GUIDELINES, *supra* note 30, § 2, *with* U.S. DEP'T OF JUSTICE & FTC, HORIZONTAL MERGER GUIDELINES § 1 (1992), *available at* http://www.justice.gov/atr/public/guidelines/hmg.pdf.

444. *See, e.g.*, Werden et al., *supra* note 7, at 136 (arguing that "[i]n estimating the parameters of the demand system from data on actual choices, merger assessment accounts for the actual decisions made in the marketplace, normally with high-frequency aggregate data collected at the point of sale" even while conceding that behavioral deviation can complicate demand estimation); *supra* section III(A)(1) (examining this argument when considering whether and when market institutions promote demand-side rationality).

445. For a discussion of the main approaches to merger simulation and some of their limitations see, for example, Budzinski & Ruhmer, *supra* note 81; Coate & Fischer, *supra* note 443; Aviv Nevo & Michael D. Whinston, *Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference*, J. ECON. PERSP., Spring 2010, at 69; and Daniel L. Rubinfeld, *Current Issues in Antitrust Analysis*, *in* COMPETITION POLICY AND THE ECONOMIC APPROACH: FOUNDATIONS AND LIMITATIONS 81, 88–91 (Josef Drexl et al. eds., 2011).

446. DAVIS & GARCÉS, *supra* note 80, § 9.4.

447. *See* Coate & Fischer, *supra* note 443, at 61–62 (noting the limited usefulness of structural models when empirical support is lacking).

448. *See* Budzinski & Ruhmer, *supra* note 81, at 304–14 (reviewing "problems that have occurred in [merger analysis] in the literature as well as in the actual case proceedings"); Coate & Fischer, *supra* note 443, at 61–62; Nevo & Whinston, *supra* note 445, at 72–77 (analyzing the criticisms levied against merger simulations such as how to define "similar" mergers and the hazards of failing to account for changes in industry circumstances over time); Rubinfeld, *supra* note 445 (discussing trade-offs involved in several merger simulation methods); *see also* Craig Peters, *Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry*, 49 J.L. & ECON. 627, 647 (2006) (finding, in a set of airline mergers, that standard, unilateral-effects simulations did not accurately forecast price effects); sources cited *supra* note 81 (explaining the limits of some common methods of demand estimation that rely on the assumption of consumer rationality).

449. See *supra* notes 270–71 and accompanying text, which also discuss the broader challenge consumer bias poses for antitrust policy.

3. Accounting for Behavioral Irregularities in Specific Cases.— Behavioral variability, heterogeneity, and institutional effects indicate that courts and agencies also need to avoid the fundamental methodological error when evaluating allegations of anticompetitive effects in specific cases.[450] Specifically, courts and agencies cannot automatically assume—without adducing evidence for the conduct or effects that constitute the relevant violation of the antitrust laws—that market participants will exhibit any particular deviation from rationality in a given instance.[451] They cannot assume defendants' conduct had an anticompetitive effect for behavioral reasons, nor can they assume that bounded rationality prevented competitive harm, without accounting for behavioral irregularities.

Unlike those scholars who fall prey to the fundamental methodological error, however, attention to behavioral irregularities should come naturally to courts and agencies that primarily evaluate specific instances of potentially anticompetitive behavior.[452] After all, when evaluating the merits of antitrust cases, these decision makers must determine whether the particular conduct of a particular defendant violated the antitrust laws, not how market participants generally behave.[453]

The risk of failing to account for behavioral irregularities is small where the antitrust laws do not require an evaluation of the competitive effects of the alleged conduct. Criminal prosecutions of cartels are a primary example here since these horizontal restraints are per se illegal.[454] When courts face evidence of cartelization they need not examine whether market behavior and outcomes comport with rationality or bounded rationality.[455]

In most areas of antitrust law, however, defendants' conduct is judged under a rule of reason, which requires the courts, in principle, to evaluate its

---

450. *See supra* subpart I(D).

451. *See* Christopher R. Leslie, *Can Antitrust Law Incorporate Insights from Behavioral Economics?*, 92 TEXAS L. REV. SEE ALSO (forthcoming 2014) (manuscript at 7) (on file with the Texas Law Review) ("Theory deals in aggregates; litigation deals with individual episodes of anticompetitive behavior.").

452. Insofar as the Court is tasked with formulating broader antitrust doctrines and the agencies need to offer guidance, both sets of institutions must consider the effects of behavioral regularities as well. *See infra* subpart V(B).

453. Leslie, *supra* note 451 (manuscript at 9–10).

454. *See, e.g.*, Texaco Inc. v. Dagher, 547 U.S. 1, 5 (2006) (noting that "horizontal price-fixing agreements[] fall into the category of arrangements that are *per se* unlawful"); *see also* Verizon Commc'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 408 (2004) (referring to collusion as the "supreme evil of antitrust"); Palmer v. BRG of Ga., Inc., 498 U.S. 46, 49–50 (1990) (per curiam) (applying per se illegality to the division of markets between competitors).

455. *See* 7 PHILLIP E. AREEDA & HERBERT HOVENKAMP, ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION ¶ 1500 (3d ed. 2010) (noting that price-fixing is per se unlawful).

competitive effects.[456] Yet in practice, the dramatic majority of rule of reason cases are disposed of due to the plaintiff's failure to make the case that the defendant's conduct generated the requisite anticompetitive effects.[457] Since the rule of reason requires courts to consider anticompetitive effects, however, they should be able to avoid the fundamental methodological error.

The Court's recent adoption of a rule of reason approach to RPM[458] is a case in point. The behavioral approach revealed that manufacturers excessively rely on RPM to control the resale prices of their products.[459] Yet further analysis supported *Leegin*'s discarding of the long-standing but outdated per se rule and offered behaviorally informed foundations for a structured rule of reason in this area.[460] Behavioral antitrust did not suggest, however, that courts rely on the behavioral evidence to assume that RPM is always anticompetitive, or even just inefficient—a move that would have led courts falsely to declare the practice illegal in some cases where its effects are benign or even procompetitive.[461] Instead, the behavioral approach sought to account for behavioral irregularities, fashioning a structured rule for markets inhabited by rationally anticompetitive and procompetitive, as well as by boundedly rational, inefficient, and sometimes anticompetitive, instances of RPM.[462] Under this approach, courts would seek case-specific evidence that sheds light on the nature of defendants' RPM and its competitive effects, assigning liability only to cases in which the practice—rational or boundedly rational—is anticompetitive.[463]

Monopolization by predatory pricing offers another familiar example of how courts can account for behavioral irregularities. Behavioral findings

---

456. *See, e.g.*, *Dagher*, 547 U.S. at 5 (applying the rule of reason presumptively). The various rule of reason approaches still require the courts to determine the effects of the conduct in question. *See generally* Andrew I. Gavil, *Moving Beyond Caricature and Characterization: The Modern Rule of Reason in Practice*, 85 S. CAL. L. REV. 733 (2012) (providing a survey of the current rule of reason doctrine).

457. *See* Michael A. Carrier, *The Real Rule of Reason: Bridging the Disconnect*, 1999 BYU L. REV. 1265, 1267–68 [hereinafter Carrier, *The Real Rule of Reason*] (finding that in 96% of rule of reason cases courts do not conduct the balancing test contemplated by the rule); Michael A. Carrier, *The Rule of Reason: An Empirical Update for the 21st Century*, 16 GEO. MASON L. REV. 827, 828 (2009) [hereinafter Carrier, *An Empirical Update*] (finding that 97% of reported cases are disposed of based on the plaintiff's failure to prove anticompetitive effects).

458. Leegin Creative Leather Prods., Inc. v. PSKS, Inc., 551 U.S. 877, 882 (2007).

459. Tor & Rinner, *supra* note 7, at 834.

460. *Id.* at 854–55.

461. *See id.* at 855 ("At times, [RPM] generates significant consumer harm, while in other instances RPM may be used excessively but cause only limited consumer harm. Yet on other occasions, the practice reflects those rational pro- or anticompetitive calculations assumed by its past analyses.").

462. *See id.* at 858–64 (proposing a new rule of reason analysis that takes into account the behavioral evidence).

463. *Id.*

suggest that some monopolists may engage in predation that would not have been rational in a world populated only by strictly rational actors.[464] But courts must account for behavioral irregularity—namely, the possibility that some monopolists engage, successfully or unsuccessfully, in boundedly rational predation while at other times monopolists act rationally. Hence courts should neither assume that predatory pricing allegations are never credible when recoupment would have been unlikely in a strictly rational world nor that such predation is always credible or likely, taking care instead to evaluate the evidence in the specific case.[465]

In the merger area, due to the parties' strong incentives to overstate future efficiencies, as well as their tendency otherwise to overestimate them, and given the evidence of the high frequency of unsuccessful mergers, the agencies should not simply assume that proffered merger efficiencies will materialize in fact.[466] At the same time, however, because some mergers do generate meaningful efficiencies, a blanket assumption that the parties' arguments are overstated would be unwarranted as well. Arguably, the agencies' traditional skepticism towards efficiency claims is appropriate, though currently it is based on considerations other than the contribution of behavioral forces to biased predictions.[467] To determine how behavioral factors should be accounted for when they evaluate efficiency claims, the agencies would need further to identify specific market and firm factors that facilitate or inhibit biases in managerial predictions of future efficiencies.

Similarly, some empirical evidence indicates that, in contrast to the standard, rationality-based assumption, fixed and sunk costs may affect firms' pricing decisions.[468] Firms that take into account fixed costs, however, are more likely than traditional models predict to reduce product prices following a merger that achieves such cost reductions.[469] Some commentators point to this evidence as favoring a less critical approach

---

464. *See* Leslie, *supra* note 3, at 284–85, 319–24 (attacking the use of the rationality assumption in predatory pricing cases). *See generally* Tor, *Predatory Pricing*, *supra* note 2 (arguing that more predatory pricing occurs than neoclassical economics would predict).

465. Further research could be helpful in identifying circumstances that make boundedly rational predation more or less likely and for fashioning appropriate hurdles for plaintiffs to overcome to reach trial. *See infra* subpart V(B).

466. *See, e.g.*, Stucke, *Reconsidering*, *supra* note 7, at 155 & n.210 (discussing overconfidence in the merger context and noting the high frequency of failed mergers).

467. *See, e.g.*, Reeves & Stucke, *supra* note 2, at 1560–63 (challenging the assumption that mergers are consummated to generate significant efficiencies); Werden et al., *supra* note 7, at 130 (noting that agency skepticism is consistent with analysis showing that mergers "rarely create a significant profit incentive for entry").

468. Bennett et al., *supra* note 7, at 124–25; Werden et al., *supra* note 7, at 130–31.

469. *Cf.* Bennett et al., *supra* note 7, at 125 (questioning the assumption that sunk costs would not affect pricing decisions in mergers).

towards claims of merger efficiencies.[470] Yet this argument again manifests the fundamental attribution error: some merging parties will indeed take into account fixed-cost reductions in their post-merger prices, reducing some of the merger's anticompetitive effects, while other firms will ignore these costs in the short term, as traditional models assume. Consequently, in this case as well, a blanket approach that always assumes or ignores fixed-cost effects is likely to miss the mark. Instead, case-specific evidence about the parties' track record with respect to fixed-cost changes, customer expectations, and so on would be more informative for agency predictions of whether such cost savings will be passed on.

B. Lesson Two: Accounting for Behavioral Regularities in Antitrust Law

Beyond highlighting the need to account for behavioral irregularities in antitrust adjudication and merger-enforcement practices, the behavioral approach also offers important lessons for the design of the antitrust laws. For some commentators, the complex reality of market behavior seems sufficiently overwhelming to justify a conscious reliance on radically simplified assumptions of rationality throughout antitrust law without exception.[471] But willful ignorance of systematic, predictable deviations from strict rationality can produce antitrust doctrines and policies that harm, rather than benefit, competition. At the same time, care also must be taken to avoid the fundamental methodological error when formulating behaviorally informed antitrust doctrine. Instead, the regularities identified by the behavioral approach can advance antitrust law in a number of distinct ways: First, behavioral regularities can help guide the choice among competing antitrust rules in different areas.[472] Second, a better understanding of such regularities can assist in structuring these rules to promote rather than retard competition. And, finally, behavioral regularities sometimes may be carefully and usefully generalized as stylized observations, which economic models can incorporate when seeking to predict and explain market outcomes.

For those manifesting the fundamental methodological error, behavioral antitrust may appear to provide clear, general policy implications. After all, if one assumes that all market participants—or at least all actors of a given class of participants—are always and equally biased, the necessary modifications of antitrust doctrine seem

---

470. *See, e.g.*, Werden et al., *supra* note 7, at 130–31 (citing commentators who suggest that merger efficiencies should be given more weight in light of behavioral-research findings).

471. *See, e.g.*, Ginsburg & Moore, *supra* note 7, at 96–98 (concluding that judges are unlikely to rely on behavioral research because "the central theme of the discipline[,] . . . rather than foreclosing possibilities, opens them up and thereby increases the degrees of freedom with which a court may pursue personal, idiosyncratic goals").

472. Importantly, the rules advocated by the behavioral approach can be either simple or complex, irrespective of the complexity of the market behavior they address.

straightforward. For example, if one were to assume that manufacturers always impose RPM excessively on their retailers when the practice is legal, the law would be justified in reverting to the rule of per se illegality that *Leegin* discarded.[473] Universally excessive RPM would be either anticompetitive—for traditional or behavioral reasons—or a competitively neutral yet socially costly, systematic mistake on the part of manufacturers that harms them and some of their retailers. If this were the case, per se illegality would have benefited competition, risking no chilling of efficient vertical arrangements—which are absent in this case by assumption—and providing the great benefits and cost savings to business and the legal system that flow from clear and simple, bright-line antitrust rules.[474]

In reality, of course, the behavioral evidence on RPM revealed only a strong tendency of some manufacturers to employ this practice excessively.[475] Yet excessive, boundedly rational uses of this vertical arrangement can coexist with rational, anticompetitive ones as well as with beneficial, procompetitive arrangements that promote the provision of output-increasing dealer services.[476] Once the heterogeneity and variability of market behavior are taken into account, behavioral regularities advocate for a different rule from the one supported by assumed universal bias, justifying a rule of reason approach to RPM.[477] Per se illegality, which initially appeared attractive, turns out to be inappropriate in the face of a behavioral regularity that falls short of universality.[478]

Besides tipping the scales in favor of one candidate rule over another, the behavioral approach can assist in a more nuanced structuring of the chosen rule. This contribution is particularly important since current-day antitrust applies a rule of reason analysis in areas ranging from many horizontal restraints through nearly all vertical ones, to monopolization, and

---

473. *See supra* notes 145, 229 and accompanying text (describing the tendency of *some* manufacturers to excessively impose RPM).

474. *See generally* Atl. Richfield Co. v. USA Petroleum Co., 495 U.S. 328, 342 (1990) ("The *per se* rule is a presumption of unreasonableness based on 'business certainty and litigation efficiency.'" (quoting Arizona v. Maricopa Cnty. Med. Soc'y, 457 U.S. 332, 344 (1982))); *Maricopa Cnty.*, 457 U.S. at 343–44 ("The costs of judging business practices under the rule of reason, however, have been reduced by the recognition of *per se* rules."); AREEDA & HOVENKAMP, *supra* note 455, ¶ 1510a ("The root meaning of 'per se illegality' is that courts refuse to consider one or more factors that would ordinarily bear on the reasonableness of challenged conduct."); Richard A. Posner, *The Next Step in the Antitrust Treatment of Restricted Distribution: Per Se Legality*, 48 U. CHI. L. REV. 6 (1981) (advocating for the benefits of per se legality in vertical restraints).

475. *See* Tor & Rinner, *supra* note 7, at 839–47 (finding that some manufacturers will overuse RPM for extensive periods of time when there are no legal constraints).

476. *See id.* at 859–62 (arguing that behavioral antitrust supports the view that neither the proplaintiff nor the prodefendant visions of RPM are entirely correct).

477. *See id.* at 860–62 (arguing for an improved rule of reason approach to RPM).

478. The behavioral account also makes clear that the per se legality that some advocate for RPM is equally inappropriate. *See generally* Posner, *supra* note 476 (arguing for per se legality of vertical restraints).

more.[479] In each of these areas, antitrust courts are required to determine whether a particular form of market conduct is on balance pro- or anticompetitive in every case.

Because open-ended inquiries under the rule of reason are notoriously difficult to implement, courts have long sought to structure them. Frequently they require plaintiffs to make some initial showing—most notably a clearly defined market where the allegedly anticompetitive conduct took place—without which the case will not proceed beyond summary judgment.[480] In other cases, courts have established elaborate structures that require plaintiffs and defendants in turn to bear the burden of proving different elements of the case.[481] The specific structure of the rule of reason significantly impacts antitrust plaintiffs' likelihood of success. Private plaintiffs routinely flounder, for instance, when required by courts to offer a market definition before proceeding with other evidence of anticompetitive effects, as the case has been with allegations of vertical nonprice restraints under Section 1 of the Sherman Act since *Sylvania*.[482]

Behavioral antitrust sometimes can guide the all-important structuring of the rule of reason to insure that plaintiffs' antitrust actions will face appropriate hurdles—neither insufficient nor excessive—based on a better understanding of market behavior. To continue with the RPM illustration,

---

479. *See, e.g.*, Leegin Creative Leather Prods., Inc. v. PSKS, Inc., 551 U.S. 877, 899, 907 (2007) (overruling Dr. Miles Med. Co. v. John D. Park & Sons Co., 220 U.S. 373 (1911) and holding that minimum RPM should be evaluated under the rule of reason); State Oil Co. v. Khan, 522 U.S. 3, 22 (1997) (overruling Albrecht v. Herald Co., 390 U.S. 145 (1968), and holding that maximum RPM should be evaluated under the rule of reason); Cont'l T.V., Inc. v. GTE Sylvania Inc., 433 U.S. 36, 58–59 (1977) (overruling United States v. Arnold, Schwinn & Co., 388 U.S. 365 (1967), and holding that vertical nonprice restraints should be evaluated under the rule of reason); Standard Oil Co. of N.J. v. United States, 221 U.S. 1, 60–66 (1911) (adopting and announcing the rule of reason).

480. *See Leegin*, 551 U.S. at 885–86 ("Whether the businesses involved have market power is a further, significant consideration."); Spectrum Sports, Inc. v. McQuillan, 506 U.S. 447, 459 (1993) (holding that a plaintiff must prove a dangerous probability that the defendant would monopolize a particular market to prevail in a Section 2 attempted monopolization claim). The Court even uses market power as a screening mechanism in cases judged under a per se rule. *See* Jefferson Parish Hosp. Dist. No. 2 v. Hyde, 466 U.S. 2, 13–14 (1984) ("[W]e have condemned tying arrangements when the seller has some special ability—usually called 'market power'—to force a purchaser to do something that he would not do in a competitive market.").

481. *See, e.g.*, Polygram Holding, Inc. v. FTC, 416 F.3d 29, 35–39 (D.C. Cir. 2005) (upholding the FTC's order in the "Three Tenors" case and applying a burden-shifting framework to a horizontal-restraint claim); United States v. Microsoft Corp., 253 F.3d 34, 58–59 (D.C. Cir. 2001) (applying a burden-shifting structure in a rule of reason inquiry for alleged monopolization under Section 2 of the Sherman Act).

482. Cont'l T.V., Inc. v. GTE Sylvania Inc., 433 U.S. 36 (1977); *see also* Carrier, *An Empirical Update*, *supra* note 457, at 834–36 (describing the challenges of burden-shifting tests applied by each circuit); Carrier, *The Real Rule of Reason*, *supra* note 457, at 1268 (summarizing the burden-shifting exercise and its demands on plaintiffs); Douglas H. Ginsburg, *Vertical Restraints: De Facto Legality Under the Rule of Reason*, 60 ANTITRUST L.J. 67, 71 (1991) (showing very low percentages of success for plaintiffs under the *Sylvania* rule).

the *Leegin* Court explicitly acknowledged the need for structuring its new rule of reason for the practice, suggesting that lower courts may "devise rules over time for offering proof, or even presumptions where justified, to make the rule of reason a fair and efficient way to prohibit anticompetitive restraints and to promote procompetitive ones."[483] Following *Leegin*'s call, and occasionally even prior to it, courts, enforcement agencies, and scholars offered a variety of structured rule of reason approaches to RPM.[484] Notwithstanding their many fundamental differences, these approaches all sought to distinguish rationally procompetitive instances of the practice from its rationally anticompetitive uses.[485] None of the proposed structures, however, accounted for the behavioral regularity of boundedly rational manufacturers excessively employing RPM that is neither pro- nor anticompetitive, as traditionally understood.[486]

In contrast, when structuring the rule of reason, behavioral antitrust takes into account the incidence and market effects of boundedly rational RPM as well as the possibility of rationally pro- and anticompetitive uses of the practice. These factors suggest that plaintiffs should be required first to make their prima facie case in one of two ways: They could directly show that the defendant's output decreased following the employment of RPM, which indicates that the practice was anticompetitive or boundedly rational and excessive.[487] Alternatively, they could offer indirect evidence of the dangers of the practice in the specific case by establishing the presence of factors like those cited by the *Leegin* Court, especially the prevalence of RPM in the industry, concentration, or market power at either manufacture or retail.[488] Thus, where a plaintiff makes either prima facie case:

> [T]he defendant should be allowed to rebut . . . . show[ing] not only that the practice sought to address a real business problem—such as free riding—but also that the problem generated measurable harm to the manufacturer. Defendants would also have to show, moreover, that less restrictive means for addressing this problem were significantly more costly or less effective. Otherwise . . . manufacturers could routinely proclaim the various theoretical harms of price-cutting . . . without [offering specific evidence].[489]

Last, beyond assisting in choosing and fashioning more effective antitrust rules, behavioral regularities can be used as stylized facts, allowing economists to develop more accurate models and predictions of market

---

483. 551 U.S. at 898–99.

484. *See* Tor & Rinner, *supra* note 7, at 858 & nn.309–11, 859 & n.312 (providing examples for rule of reason structures in judicial and agency decisions as well in academic commentary).

485. *See id.* at 859.

486. *See id.*

487. *Id.* at 862.

488. *Id.*

489. *Id.* (footnote omitted).

behavior.[490] Although such models must ignore some behavioral irregularities and should therefore be used with care, they can complement more nuanced analyses. The potential contribution of such models is illustrated by the extensive research already employing them in other fields that study market behavior, like behavioral finance. Economists drawing on behaviorally informed models have offered both theoretical and empirical insights into firms' dividend policies, IPOs, mergers and acquisitions activity, and financing and investment decisions, to name but a few key areas.[491] In antitrust, empirical studies similarly could compare the predictive power of behaviorally informed models to that of traditional, rationality-based ones. In cases where they are shown to offer better predictions, specific stylized behavioral models also could be used by the agencies and the parties before them when assessing the competitive effects of mergers as well. Although promising within their inherent constraints, however, formal models based on stylized behavioral facts are only beginning to develop in industrial organization—the economic field most directly relevant for antitrust—and to date this new literature primarily focuses on the reactions of rational firms to consumer limitations and bias.[492]

Conclusion

This Article finds that behavioral antitrust can advance the law by offering a better understanding of the behavior of antitrust actors, though the approach still is nascent. The recent outpour of interest from the antitrust community—with scholars, practitioners, enforcement officials, and judges joining the behavioral antitrust fray—is a clear indication of both the significance of this new approach and the concerns over its future impact on the field. The Article showed that the behavioral approach is

490. *Cf.* Jolls et al., *supra* note 1, at 1475 (making a similar point with respect to behavioral law and economics more generally).

491. *See, e.g.*, BEHAVIORAL FINANCE, *supra* note 189 (offering a series of reviews in these and other areas of behavioral finance that discuss both theoretical and empirical findings in the field, many of which rely on stylized models); *see also* Malcolm Baker & Jeffrey Wurgler, *Behavioral Corporate Finance: An Updated Survey*, *in* 2A HANDBOOK OF THE ECONOMICS OF FINANCE: CORPORATE FINANCE 357 (George M. Constantinides et al. eds., 2013) (providing a more recent general survey of the field).

492. *See, e.g.*, STEFFEN HUCK ET AL., OFFICE OF FAIR TRADING, CONSUMER BEHAVIORAL BIASES IN COMPETITION: A SURVEY (2011) (U.K.), *available at* http://www.oft.gov.uk/shared_oft/research/OFT1324.pdf (reviewing both empirical and theoretical literature on behavioral economics and examining the influence of consumer behavioral biases on firm decisions); RAN SPIEGLER, BOUNDED RATIONALITY AND INDUSTRIAL ORGANIZATION (2011) (summarizing and synthesizing recent theoretical developments in models that incorporate some behaviorally informed stylized facts with respect to consumer behavior and rational firms' strategic responses to it); Armstrong & Huck, *supra* note 7 (discussing some research of this nature); Ellison, *supra* note 95 (reviewing much of the new industrial organizational literature that draws on stylized behaviorally informed models).

poised to advance antitrust law and policy in myriad ways but can and should only supplement, not substitute for, the apparatus of standard antitrust law and economics.

As behavioral antitrust continues to develop, proponents and critics alike must beware the common pitfalls associated with the fundamental methodological error. Without a better appreciation of the empirically driven nature of the behavioral approach, analysts will continue to confuse concrete behavioral phenomena with broad hypothetical assumptions, repeatedly making the three classes of mistakes examined in this Article. Some will fail to appreciate that human judgment and decision behavior is variable and heterogeneous, neither constant nor uniform. Others will ignore the various ways in which antitrust institutions facilitate behavior that resembles the assumptions of rationality to greater or lesser degrees. And many commentators will continue to conflate bounded rationality with an automatic license for more assertive antitrust policies.

Scholars should recognize instead that both the unique contribution of behavioral antitrust and its inherent limitations manifest the fundamentally empirical character of this approach. A greater attention to the specific contours of the behavioral evidence generally and its likely manifestation in antitrust settings in particular, combined with a continued effort to generate further antitrust-relevant findings, will go a long way towards helping the antitrust community overcome the fundamental methodological error in behavioral antitrust.

Finally, beyond providing a better understanding of the nature of behavioral antitrust, this Article discussed two essential sets of lessons this new approach offers for doctrine and policy even now. One concerned the important function of case-specific evidence in both antitrust adjudication in the courts and agency enforcement actions. The other showed how antitrust doctrine can incorporate the evidence of systematic and predictable behavioral regularities in the market that still fall short of universal propositions. If followed, these essential lessons already can promote significantly the antitrust laws' mission of protecting competition.

* * *

Book Reviews

From Interraciality to Racial Realism

ACCORDING TO OUR HEARTS: *RHINELANDER V. RHINELANDER* AND THE LAW OF THE MULTIRACIAL FAMILY. By Angela Onwuachi-Willig. New Haven, Connecticut: Yale University Press, 2013. 344 pages. $38.00.

Reviewed by L. Song Richardson* & Phillip Atiba Goff**

Introduction

As a result of our nation's progress in breaking down de jure and de facto discrimination, there is significant debate about how to advocate for racial justice. Since explicit prejudice is in such stark decline,[1] some argue that the fight for racial equality is almost complete. All that remains is for the achievements of stigmatized racial groups to catch up to our nation's enlightened hearts and minds.

However, as Professor Angela Onwuachi-Willig demonstrates in her important and groundbreaking new book, *According to Our Hearts:* Rhinelander v. Rhinelander *and the Law of the Multiracial Family*, it is a mistake to believe that the fight for racial justice is over. Through her close

* Professor, University of Iowa College of Law. J.D., Yale Law School; B.A., Harvard University. I wish to thank Lauren Sudeall Lucas for her insightful comments on an earlier draft and to extend my gratitude to Elizabeth A. Etchells and Franccesca Kazerooni for their invaluable research assistance. Finally, I appreciate the valuable work performed by Kelsie Krueger, Michael Deane, Spencer Patton, and the Texas Law Review staff.

** Assistant Professor, UCLA, Department of Psychology. Ph.D., Stanford University; B.A. Harvard University.

1. John F. Dovidio, *On the Nature of Contemporary Prejudice: The Third Wave*, 57 J. SOC. ISSUES 829, 833–34 (2001).

examination of relationships between black and white individuals, she exposes how the remnants of our nation's sordid history of racial bias continue to affect both law and culture today. More importantly, what Onwuachi-Willig conveys is the need to challenge dominant assumptions regarding how racism operates in order to improve the laws and customs that enable equitable outcomes. Specifically, she demonstrates that although de jure prohibitions on interracial marriages are unconstitutional after the Supreme Court's decision in *Loving v. Virginia*[2] and societal approval towards these couplings is higher than ever,[3] the dearth of mixed-race relationships and families, especially black–white couples,[4] belies the claim that our society has overcome the deeply engrained taboo.

Despite progress, outright racism against mixed-race couples is troublingly evident, as the violent discourse in response to a recent Cheerios commercial[5] and the violent assaults and murders perpetrated against such couples attest.[6] However, Onwuachi-Willig's chief accomplishment in this book is not simply pointing out where racism still exists but rather in revealing why our thinking about race must evolve in order to address it. As Onwuachi-Willig masterfully demonstrates, unconscious and invisible biases are equally, if not more, pernicious than outright bigotry because their invisibility in both law and society makes them difficult to recognize, confront, and address.[7] In fact, the law often provides no remedy for even the most obvious instances of discrimination against mixed-race couples.[8]

According to Our Hearts is on the pulse of the central paradox facing the fight for civil rights today, namely, how to reconcile the broad gentling of racial animus with the continued persistence of racial inequalities of all types. Onwuachi-Willig gives us some insight into this puzzle as it relates to multiracial families. Through her exploration of how law and society together create and maintain the normative ideal of family as monoracial and heterosexual she "challenge[s] the commonly accepted notion that legal

---

2. 388 U.S. 1, 12 (1967).

3. ANGELA ONWUACHI-WILLIG, ACCORDING TO OUR HEARTS: *RHINELANDER V. RHINELANDER* AND THE LAW OF THE MULTIRACIAL FAMILY 166 (2013).

4. *Id.* at 124–26, 128–31. Professor Onwuachi-Willig primarily focuses her attention on heterosexual relationships.

5. Colleen Curry, *Interracial Family in Cheerios Ad Gets Hate Comments*, ABC NEWS BLOGS (May 31, 2013, 1:08 PM), http://abcnews.go.com/blogs/headlines/2013/05/interracial-family-in-cheerios-ad-gets-hate-comments.

6. *See, e.g.*, Smokey D. Fontaine, *Interracial Couple Murder Was a Hate Crime*, NEWSONE FOR BLACK AM. (Nov. 7, 2008), http://www.newsone.com/33081/interracial-couple-murder-was-a-hate-crime (detailing a crime committed against interracial couple); Sasha King, *3 Men Charged with Hate Crime After Beating Up Interracial Couple*, BLACK ENTERPRISE (Feb. 4, 2013), http://www.blackenterprise.com/news/interracial-hate-crime-men-charged (same).

7. *See* ONWUACHI-WILLIG, *supra* note 3, at 134–36, 154–55 (discussing and critiquing the assumption that attraction is personal and not affected by societal norms).

8. *See, e.g.*, *id.* at 207–08, 262–63 (describing courts' rejection of discrimination claims premised on interraciality).

discouragement of and punishment for intimate, cross-racial heterosexual intimacy no longer exists."[9] After persuasively demonstrating that these sanctions survive today, she introduces the concept of "interraciality." This term refers to the discrimination biracial, heterosexual couples face "because of their interraciality as a couple, as opposed to the race of just one member of the couple."[10] In order to facilitate recognition of and recompense for the harm multiracial families continue to endure, Onwuachi-Willig argues that discrimination based on interraciality should be included as a protected category in current antidiscrimination law.[11]

In this Review, we argue that her concept of interraciality is not only an important intervention for a group currently left unprotected by antidiscrimination law but that it is also important within the broader context of civil rights. Her insights not only explain why adding protected *categories* to antidiscrimination law is necessary, but—when combined with lessons from the social psychology of contemporary bias—they show why the law should also account for the additional *mechanisms* through which racial bias manifests itself, whether conscious or unconscious, visible or invisible.

In Part I, we briefly summarize the book, emphasizing Professor Onwuachi-Willig's discussion of how societal attitudes influence the law's construction of the family as monoracial and heterosexual and the resulting harms to multiracial families. In Part II, we situate her book within the broader cause of civil rights and explain how her concept of interraciality pushes the boundaries of the law's traditional framing of bias as primarily motivated by intentional racism. We briefly introduce the social science of contemporary bias to demonstrate that not only is the law constructed around an outdated notion of family but also that the law's current conception of racial bias is similarly constructed to make invisible all but the most obvious and purposeful forms of discrimination. We conclude by arguing that only by enlarging our discussion of the nature and causes of bias, as Onwuachi-Willig does so powerfully in her book, can law and society address "the full range of [discrimination] harms that can occur."[12]

I. Interraciality: Protecting an Invisible Class

While today no laws prohibiting interracial marriages exist and attitudes towards these relationships are generally positive, our society is still dealing with the effects of centuries of legal and de facto discrimination against these couples. This is reflected not only in the relative rarity of

9. *Id.* at 201.
10. *Id.* at 21 (footnote omitted).
11. *Id.* at 264–66.
12. *Id.* at 212.

black–white intimacy[13] but also in the invisible, but harmful, ways that law and society continue to treat these relationships as nonnormative. Professor Onwuachi-Willig's theory of interraciality accounts for this legacy of history by suggesting that antidiscrimination laws should explicitly recognize multiracial families as a new protected class. Only by doing so, she argues, can the law, and society in general, begin to ameliorate the often invisible punishments these couples and families confront. This Part traces Onwuachi-Willig's illuminating discussion of the historical and contemporary punishments of heterosexual, mixed-race couples and unpacks her theory of interraciality.

A. The Rhinelanders' Doomed Relationship

Onwuachi-Willig uses the ill-fated relationship of Alice and Leonard Rhinelander in 1920s New York to expose how the legal and social taboo against interracial relationships is not simply a relic of the past but continues to affect multiracial relationships today. In part I of the book, she examines their relationship from their initial meeting in 1921, through their two-month marriage, to the end of the annulment trial in which Leonard lost his claim of marital fraud.[14] His allegation was that Alice had deceived him about her racial identity, and had he known she was not white, he would not have married her.[15] Through her analysis, Onwuachi-Willig reveals "how law and society have often functioned together to frame the normative ideal of family as monoracial, both in our history and in our present."[16]

The Rhinelander story reflects the symbiotic relationship between societal norms and the law. Despite the absence of de jure prohibitions on interracial marriages in 1920s New York,[17] racism made the crossing of racial boundaries in intimate relationships unacceptable. The depth of the then-existing racial intolerance is made starkly apparent by the fact that neither Alice's lawyers nor the judge questioned whether racial misrepresentation was a cognizable claim under the marriage fraud statute.[18] It was so obvious that unknowingly marrying a black person was harmful and degrading to the white spouse that neither the judge nor Alice's lawyer challenged the underlying racist assumption embedded in the very nature of the lawsuit itself. The language of the statute did not require this interpretation, but prevailing attitudes about black inferiority did.

Furthermore, existing beliefs about "racial purity" also foreclosed Alice's ability to defend against the lawsuit by claiming a white racial

13. *Id.* at 124–26.
14. *Id.* at 19–20.
15. *Id.* at 33.
16. *Id.* at 121.
17. *Id.* at 36.
18. *Id.*

identity. Although Alice's mother was white and the rule of hypodescent did not officially apply in New York, it was also apparent that just one drop of "black" blood was sufficient to exclude her from being classified as white.[19] Finally, Alice's lawyer took advantage of and reinforced the predominant view of race as biological. He argued that no fraud occurred because Leonard had seen Alice naked prior to their marriage and thus had to know that she was black.[20] To prove his point, in a shocking and degrading move that would have been unthinkable had Alice been considered white, her lawyer had Alice bare her upper and lower body in front of the all-male jury.[21] This display was necessary, in his view, to demonstrate that she was "'dark' all over."[22]

Although racial attitudes have become more outwardly tolerant since the Rhinelander trial in 1924, in the remainder of the book, Professor Onwuachi-Willig reminds the reader that it is not so easy to escape the legacy of the racist history reflected in the Rhinelander case. She masterfully weaves the story throughout, persuasively demonstrating that its lessons still have relevance today. Next, subpart I(B) shows that despite the gentling of racial animus, multiracial families are still punished, both explicitly and implicitly, for crossing the color line.

B. Current Manifestations

In part II of the book, Onwuachi-Willig challenges the belief that "legal discouragement of and punishment for . . . [interracial] intimacy no longer exists."[23] She demonstrates that while the law does not explicitly discriminate against these relationships, its implicit assumptions about the monoraciality of families are equally pernicious. That is because, by framing family as monoracial, antidiscrimination law fails to recognize multiracial families and the unique harms they suffer as a result of their interraciality.[24]

One example she uses to illustrate her point comes from how the Fair Housing Act defines familial status.[25] This statute provides a compelling illustration because, unlike the other civil rights statutes, its goal is to safeguard families, instead of just individuals, from unlawful discrimi-

---

19. *Id.* For a general discussion of hypodescent, see Lauren Sudeall Lucas, *Undoing Race? Reconciling Multiracial Identity with the Pursuit of Racial Equality*, 102 CALIF. L. REV. (forthcoming 2014) (manuscript at 6–7) (on file with authors).

20. ONWUACHI-WILLIG, *supra* note 3, at 77.

21. *Id.* at 77–79.

22. *Id.* at 79 (quoting RENEE C. ROMANO, RACE MIXING: BLACK-WHITE MARRIAGE IN POSTWAR AMERICA 167 (2003)).

23. *Id.* at 201.

24. *See id.* at 20.

25. *See* 42 U.S.C. § 3602(k) (2006) (defining "familial status" only in terms of one's relation to his or her dependents, not on the relationship between spouses).

nation.[26] However, since Congress assumed monoraciality when drafting the statute, its text inadvertently excludes multiracial families from protection. Thus, if a landlord unlawfully refuses to rent to a monoracial black family, the family unit can file a suit together, alleging discrimination against them as a family.[27] However, the statute does not explicitly cover situations where "the couple is being discriminated against, not because of the race of a specific member or specific members of the family—that is, the black person in the family—but rather because of the family's status as a multiracial family, as a unit that was formed through race mixing."[28] Multiracial families are not a protected category under the statute,[29] and the fact that these families were overlooked in the statute's drafting demonstrates just how deeply embedded the norm of the monoracial family is.

A few courts have attempted to work around the invisibility of multiracial families in the law by using a "discrimination by association" analysis.[30] When seeking relief under the Fair Housing Act, for instance, this analysis requires each member of a multiracial family to claim "separately and individually that, but for the race of their spouse, they each would have been treated differently."[31] Unlike monoracial couples, who can file their claims together, the discrimination by association analysis requires multiracial families to "engage in wordplay" and "divide up their family unit" in order to bring their suit.[32]

This construction of family as monoracial does not just exist in antidiscrimination law; it exists in society as well. For instance, Onwuachi-Willig gives numerous examples of multiracial relationships being rendered invisible in public "even when all the social cues point to marriage or commitment."[33] The monoracial ideal is so strong that people simply do not see the relationship.[34] Worse yet, even the children of these couples are often perceived as not being part of the family unit.[35] The situation is not

26. ONWUACHI-WILLIG, *supra* note 3, at 194.

27. *See id.* at 197.

28. *Id.* at 196.

29. *Id.*

30. *Id.* at 197–98.

31. *Id.* at 197.

32. *Id.*

33. *Id.* at 175.

34. *See id.* at 175–78 (discussing this phenomenon and describing it as "relationship erasing").

35. *See id.* at 179–84 (providing examples of instances in which parents of biracial children are questioned about their position as a parent). The number of multiracial children in the United States has grown exponentially since the 1970s. *See* Lucas, *supra* note 19 (manuscript at 10–11) (reporting that the number of people who reported multiracial status on the census grew from forty-six thousand in 1970 to almost one million in 1980 and that by 2000 that number had reached 6.8 million, although this was also the first year that people were instructed to check two racial categories on the census form if applicable).

improved when the family becomes visible because then these couples or families are interrogated about their relationship[36] or their children,[37] as their interracial coupling is perceived as the most salient aspect of their relationship. Thus, society's normative conception of family as monoracial and heterosexual renders multiracial families at once both invisible and remarkable.[38] These contrasting treatments coexist because these families exist outside the norm.

Some may dismiss "all the slights, mistreatments, and intrusions toward multiracial families [as] purely social,"[39] reminiscent of the Supreme Court's sharp distinction between social and legal discrimination in the now infamous *Plessy v. Ferguson*[40] decision.[41] The Court's analysis in *Plessy* reflected the belief of many sociologists that "stateways cannot change folkways,"[42] and Justice Powell's later opinion in *Wygant v. Jackson Board of Education*[43] expressed the view that de facto social discrimination was "too amorphous" to be addressed by antidiscrimination law.[44] However, as Onwuachi-Willig explains, "These responses ignore the way that law can work and has worked to perpetuate and reify norms and practices that continue to place interracial families on the margins of our society."[45] Thus, the law and social norms together establish and maintain the monoracial ideal, exposing multiracial families not only to social stigma but also to legal erasure.

The problem, as Onwuachi-Willig demonstrates, is that while the discrimination by association analysis gives these families a vehicle to recover damages and other forms of recompense, it does not capture the "true nature of discrimination against the mixed-race, heterosexual couple . . . . Specifically, [the discrimination] is based on interraciality and the particular stereotypes targeted at people who *together* intimately cross

---

36. *See* ONWUACHI-WILLIG, *supra* note 3, at 169 (describing how interracial couples are often asked questions such as "How did you *ever* meet?"); *see also id.* at 175–79 (providing other examples of questions multiracial couples have received about their relationship).

37. *Id.* at 170, 179–83.

38. And when it becomes apparent that the group is actually a unit, then, she writes, "In that moment, the questioner gets to exercise her privilege in discerning and naming who is family and who is not." *Id.* at 176.

39. *Id.* at 184.

40. 163 U.S. 537 (1896).

41. *See id.* at 544 ("The object of the amendment was undoubtedly to enforce the absolute equality of the two races before the law, but in the nature of things it could not have been intended to abolish distinctions based upon color, or to enforce social, as distinguished from political equality . . . .").

42. *E.g.*, Barton J. Bernstein, Plessy v. Ferguson: *Conservative Sociological Jurisprudence*, 48 J. NEGRO HIST. 196, 201 (1963).

43. 476 U.S. 267 (1986).

44. *Id.* at 276 (plurality opinion).

45. ONWUACHI-WILLIG, *supra* note 3, at 184.

racial boundaries."[46] Thus, the law "fails to address the 'expressive harms' or lack of dignity in the continued assumption of monoraciality among families in housing discrimination statutes."[47]

Throughout the book, Onwuachi-Willig carefully examines how all the civil rights statutes suffer from similar shortcomings. By rendering the multiracial family invisible, antidiscrimination law not only constructs and reinforces the normative ideal of the monoracial family but also fails to recognize the unique discrimination faced by multiracial families unless they force themselves into the monoracial ideal. Thus, antidiscrimination law reifies existing racial categories.

C. *The Remedy*

Onwuachi-Willig suggests that to solve the problem of "law's invisible role in reifying race and racial hierarchies among individuals and families,"[48] Congress should add the term "interraciality" as a protected category to antidiscrimination statutes.[49] She argues that this "modest"[50] change could have a number of important effects. First, it would provide much-needed redress for the unique harms multiracial families face and signal to advocates that they should recognize and raise these claims.[51] Furthermore, it would help to weaken society's conception of fixed racial categories, thereby creating a place for these families within society.[52] Finally, she concludes that:

> [A]dding the term *interraciality* to antidiscrimination statutes would make visible the ways in which narrowly framing the ideal for families has stifled us in our growth as individuals and communities, and it would help to expose all the differences in the lives and lived experiences of families . . . .
>
> In the end, as the assumptions that have undergirded our social norms and laws slowly begin to catch up to our nation's aspirations for equality, the text within our antidiscrimination laws should also begin to reflect those very goals.[53]

---

46. *Id.* at 197.
47. *Id.* at 198.
48. *Id.* at 233.
49. *Id.* at 234.
50. *Id.* at 233.
51. *Id.* at 264.
52. *Id.* at 265–66.
53. *Id.* at 266. There are, of course, many reasons to be skeptical that Congress or the courts would be sympathetic to such a change in antidiscrimination law. *See* Jeb Rubenfeld, Essay, *The Anti-Antidiscrimination Agenda*, 111 YALE L.J. 1141, 1143 (2002) ("[A] good deal of the present Supreme Court's groundbreaking constitutional case law makes better sense when viewed not in the doctrinal terms in which it presents itself, but in terms of an anti-antidiscrimination agenda . . . ."); Kenji Yoshino, *The New Equal Protection*, 124 HARV. L. REV. 747, 748 (2011)

In the next Part, we situate Onwuachi-Willig's book within the larger struggle for civil rights and racial equity in the United States. We argue that her insights pave the way for recognition of all the ways in which the law fails to remedy discrimination, even beyond the domain of family.

II. From Interraciality to Racial Realism

Onwuachi-Willig's close examination of multiracial families demonstrates the symbiotic relationship between the law and prevailing norms. On the one hand, the law embeds societal norms, such as the norm of the monoracial family. However, societal norms do not remain static. As societal norms change, the law also changes to incorporate new understandings. On the other hand, the law does not simply reflect societal norms, it also can play a role in legitimizing them. Thus, once the problematic assumptions embedded in the law are brought to light, the law must change to reflect new understandings. Otherwise, the law will continue to play a role in legitimizing the problematic norm.

Onwuachi-Willig demonstrates this relationship in the context of the interracial family, showing that current antidiscrimination law reflects societal expectations of monoraciality among families. Yet, as these multiracial families have become more numerous, the assumption of monoraciality has slowly begun to change, bringing the embedded assumption to light and making it clear that the law is no longer adequate to reach all the harms it was meant to redress. Some judges are beginning to recognize that antidiscrimination law is inadequate to safeguard multiracial families and have devised innovative interpretations of the law to bring these families under its protection.[54]

However, while these broader interpretations are commendable, working within existing laws rather than changing the law also has problematic consequences. As Onwuachi-Willig points out, leaving the framework in place legitimizes and reinforces the monoracial norm by requiring multiracial families to erase their multiracial status in order to fit themselves within the monoracial ideal.[55] Only by changing the law to provide explicit recognition of and protection to multiracial families will the law begin to facilitate changes in the way that society views family. Thus, the law influences norms, and norms influence the law. Furthermore, while some judges have recognized the discrimination that multiracial families endure, others have not. Now that the problematic norm has been

("[T]he [Supreme] Court has systematically denied constitutional protection to new groups, curtailed it for already covered groups, and limited Congress's capacity to protect groups through civil rights legislation." (footnotes omitted)).

54. *See supra* notes 30–32 and accompanying text.

55. *See* ONWUACHI-WILLIG, *supra* note 3, at 198. *See generally*, Lucas, *supra* note 19 (highlighting this problem in her discussion of multiracial individuals and antisubordination).

exposed, the protection these families receive should not be dependent upon the judge who happens to hear the case.

In this Part, we argue that Onwuachi-Willig's insights about the symbiotic relationship between law and society are not only important in the domain of interracial families but also in the broader cause of civil rights. This is because the antidiscrimination statutes upon which she relies also embed societal assumptions about how discrimination operates. As a result, they not only fail to remedy discrimination against interracial families, but they also fail to remedy all but the most invidious and intentional forms of discrimination.

In the short term, then, the addition of multiracial families as a protected class is important because currently they are unprotected against even intentional discrimination. Adding them to the statutes will give them the same recognition and protection that other protected classes already receive. However, these already existing protected classes only receive a remedy when they can provide evidence of malicious motives.[56] Thus, as we argue next, in addition to adding protected categories, antidiscrimination law must expand to cover less obvious, but more ubiquitous, contemporary mechanisms of bias as well. Otherwise, multiracial families, just like the protected classes that already exist, will not obtain a remedy for the more prevalent forms of bias that exist today. In subpart II(A), we discuss how antidiscrimination law currently frames discrimination. In subpart II(B), we introduce the social science of contemporary bias to highlight the disconnect between the law's construction of discrimination and how discrimination actually operates.

A. The Law's Construction of Discrimination

Currently, antidiscrimination law foregrounds intent and malicious motivations. This construction of discrimination is understandable when we consider the societal context in which the civil rights statutes were enacted. These statutes were passed in the 1960s[57] to provide remedies for the malign effects of centuries of intentional bigotry, including slavery and Jim Crow laws and policies.[58] Unsurprisingly, then, they reflect the

---

56. By malicious motives we mean that plaintiffs must show a discriminatory intent, even when the claim is based upon disparate impact. *See* Washington v. Davis, 426 U.S. 229, 239–41 (1976) (holding that laws that have a racially discriminatory effect, but that were not adopted to advance discriminatory motives, are constitutional); *see also* Bernie D. Jones, *Critical Race Theory: New Strategies for Civil Rights in the New Millennium?*, 18 HARV. BLACKLETTER L.J. 1, 14–15 (2002) ("With respect to disparate impact determinations, plaintiffs gained relief only if they could show that an allegedly discriminatory policy had discrimination as its intent.").

57. Fair Housing Act of 1968, Pub. L. 90-284, §§ 801–819, 82 Stat. 81, 81–89; Voting Rights Act of 1965, Pub. L. No. 89-110, 79 Stat. 437; Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241.

58. *See* Angela P. Harris, *Equality Trouble: Sameness and Difference in Twentieth-Century Race Law*, 88 CALIF. L. REV. 1923, 1991–96 (2000) (describing the fall of Jim Crow laws from

understanding of bias as primarily intentional and motivational. Hence, just as antidiscrimination law reflects society's framing of family as monoracial, so too does it reflect society's understanding of how discrimination manifests itself.

For instance, the employment discrimination statute, Title VII of the Civil Rights Act of 1964, prohibits discrimination based upon disparate treatment.[59] Courts typically "approach every disparate treatment case as a search for discriminatory motive or intent."[60] While this interpretation of the statute is not inevitable,[61] it reflects the common understanding of bias as arising primarily from consciously held negative attitudes and beliefs. As Professor Angela Harris has noted, this focus on intent grows out of the "essentially moralistic discourse of discrimination [that] condemns the racialist ideologies that pervaded most of twentieth century law and public policy, . . . [and] place[s] a premium on proving individual intent to harm and distinguishing innocent victims from evil victimizers."[62]

However, while the foregrounding of intent and motive made sense at a time when most forms of discrimination were overt and obvious, today discrimination no longer primarily functions this way.[63] Consequently, the

---

the end of World War II through the 1960s and the statutes passed to remedy the effects of those laws).

59. *See, e.g.*, Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 985–86 (1988) (citing to Title VII for a disparate treatment theory). While Title VII and the other civil rights statutes also prohibit disparate impact discrimination, *id.* at 990, over 90% of the cases brought under Title VII focus on disparate treatment. Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1163 (1995). For a discussion of the almost insurmountable hurdles posed by the disparate impact prong of the civil rights statutes, see Ricci v. DeStefano, 557 U.S. 557, 589–92 (2009). *See also* Sheila R. Foster, *Causation in Antidiscrimination Law: Beyond Intent Versus Impact*, 41 HOUS. L. REV. 1469, 1478 n.25 (2005) ("Since its decision in *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), the Court has yet to find disparate impact discrimination, and lower federal courts have not had much better success."); Krieger, *supra*, at 1162 n.3 (noting that an American Bar Foundation study found that "disparate impact cases comprised only 1.84% of all employment-related civil rights cases in the federal court docket between 1985 and 1987").

60. Krieger, *supra* note 59, at 1164–65.

61. *See, e.g.*, McDonnell Douglas Corp. v. Green, 411 U.S. 792, 801 (1973) ("Title VII tolerates no racial discrimination, subtle or otherwise."); Sheila Foster, *Intent and Incoherence*, 72 TUL. L. REV. 1065, 1084–85 (1998) ("The term discriminatory intent is hardly self-defining, as its meaning can range anywhere from unconscious bias to conscious bias to conscious desire to harm." (internal quotation marks omitted)); Michael Selmi, *Proving Intentional Discrimination: The Reality of Supreme Court Rhetoric*, 86 GEO. L.J. 279, 288 (1997) ("[I]t is not the Court's doctrine that has limited its vision, but the Court's vision has limited its doctrine."); Rebecca Hanner White & Linda Hamilton Krieger, *Whose Motive Matters?: Discrimination in Multi-Actor Employment Decision Making*, 61 LA. L. REV. 495, 499 (2001) ("[T]he disparate treatment inquiry should focus on causation, not conscious discrimination.").

62. Harris, *supra* note 58, at 2003.

63. *See* Dovidio, *supra* note 1, at 845 ("[A]lthough overt expressions of prejudice have declined steadily and significantly over time, subtle—often unconscious and unintentional—forms continue to exist."); *see also* Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489, 1493–94 (2005) ("[M]ost of us have implicit biases in the form of negative beliefs (stereotypes) and

traditional civil rights model does not describe either the experiences of those often targeted for disparate treatment or our best scientific understanding of the mechanisms that undergird discrimination. As a result, the law's blind spot has harmful and unintended effects similar to those identified by Onwuachi-Willig in her book. Next, we briefly discuss what social psychology reveals about the operation of bias.

B. The Mind–Science–Justice Model of Discrimination

In stark contrast to the traditional civil rights model of discrimination, which relies on lay assumptions of discrimination's origins, the mind–science–justice (MSJ) model[64] of discrimination adopts a behavioral–realist approach to questions of race and fairness in the law.[65] This behavioral–realist approach uses scientific consensus in the social sciences to develop accurate models of human behavior from which legal doctrine may then be derived.[66] It calls upon legislators, jurists, and legal scholars to identify empirically verifiable mechanisms for the causes of objectionable inequalities and to account for them in law and policy.[67]

When behavioral realism is applied specifically to domains of race and antidiscrimination legislation, the science reveals a radically different theory of how discrimination occurs and what remedies it. Specifically, while bigotry can play a role in intergroup injuries (e.g., employment discrimination, police abuses, educational inequalities), the science reveals that attitudes are a notoriously capricious predictor of behaviors—accounting for 10% of behaviors at best.[68] This suggests that the traditional

---

attitudes (prejudice) against racial minorities. . . . [T]hese implicit biases have real-world consequences . . . ."); Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997, 1004 (2006) ("[M]any scholars have drawn on advances in the empirical social sciences to demonstrate that what the law refers to as intentional discrimination can just as easily result from the uncontrolled application of implicit, unconscious, or automatic stereotypes . . . as from the operation of conscious discriminatory designs." (internal quotation marks omitted)).

64. This term was first used in Phillip Atiba Goff et al., *(The Need for) A Model of Translational Mind Science Justice Research*, 153 J. SOC. PSYCHOL. (forthcoming 2014).

65. *See id.* (manuscript at 9–11, 15–16) (stressing the need for a model, such as the MSJ model, that attempts to ground theoretical abstractions in real word experiences); *see also* Krieger & Fiske, *supra* note 63, at 1000–01 (describing the behavioral–realist approach as requiring judges to develop substantial theories without basing their analysis on inaccurate concepts of real world phenomena).

66. *See* Gary Blasi & John T. Jost, *System Justification Theory and Research: Implications for Law, Legal Advocacy, and Social Justice*, 94 CALIF. L. REV. 1119, 1120–21 (2006) (discussing how behavioral realism seeks to bring in more complete models of human behavior derived from the social sciences to help guide the business of lawyering).

67. *See id.* (identifying the provision of "more complete, empirically satisfying foundations" to models of legal behavior as a goal of behavioral realism).

68. *See generally* Dovidio, *supra* note 1 (illustrating the shifting focus from overt prejudice studied in first half of the twentieth century to the more subtle forms of prejudice that are currently being studied); Richard T. LaPiere, *Attitudes vs. Actions*, 13 SOC. FORCES 230 (1934) (illustrating

civil rights model, which presupposes that racial prejudice is nearly the sole source of discriminatory outcomes, builds from a faulty conception of discrimination. Further, the traditional model is not able to account for research innovations, particularly in the mind sciences, over the past quarter century such as the ways that implicit bias,[69] color blindness,[70] stereotype threat,[71] and masculinity threat[72] can conspire to produce racially disparate outcomes even in the absence of old-fashioned racial bigotry.

For instance, research demonstrates that as a result of a psychological process known as stereotype threat, even egalitarian individuals can engage in behaviors that create problematic racial disparities and sometimes produce deadly consequences.[73] Stereotype threat refers to the fear and anxiety that is produced when one fears being evaluated in relationship to a negative stereotype about one's group.[74] The influence of stereotype threat on interracial interactions has been well documented.[75]

---

the difference between attitudes and overt behavior in a study comparing whether hotels and restaurants would serve individuals of Chinese descent with the results of a questionnaire sent to the same establishments asking whether they would serve Chinese individuals); Allan W. Wicker, *Attitudes Versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects*, J. SOC. ISSUES, Autumn 1969, at 41, 65 (examining various aspects of the relationship between attitudes and actions and finding that "[o]nly rarely can as much as 10% of the variance in overt behavioral measures be accounted for by attitudinal data").

69. *See* John F. Dovidio et al., *Implicit and Explicit Prejudice and Interracial Interaction*, 82 J. PERSONALITY & SOC. PSYCHOL. 62, 66 (2002) (finding that the implicit attitudes of white individuals toward black individuals predicted the white individuals' spontaneous behavior).

70. *See* Victoria C. Plaut et al., *Is Multiculturalism or Color Blindness Better for Minorities?*, 20 PSYCHOL. SCI. 444, 444–45 (2009) (concluding that color blindness, "an assimilationist ideology . . . [that] stresses ignoring or minimizing group differences," can "reinforce[] majority dominance and minority marginalization" and "may promote interpersonal and institutional discrimination through social distancing").

71. *See* Phillip Atiba Goff et al., *The Space Between Us: Stereotype Threat and Distance in Interracial Contexts*, 94 J. PERSONALITY & SOC. PSYCHOL. 91, 91, 104 (2008) (demonstrating that for white individuals, stereotype threat, "the fear of being stereotyped as racially prejudiced," can lead to racial distancing); Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*, 69 J. PERSONALITY & SOC. PSYCHOL. 797, 797, 808 (1995) (defining stereotype threat and finding that stereotype threat with regard to intelligence affects the behavior of African-Americans in intellectual situations).

72. *See* Phillip Atiba Goff et al., *Racism Leads to Pushups: How Racial Discrimination Threatens Subordinate Men's Masculinity*, 48 J. EXPERIMENTAL SOC. PSYCHOL. 1111, 1115 (2012) (finding that "[f]or Black participants [in a study], the experience of discrimination threatened their masculinity, which led them to engage in compensatory masculine behaviors").

73. PHILLIP ATIBA GOFF ET AL., CONSORTIUM FOR POLICE LEADERSHIP IN EQUITY, PROTECTING EQUITY: THE CONSORTIUM FOR POLICE LEADERSHIP IN EQUITY REPORT ON THE SAN JOSE POLICE DEPARTMENT 11 (2013); Goff et al., *supra* note 71, at 91.

74. Steele & Aronson, *supra* note 71, at 797.

75. *See* Goff et al., *supra* note 71, at 104 (positing that stereotype threat affects the interaction between white and black individuals by causing racial distancing); Jennifer A. Richeson & J. Nicole Shelton, *Negotiating Interracial Interaction: Costs, Consequences, and Possibilities*, 16 CURRENT DIRECTIONS PSYCHOL. SCI. 316, 317–19 (2007) (arguing that members of dominant groups engage in self-regulation—careful monitoring of thoughts and behaviors—in order to avoid appearing prejudiced, which can lead to negative intrapersonal outcomes).

In interactions between black and white strangers, for example, egalitarian-minded whites might fear that their words and actions will be assessed as racist.[76] Meanwhile, black individuals might fear that they will be evaluated as criminal or unintelligent.[77] Researchers have found that this anxiety can affect the verbal and nonverbal behaviors of both individuals,[78] leading them to physically distance themselves from each other, make less eye contact, and use unfriendly tones of voice.[79] It is not difficult to imagine that the resulting interaction is negative for both individuals, confirming the negative stereotypes, either conscious or unconscious, that they had for each other.

We can see the relevance of stereotype threat in the context of a job interview. Imagine a white employer who interviews two equally qualified candidates who differ only in race. The effects of stereotype threat on both the employer and the black candidate can adversely affect the dynamics of the interview. As a result, the black applicant may be judged more negatively and therefore fail to obtain the job, not because of any explicit racist beliefs on the part of the decision maker but because stereotype threat adversely affected the interaction.[80] In this situation, the black candidate has been disadvantaged by bias, but not the intentional bias that current antidiscrimination law recognizes.[81] Furthermore, if it becomes apparent that this particular employer has historically hired equally qualified whites over blacks, others may ascribe his or her behavior to conscious racism even when this label is inaccurate.

We can even imagine how stereotype threat might affect the interactions of individuals with multiracial couples. An individual might want to ask an interracial couple about how they met simply because she is always curious about what brought two people together and not because she wants to interrogate the couple's interraciality. However, she may fear that this question, or some other question, might lead the couple to perceive her as having a problem with their relationship. Thus, she becomes

---

76. Goff et al., *supra* note 71, at 92.

77. Cynthia J. Najdowski, *Stereotype Threat in Criminal Interrogations: Why Innocent Black Suspects Are at Risk for Confessing Falsely*, 17 PSYCHOL. PUB. POL'Y & L. 562, 564–65 (2011); Steele & Aronson, *supra* note 71, at 797.

78. *E.g.*, Dovidio et al., *supra* note 69, at 62.

79. *See* Goff et al., *supra* note 71, at 91; *see also* NAT'L RESEARCH COUNCIL, MEASURING RACIAL DISCRIMINATION 95 (Rebecca M. Blank et al. eds., 2004) (discussing tone of voice); Dovidio et al., *supra* note 69, at 63 (discussing eye contact).

80. For a general discussion of the effects of negative stereotypes on interview interactions, see Carl O. Word et al., *The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction*, 10 J. EXPERIMENTAL SOC. PSYCHOL. 109 (1974); and also see John F. Dovidio & Samuel L. Gaertner, *Aversive Racism and Selection Decisions: 1989 and 1999*, 11 PSYCHOL. SCI. 315 (2000).

81. *See* 42 U.S.C. § 2000e–2(a)(1) (2006) (stating that it is illegal to discriminate on the basis of race in hiring).

uncomfortable, anxious, and self-conscious about everything she says, leading her to act in ways that might lead the couple to assume incorrectly that she is disturbed by the fact that they are an interracial couple. Again, stereotype threat has led to attributions of racism when, in fact, no such bigotry existed.

Similarly, stereotype threat may also be an additional reason for why interracial relationships, especially black–white interracial relationships, are rare. A white man who is attracted to a black woman may be apprehensive about approaching her because of the fear that he will be evaluated negatively because of predominant racial stereotypes of white racism. He may be concerned that the anxiety he feels about approaching someone to whom he is attracted will be mistakenly interpreted as arising out of racism. If he does engage her in conversation, she may attribute his anxious behaviors to bias rather than to jitters. Thus, the resulting unpleasant interaction may prevent a relationship from occurring.

In addition to stereotype threat, another psychological process that can cause racial disparities is implicit racial bias.[82] Implicit racial bias is a pernicious result of the way that our brains process information.[83] In order to make sense of all the information that bombards us, our brains automatically and unconsciously categorize people and objects with lightning speed.[84] Once these people or objects are placed into categories, anything that is associated with that category also comes to mind.[85] Again, this occurs unconsciously and automatically.[86] Then, these unconscious associations can predispose us to evaluate the person or object in a manner consistent with the unconsciously activated associations.[87]

---

82. *See generally* John F. Dovidio et al., *supra* note 69 (studying the effects of implicit bias on behavior during interactions between black and white individuals and finding that the implicit bias of white individuals predicted their nonverbal friendliness toward black individuals); Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCHOL. 1464, 1464 (1998) ("Implicit attitudes are manifest as actions or judgments that are under the control of automatically activated evaluation, without the performer's awareness of that causation.").

83. *See* sources cited *supra* note 82.

84. *See* SUSAN T. FISKE & SHELLEY E. TAYLOR, SOCIAL COGNITION 71 (2d ed. 1991) (discussing the "categorization" stage of information management and suggesting that the categorization of a particular action occurs spontaneously and rapidly).

85. *See id.* at 69 (discussing how fundamental attribution error occurs when one attributes behavior of another person to his or her own dispositional qualities, rather than to situational forces).

86. *See id.* at 67–69 (suggesting that fundamental attribution error is automatic and unconscious).

87. *See* Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 5 (1989) (discussing a model whereby "stereotypes are automatically (or heuristically) applied to members of the stereotyped group" even without a conscious element).

For instance, research over the past four decades demonstrates that, regardless of our race and our consciously held beliefs, most of us unconsciously associate young black men with criminality.[88] We are socialized to have this association because of what we read in the newspapers, hear on the radio, or see on television.[89] As a result, in contexts where this association is relevant, we will interpret ambiguous behaviors of black men more negatively, and as more consistent with criminality, than ambiguous behaviors of young white men in identical situations.[90] Unconscious racial bias, then, refers to how our unconscious stereotypes and attitudes can affect our behaviors and judgments in ways that we are unaware of and thus cannot control.

In her book, Onwuachi-Willig reveals how unconscious racial bias can influence our dating preferences.[91] She references a study by Professor Goff, one of this Review's authors, in which black women were rated as more masculine than their white counterparts.[92] As a result, black women were literally miscategorized as black men between 9% and 13% of the time (depending on stimulus modality, e.g., videos versus pictures).[93] Additionally, and unsurprisingly, the more masculine a woman was rated, the less attractive she was rated, regardless of race, leading to racial differences in ratings of attractiveness.[94] Importantly, when black women were not perceived as black, these effects virtually disappeared, suggesting that it is the coding of "blackness" that masculinizes black women.[95]

Furthermore, unconscious biases can also explain the association between family and monoraciality that Onwuachi-Willig uncovers in her book. As she writes, "Strangers often try to fit such couples into comfortable tropes of blackness and whiteness."[96] This observation is consistent with lessons from social psychology. We have been socialized

88. For a discussion of this phenomenon, see L. Song Richardson & Phillip Atiba Goff, Essay, *Self-Defense and the Suspicion Heuristic*, 98 IOWA L. REV. 293, 303–04, 306–07, 310–14 (2012).

89. *See* Phillip Atiba Goff et al., *Not Yet Human: Implicit Knowledge, Historical Dehumanization, and Contemporary Consequences*, 94 J. PERSONALITY & SOC. PSYCHOL. 292, 303–05 (2008).

90. *See* Richardson & Goff, *supra* note 88, at 303–07, 310–14 (referring to numerous studies that reveal an implicit association between young black men and criminality and discussing that association's effect on the way people perceive seemingly benign behaviors); *see also* L. Song Richardson, *Arrest Efficiency and the Fourth Amendment*, 95 MINN. L. REV. 2035, 2045–48 (2011) (recounting a study that demonstrated that the race of the actors dramatically influenced how observers interpreted the actors' actions).

91. ONWUACHI-WILLIG, *supra* note 3, at 128–29.

92. *Id.* at 146 (citing Phillip Atiba Goff et al., *"Ain't I a Woman?": Towards an Intersectional Approach to Person Perception and Group-Based Harms*, 59 SEX ROLES 392, 392 (2008)).

93. Goff et al., *supra* note 92, at 397 fig.2, 400 fig.4, 401.

94. *Id.* at 397–98.

95. *See id.* at 401.

96. ONWUACHI-WILLIG, *supra* note 3, at 175.

through television shows and other sources to view families as monoracial. When multiracial families are depicted, their non-normativity is highlighted, thereby reinforcing the monoracial norm.[97] Thus, our unconscious association of family with monoraciality prevents us from categorizing multiracial individuals as families in our minds.

The effects of unconscious racial bias and stereotype threat can cause problematic racial disparities in employment, housing, and a host of other legally relevant domains. Yet, antidiscrimination law fails to provide a remedy.[98] The Fourteenth Amendment's Equal Protection Clause provides that no state shall "deny to any person within its jurisdiction the equal protection of the laws."[99] While the language of the Amendment is broad enough to cover both conscious bigotry and unconscious racial bias, the Supreme Court's equal protection jurisprudence requires plaintiffs to prove that "a discriminatory purpose [was] a motivating factor" in the challenged conduct.[100] And discriminatory purpose means that "the decisionmaker . . . selected or reaffirmed a particular course of action at least in part 'because of,' not merely 'in spite of,' its adverse effects upon an identifiable group."[101] While scholars have argued that the intent standard is broad enough to cover unconscious biases,[102] thus far, the Supreme Court has not adopted this interpretation.[103]

Acknowledgment of racially disparate *impacts* might provide a remedy for the effects of unconscious biases and stereotype threat because these psychological processes result in racially disparate impacts. The Court has held that claims based upon disparate impacts must be brought under the civil rights statutes rather than under the Equal Protection Clause.[104] However, while antidiscrimination statutes purport to recognize claims of

97. Consider the recent reaction to a Cheerios commercial that depicted an interracial family—a black father, white mother, and biracial child. The reaction to the ad was so negative that Cheerios asked that the comments section on YouTube be turned off. Leanne Italie, *Cheerios Exec on Ad Featuring Mixed Race Couple: 'We Were Reflecting an American Family,'* HUFFINGTON POST (June 5, 2013, 5:14 PM), http://www.huffingtonpost.com/2013/06/05/cheerios-ad-mixed-race-couple_n_3390520.html. A Cheerios executive defended the commercial, stating that they "were reflecting an American family." *Id.*

98. *See supra* notes 26–32 and accompanying text.

99. U.S. CONST. amend. XIV, § 1.

100. Vill. of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 265–66 (1977).

101. Pers. Adm'r of Mass. v. Feeney, 442 U.S. 256, 279 (1979).

102. *See, e.g.*, Krieger, *supra* note 59, at 1243 (describing the need for coherence in distinguishing between intentional and unintentional forms of discrimination and arguing in favor of a tiered system of liability predicated on this distinction); White & Krieger, *supra* note 61, at 506–11 (arguing that some Supreme Court opinions support the idea that intent can be unconscious).

103. Harris, *supra* note 58, at 2011 ("Perhaps more disturbingly, the criminal intent rule insulates from judicial scrutiny forms of bias that are unconscious or cognitive in origin.").

104. Washington v. Davis, 426 U.S. 229, 248–52 (1976).

discrimination based upon disparate racial impacts,[105] courts still require plaintiffs to demonstrate some malicious intent.[106]

As the previous discussion demonstrates, while the Equal Protection Clause and the civil rights statutes were meant to provide remedies for discrimination, they fail because each is based upon a model of discrimination that, as shown by the science, is no longer prevalent.[107] Thus, just as the law embeds normative conceptions of family, the law also embeds norms of how discrimination operates. This construction not only reflects societal beliefs about what discrimination looks like but also legitimates them. This means that, barring significant challenges to the present conceptualization of race, racism, and how they function, both the law and society will fail to interrogate how universally human mental processes can provoke racial injustice even in the absence of conscious bias. This is why a new model of racism, broader conceptualizations of how it functions, and more inclusive understandings of protected classes are critical to making continued progress on racial equity—something that the MSJ model advances and Onwuachi-Willig's book pushes us to do.

Conclusion

Professor Onwuachi-Willig's compelling analysis reveals the importance of considering how the law's embedded assumptions can stymie progress towards racial equality. In the domain of interracial relationships, her proposal to add multiracial families as a protected category to antidiscrimination statutes is both well-supported and necessary in order to protect these families from discrimination and to broaden societal conceptions of family. Furthermore, her insights also open up space to consider more broadly how the law's and society's conceptions of discrimination also create harm. Consequently, not only should the law expand to provide protection to categories of individuals previously excluded, as Onwuachi-Willig rightly suggests, but the law should also

---

105. *See, e.g.*, 42 U.S.C. § 2000e–2(a), (k) (2006).

106. Harris, *supra* note 58, at 2010 ("Even where evidence of 'disparate impact' can establish a *prima facie* case, the courts have emphasized the plaintiff's ultimate burden of proving invidious intent."). Furthermore, claims based upon disparate impact are rare because "very few Title VII cases are actually amenable to disparate impact treatment." Krieger, *supra* note 59, at 1162 n.3 (noting that it is a "common misperception . . . that a plaintiff can prevail in virtually any type of case by making an unrebutted showing of disparate impact on a group protected by Title VII"); *see also* Linda Hamilton Krieger, *Civil Rights Perestroika: Intergroup Relations After Affirmative Action*, 86 CALIF. L. REV. 1251, 1302–03 (1998) ("Claims of disparate impact[] . . . constituted less than two percent of the courts' Title VII caseload . . . ."). In fact, since 1971, the Supreme Court "has yet to find disparate impact discrimination, and lower federal courts have not had much better success." Foster, *supra* note 59.

107. *See, e.g.*, Krieger, *supra* note 106, at 1304 (arguing that "from a social and cognitive psychological perspective, . . . reliance on the individual disparate treatment adjudications will result in the significant underidentification of discrimination, not only by decision makers, . . . but by victims and fact finders as well").

safeguard these protected categories from the harms that arise from unintentional biases. Put another way, her insights require an expansion of antidiscrimination law to include protections against these seemingly "nonbiased" sources of pervasive intergroup discrimination as well.

Finally, Onwuachi-Willig's proposal has an additional benefit. As previously discussed, unconscious stereotypes and attitudes can affect the behaviors and judgments of even the most egalitarian individuals. However, despite their ubiquity, these unconscious biases are also malleable. In other words, it is possible to overcome their effects. Lessons from the mind sciences demonstrate that one effective way to overcome the effects of unconscious biases is to make people aware of their existence and the possibility for overcoming them.[108] As Onwuachi-Willig notes throughout her book, people likely harbor unconscious biases against multiracial families. Thus, the addition of interraciality as a protected category to antidiscrimination statutes will help highlight to individuals the ways in which their unconscious biases may disadvantage multiracial families. Such awareness can help people overcome their automatic and unconscious association of family with monoracial individuals and thereby serve to mitigate some of the harms currently affecting multiracial families that Onwuachi-Willig emphasizes in her pioneering book.

108. For a discussion of these studies, see Richardson, *supra* note 90, at 2088–91, and L. Song Richardson & Phillip Atiba Goff, Essay, *Implicit Racial Bias in Public Defender Triage*, 122 YALE L.J. 2626, 2644–46 (2013).

* * *

Constitutionalism and War Making

WAR POWERS: THE POLITICS OF CONSTITUTIONAL AUTHORITY. By Mariah Zeisberg. Princeton, New Jersey: Princeton University Press, 2013. 288 pages. $29.95.

LONG WARS AND THE CONSTITUTION. By Stephen M. Griffin. Cambridge, Massachusetts: Harvard University Press, 2013. 376 pages. $39.95.

Reviewed by Peter M. Shane*

When President Obama, in August 2012, explicitly requested statutory authority for a military strike against Syria, Americans witnessed a political event without modern precedent: a president explicitly inviting full-blown congressional deliberation regarding a discrete military strike calculated *not* to start all-out war.[1] The gambit was derided by some as a signal of political weakness[2] and applauded by others as a gesture of constitutional respect.[3] Either way, the debate tested the coherence of the President's war aims, vetted the intelligence on which his strategy was based, and probed the sufficiency of his planning for both the strike and its aftermath. The substance of the debate shaped not only the committee's draft authorization for the use of military force[4] but, quite likely, the President's subsequent turn to multilateral diplomacy as an alternative strategy for deterring the Assad government's use of chemical weapons.[5]

Two thoughtful new books on war powers advocate just this kind of interbranch dialogue as a way of fulfilling the original constitutional design regarding the deployment of military force. Such debate exemplifies what Tulane law professor Stephen M. Griffin considers the constitutionally

---

* Peter M. Shane is the Jacob E. Davis and Jacob E. Davis II Chair in Law, Moritz College of Law, Ohio State University. I am grateful to Garrett Epps, Marty Lederman, Dakota Rudesill, Chris Walker, and participants in a September 19, 2013, faculty workshop at the University of Texas School of Law for thoughtful comments on an earlier version of this Review.

1. Carol E. Lee & Janet Hook, *Obama Presses Syria Gamble*, WALL ST. J., Sept. 2, 2013, http://online.wsj.com/news/articles/SB10001424127887323932604579051190561274858?KEYWORDS=Obama+Syria+Congress.

2. Charles R. Kesler, *No Thanks, Mr. President*, NAT'L REV. ONLINE (Sept. 5, 2013, 4:00 AM), http://www.nationalreview.com/article/357658/no-thanks-mr-president-charles-r-kesler.

3. Jack Goldsmith, *Congratulations President Obama*, LAWFARE (Aug. 31, 2013, 3:19 PM), http://www.lawfareblog.com/2013/08/congratulations-president-obama/.

4. *See* S.J. Res. 21, 113th Cong. (as reported by S. Comm. on Foreign Relations, Sept. 6, 2013).

5. Mark Landler & Jonathan Weisman, *Obama Delays Syria Strike to Focus on a Russian Plan*, N.Y. TIMES, Sept. 10, 2013, http://www.nytimes.com/2013/09/11/world/middleeast/syrian-chemical-arsenal.html?_r=0.

intended "cycle of accountability."[6] Likewise, had a military strike ensued, the deliberations, as much as any formal outcome, would have enhanced what University of Michigan political scientist Mariah Zeisberg calls the "constitutional authority" of such an initiative.[7] Although plainly aware that the Constitution's allocation of war powers is intended as law, both authors argue that the robustness of deliberations such as these, rather than the legality of military intervention per se, ought to be the focus of the public's concern with regard to war making. Even as a constitutional matter, they contend, the quality of our elected officials' decision-making process should matter more to Americans than whether military action is "legal" or "illegal."

As I discuss below, both the Zeisberg and Griffin volumes creatively attack the problem of constructing constitutional meaning for a critical aspect of the government's founding document—its allocation of war authorities—that generally escapes judicial interpretation and enforcement. But each, as I will also elaborate, springs too quickly to dismiss law as a significant force for the very kind of institutional behavior they so urgently seek. Congress and the Executive are routinely involved in the production, review, and application of law, even when their interpretations are not likely to be reviewed in court.[8] Actors within the political branches consequently do and should think of themselves as obligated to frame within a principled legal framework their deliberations over the exercise of government power, and this legal framing is critical to understanding the institutional dynamics and substantive outcomes that attend such deliberations.[9] War making is no exception. The Syria episode is a powerful illustration of how the legal framing of war powers decision making can help catalyze productive political deliberation. Our challenge is how to institutionalize that law-driven interbranch dialogue.

---

6. *See* STEPHEN M. GRIFFIN, LONG WARS AND THE CONSTITUTION 4–5 (2013) ("The cycle is a handy way of capturing what occurs when ordinary interbranch interaction is extended over time.").

7. *See* MARIAH ZEISBERG, WAR POWERS: THE POLITICS OF CONSTITUTIONAL AUTHORITY 45–46 (2013) (suggesting that "constitutional authority" derives from "the processes of knowledge production and interbranch deliberation").

8. *See id.* at 6–7 (suggesting that some of the laws of Congress are not likely to be reviewed in court by noting that the United States Supreme Court's power to interpret the Constitution is implied, rather than explicit, and that "the judiciary has chosen to limit its scrutiny of 'political questions' like the nature of constitutionally authoritative procedures for going to war").

9. This argument has been made by Professors David Barron and Martin Lederman with specific regard to war making. David J. Barron & Martin S. Lederman, *The Commander in Chief at the Lowest Ebb — A Constitutional History*, 121 HARV. L. REV. 941, 1106–10 (2008); David J. Barron & Martin S. Lederman, *The Commander in Chief at the Lowest Ebb — Framing the Problem, Doctrine, and Original Understanding*, 121 HARV. L. REV. 689, 721–25 (2008).

I. Mariah Zeisberg and the "Relational Conception" of Constitutional Authority

Perhaps surprisingly, the question implicitly framing the political scientist's book is the more straightforwardly normative of the two. Mariah Zeisberg's *War Powers: The Politics of Constitutional Authority* seeks to recast in a more useful way the familiar question, When are presidential war making (and congressional initiatives regulating war making) consituational? Her aim is to identify, justify, and apply meaningful new "standards for assessing interpretive fidelity" to the Constitution by both Congress and the President.[10] Recognizing the centrality of the elected branches rather than courts in constructing the meaning of constitutional war powers, she advocates standards that transcend the usual norms attached to judicial decision making in favor of norms "that capture, track, and engage . . . constitutional politics."[11]

For many legal scholars, of course, politics is the conspicuous stumbling block in reconciling the elected branches' behavior with a view of the Constitution as war powers *law*. Because the elected branches disagree as to the proper interpretation of the relevant text,[12] and there is no realistic prospect of the courts or any other authoritative third party settling their disagreement,[13] it is politics, not law, that is arguably the only significant source of constraint on their behavior. This institutional reality, Professor Zeisberg suggests, is inimical to the operation of the Constitution as *law*, however, only if we accept two conventional, but dubious, premises: first, that the relevant text must have a determinate and authoritative meaning, and second, that the determinate meaning must yield a rule that relevant actors can only obey or disobey.[14] Professor Zeisberg argues, however, that both these premises are wrong. The Constitution, she asserts, can operate as law even in political contexts,[15] even in the face of intractable interbranch disagreement,[16] and even if we understand the text as

---

10. ZEISBERG, *supra* note 7, at 10.

11. *Id.*

12. *Id.* at 5–8.

13. *See id.* at 8 (stating that there is no "final arbitrator" to decide the dispute).

14. *See id.* at 8–10 (discussing how the Constitution would ideally provide a resolution for basic political disputes and contain a policy controversy within a set of uncontroversial decision procedures that actors can rely on to assess the validity of their behavior).

15. *See id.* at 222 (noting that many constitutional problems, such as "impeachment," "the debt crisis," "appointments," and "governance overseas" are handled by looking to interbranch political interactions).

16. *Id.* at 41. I have taken the same position. Peter M. Shane, *Legal Disagreement and Negotiation in a Government of Laws: The Case of Executive Privilege Claims Against Congress*, 71 MINN. L. REV. 461, 461 (1987) [hereinafter Shane, *Legal Disagreement*]. On the applicability of the argument to war powers, see Peter M. Shane, *Learning McNamara's Lessons: How the War Powers Resolution Advances the Rule of Law*, 47 CASE W. RES. L. REV. 1281, 1298–300 (1997) [hereinafter Shane, *Lessons*].

posing the possibility of acting not just legally or illegally but rather "more or less" constitutionally, as a matter of degree.[17]

The key move for Professor Zeisberg is to understand the Constitution not only as allocating war-related authorities to the elected branches but also as erecting a particular institutional structure within which decisions regarding war making will be made.[18] That structure is purposive; the relevant institutions—Congress and the President—were given different political foundations and different organizational configurations with the expectation that they would bring to governance different kinds of resources and governing capacity.[19] As Professor Zeisberg sees it, then:

> The availability of . . . textual, ideological, and institutional resources [created by the Constitution] for ordinary politics makes it appropriate to assess constitutional fidelity not only in terms of respect for a legal framework, but also in terms of officials' relationships to a structured politics that is created and sustained through constitutional language and institutions.[20]

In other words, Congress and the President should be regarded as faithful to the Constitution not only if (and perhaps not even if) they respect the rules with which the text delimits their respective powers but also if they make robust use of the capacities for governance that the Constitution intends that they exercise. Because this assessment requires an evaluation of each elected branch's interactions with the other, Professor Zeisberg calls her theory a "relational conception" of constitutional power.[21]

Under the relational conception, a person judging the elected branches' behavior will have two sets of standards in mind. One is substantive and looks to the bases on which the Constitution authorizes American war making.[22] The other set Professor Zeisberg calls "processualist,"[23] but I will call "institutional" because "institutional" is the more familiar word and it captures Professor Zeisberg's thinking as well as her neologism. The institutional standards are implicit, she argues, in the structures and processes that the Constitution put in motion—processes through which the

---

17. ZEISBERG, *supra* note 7, at 51–52.

18. *See id.* at 8–9 (suggesting that constitutional fidelity concerning war making should be determined, in part, through the ways that elected officials interact with their political institutions).

19. *See* THE FEDERALIST NO. 51, at 317–18 (James Madison) (Clinton Rossiter ed., 1961) (arguing that each branch of government should be as independent as possible and have a will of its own in order to ensure the separate and distinct exercise of different powers of government between the three branches).

20. ZEISBERG, *supra* note 7, at 9.

21. *Id.* at 41 (emphasis omitted).

22. *Id.* at 19.

23. *Id.*

elected branches construct the meaning of the constitutional text through their interactions over military affairs.[24]

The substantive standards Professor Zeisberg identifies are "capacious,"[25] to put it mildly. She thinks it a necessary, albeit not sufficient, condition for presidentially initiated war that it be defensive in nature.[26] This limitation is presumably rooted in part in the familiar history of the drafting of Article I, in which the Framers gave Congress the power to "declare war," not "make war," in order to avoid the implication that the President could not, on his own, act to repel an invasion.[27] It is rooted additionally in the presidential oath to "preserve, protect and defend the Constitution."[28] Professor Zeisberg appears also to credit the idea that presidents lack authority to initiate wars of a certain scale—"wars that are expensive, time consuming, and pose heavy risks of casualties"[29]—although she is careful to argue that the Constitution does not necessarily exempt smaller scale conflicts from a requirement of congressional authorization.[30]

Professor Zeisberg does not directly address the issue of substantive standards to govern Congress's initiation of war. Professor Zeisberg poses as a question whether Congress should be deemed authorized to initiate aggressive, as well as defensive war,[31] although her treatment of the Mexican War suggests an affirmative answer.[32] Implicitly, this means that the only substantive standard confining Congress is that authorizations of war must advance "the common Defence and general Welfare of the United States."[33]

The real punch for Professor Zeisberg thus stems from her institutional, or "processualist," standards. The Constitution, she argues, makes certain categories and distinctions substantively relevant, e.g., "war," "repelling attack," and defense.[34] Because the effective meaning of these categories and distinctions "must be politically constructed,"[35] it is of utmost importance to the idea of constitutional fidelity that the political branches give effect to those categories in ways that the Constitution intends for those institutions to function.

---

24. *Id.* at 40.
25. *Id.* at 247.
26. *Id.* at 20.
27. *See* LOUIS FISHER, PRESIDENTIAL WAR POWER 6–7 (1995).
28. U.S. CONST. art. II, § 1; ZEISBERG, *supra* note 7, at 20.
29. ZEISBERG, *supra* note 7, at 19–20.
30. *Id.* at 20.
31. *Id.*
32. *See id.* at 83 (suggesting that the War's implications for slavery required more deliberation by the legislature than was given but not that Congress lacked authority, upon due deliberation, to initiate war against Mexico).
33. U.S. CONST. art. I, § 8.
34. ZEISBERG, *supra* note 7, at 21.
35. *Id.*

Professor Zeisberg's core argument is that normative standards for judging the elected branches' performance in war-related decision making may be deduced from three structural features of the Constitution. The first is that each elected branch enjoys an independent source of political authority; though our two political branches comprise three elected institutions—House, Senate, and President—that are popularly elected, constituency preferences for each are transformed into actual institutional membership under significantly different conditions.[36] Second, both Congress and the Executive Branch have distinctive governance capacities supported by their respective particular structures.[37] Finally, the Constitution assigns to the branches shared powers.[38] As a result, the two elected branches' exercise of power brings them into a relationship that may activate the possibilities for conflict inherent in their independent sources of authority and different perspectives on governance.[39]

In Professor Zeisberg's view, each of these constitutional features yields a normative standard for assessing the elected branches' constitutional fidelity with regard to war making. The branches' structural independence yields a norm of independent judgment, requiring each branch to "view itself as authorized and equipped to judge the constitutional and policy claims that it confronts while conducting its business."[40] The distinctive governance capacities of the branches resulting from their different structures generates a norm that each branch "exercise[] its powers . . . in ways that are connected to its relative governance strengths."[41] For the broadly representative Congress, those strengths include the capacity to "create[] and express[] divergent paths of reasoning, on both policy and constitutional matters."[42] Congress also has unique capacity, Professor Zeisberg asserts, "to pool and weigh information from multiple sources, and to generate large understandings of public policy on the basis of complex information."[43] For its part, the hierarchically organized Executive can draw on its intelligence, diplomatic, and military

36. *Id.* at 25.

37. *Id.* at 26.

38. *Id.* at 29–30.

39. *Id.*

40. *Id.* at 32.

41. *Id.* at 37. At different points, Professor Zeisberg seems to characterize slightly differently what she takes to be the normative implications of the different branches' governance capacities. Initially, she infers from these capacities two norms: (1) that their reasoning over public policy be "sensitive to the security realities they encounter" and (2) that the branches "link their arguments about constitutional authority to their substantive agendas for security policy." *Id.* at 32–33 (emphasis omitted). The formulation I have quoted above encompasses both of these ideas.

42. *Id.* at 37 (emphasis omitted).

43. *Id.* (emphasis omitted).

establishments to inform its war-making policy in nuanced and evidence-based ways.[44]

The third structural feature to which Professor Zeisberg refers, the centrality of shared powers, implies an obligation that each branch deploy its authorities in ways that facilitate response by the other. As Professor Zeisberg puts it:

> Branches that accompany their uses of the shared war power with a policy and constitutional position that the other branch may judge, and that use their powers in ways that are more rather than less responsive to the positions of their rivals, generate more constitutional authority for their behavior and for the larger war-making system.[45]

What the elected branches produce when they conform their behavior to these normative standards is not "legality," but "constitutional authority."[46] Professor Zeisberg is somewhat elliptical on what "constitutional authority" means—a topic to which I will return—but it permits two features that seeking "legality" would not. The first, as suggested above, is that it is not binary.[47] If legality is our rubric, then war is either lawful or not. With constitutional authority as the lodestar, war making can have more of it or less of it; authority exists on a continuum.[48] Second, authority can be earned with regard to war making even under conditions of interbranch conflict. Conflict, according to Professor Zeisberg, does not mean either branch is in violation of the Constitution—well-structured conflict may nonetheless produce constitutional authority if both branches are deploying their capacities well.[49]

II. Stephen M. Griffin on "Constitutional Orders" and Constitutional Change

Like Professor Zeisberg, law professor Stephen M. Griffin is deeply concerned with the quality of institutional practice and interaction in connection with war making. Unlike Professor Zeisberg, however, he does not advocate the reframing of what counts as "law" to incorporate into the idea of constitutionality any particular standards of institutional effectiveness. For Professor Griffin, constitutional law comprises judicially generated and enforceable doctrine.[50] What he critically discerns, however,

---

44. *Id.* at 35.

45. *Id.* at 38 (emphasis omitted).

46. *Id.* at 46.

47. *See id.*

48. *Id.* at 46–47.

49. *Id.* at 41.

50. GRIFFIN, *supra* note 6, at 11 (asking the reader to imagine a world where war powers are legalized (or judicialized)).

is that "beginning with Korea an amendment-level constitutional change"[51] occurred with regard to the constitutional rules of war making. Specifically, he argues, after World War II, presidents and many members of Congress came to believe that the Constitution empowers the President to initiate war without prior legislative authorization.[52] Professor Griffin persuasively explains that this construction of Article II is at odds with the original understanding of the Constitution, namely, that Presidents may initiate war only with congressional authorization.[53] This is a view he further asserts was the consensus interpretation of the Constitution with regard to military affairs until the advent of the Cold War.[54]

The question thus framing Professor Griffin's book is not, When is war constitutional?; but rather, How did the rules regarding the constitutionality of war change without formal constitutional amendment? His answer is that the process of constitutional construction—determining the rules that follow from the constitutional text—depends not just on the text but also on what he calls a "constitutional order."[55] A constitutional order, in Professor Griffin's theory, comprises both the text of the Constitution and the "multiple independent and distinctive institutions, . . . which mediate[] constitutional meaning."[56] The latter specifically include the "political and policy objectives of government officials, elites and the public," and, very importantly, the "structure and capacity for action of state institutions."[57] A constitutional order arises when these elements combine in "relatively stable patterns of institutional interaction with respect to basic aspects of the Constitution such as powers and rights."[58]

At the time of our founding, the constitutional order surrounding war powers thus encompassed more than just the Article I text that so conspicuously assigns policy priority to Congress with regard to military matters. It also included, among both the general public and our early policy elite, a "pervasive fear of excessive executive power."[59] A war-weary polity was determined to guard against unilateral Executive war making. Self-defense for the new nation had nothing to do with what we would today call "global security," and everything to do, quite literally, with immediate national self-preservation. It is evident from the

51. *Id.* at 48.

52. *Id.* at 49.

53. *Id.* at 35–40 (discussing the ratifying debates and concluding that it is clear that Congress alone possesses this power).

54. *Id.* at 7.

55. *Id.* at 4.

56. *Id.* at 14.

57. *Id.*

58. *Id.*

59. *Id.* at 44 (internal quotation marks omitted).

Constitution's text that the national government was to have only limited capacity even for that purpose.[60]

Professor Griffin refers to "three dimensions of a constitutional order: [1] plausible interpretations of the text linked to [2] attractive visions of public policy that can be implemented through . . . [3] the capacities of state institutions."[61] Under the original constitutional order, unilateral, presidential, war-making authority would not have been an attractive vision; it smacked too much of monarchy. Nor did the Executive Branch have any stable or sizeable military establishment with which to implement such a vision.[62] Both the text of the Constitution and common memory of the context in which it was written would have made it implausible to argue that presidents could take the United States to war without specific legislative license.[63] Notwithstanding numerous examples of U.S. military deployment throughout the nineteenth and the first half of the twentieth century, the original position—that Congress has to authorize war—was repeatedly reaffirmed until the Cold War by numerous "important public officials, including presidents and justices of the Supreme Court."[64]

What happened during the early years of the Cold War did not entail any change to the constitutional text. Textual change was unnecessary to generate constitutional change because the effective meaning of the Constitution is worked out in what Professor Griffin calls "the nonlegalized sphere"[65] of the Constitution—that is, in the interaction of our elected institutions with little judicial involvement. In this sphere, with little regard for constitutional wording, the evolution of new policy priorities coupled with the creation of innovative state capacities effectively generated new constitutional powers.[66] Professor Griffin argues that it was the coincidence of new policy and new capabilities which gave presidents the operational capacity to initiate war after 1945, and it is only that confluence of factors

60. It is well known that the Framers did not anticipate a standing army, providing specifically that appropriations to "raise and support Armies" not be granted for terms longer than two years. U.S. CONST. art. I, § 8.

61. GRIFFIN, *supra* note 6, at 55.

62. *Id.* at 59.

63. Alexander Hamilton, the most "presidentialist" of the Founders, famously wrote that the President's war powers "would amount to nothing more than the supreme command and direction of the military and naval forces, as first general and admiral of the Confederacy; while that of the British king extends to the *declaring* of war and to the *raising* and *regulating* of fleets and armies." THE FEDERALIST NO. 69 (Alexander Hamilton), *supra* note 19, at 416. The latter powers, he continued, "by the Constitution under consideration, would appertain to the legislature." *Id.*

64. GRIFFIN, *supra* note 6, at 53–54.

65. *Id.* at 16.

66. *Id.*

that made plausible a reading of the Constitution that, contrary to the Framers' understanding, authorizes such initiative.[67]

Professor Griffin details the evolution of new policy and new capacity with great care. The emergence of the Soviet threat after World War II generated widespread support for a permanent increase in the nation's military capacity to meet that threat.[68] Military alliances became essential to a strategy of containment and global stabilization.[69] The experience of World War II substantially discredited the world view of isolationists in Congress, and post-war mobilization left intact, and under presidential command, "the most capable armed forces in the world."[70] Also, and quite profoundly, the advent of nuclear weapons—and Russia's acquisition of nuclear arms—made imaginable a kind of war in which the case for presidential unilateralism would obviously be most compelling.[71] Due to missile flight times, a decision on retaliating to an incoming nuclear attack would have to be made in minutes or seconds, as would a decision on whether to launch a first strike if intelligence suggested Moscow was about to pull the nuclear trigger and much of its arsenal could be destroyed on the ground.

Against this background, President Truman advocated a substantial military buildup to maintain U.S. global supremacy and the security of both the United States and its allies in the face of a "long war" against Communist expansionism.[72] America's newly ambitious view of its global security role more or less commanded a new reading of the Constitution:

> As the fundamental strategy of the U.S. with respect to foreign affairs and national security, containment dictated the constitutional stance of the presidency with respect to war. . . . If that strategy involved the capacity to respond quickly to multiple foreign threats with overwhelming force, that meant the president had to possess the power to initiate military action, even war, as he saw fit.[73]

In short, a strategy of containment against the Soviet Union would not have been credible, either domestically or to the Soviet Union, if unilateral presidential military deployment were not understood on both sides to be a dependable foreign policy prerogative.

At the level of conventional legal interpretation, what made plausible a new reading of the Constitution under which the President had unilateral war-making power was the conflation of war making with foreign affairs.

---

67. *Id.*
68. *Id.* at 59.
69. *Id.*
70. *Id.* at 60.
71. *See id.* at 59–60.
72. *Id.* at 60.
73. *Id.* at 63 (emphasis omitted).

As Professor Griffin notes, "the consistent later use of [*Curtiss-Wright*[74]] by executive branch lawyers in fact exemplified the post-1945 belief that war powers could be subsumed under foreign affairs,"[75] a domain in which the President concededly has important unilateral powers.[76] This conflation obliterated a categorical understanding central to the founding generation's understanding of governmental powers, namely, that, among such powers, war making is simply different.

In terms of institutional capacity, however, it was not merely *military* capacity that changed with the Cold War. The capacities of the Executive Branch for intelligence gathering and interagency decision making were also transformed. For the first time, the creation in 1947 of the Central Intelligence Agency (CIA)[77] gave the Executive Branch "a civilian intelligence agency that would function in both peace and war."[78] Congress also created the National Security Council as a new instrument for coordinating security policy among key civilian departments and leaders of the military establishment, with direct accountability to the President.[79] In 1949, Congress gave the CIA authorization to conduct covert operations, thus legitimating Executive initiative with regard to a specific form of adversarial operation that blurs the line between military and civilian engagement.[80]

---

74. *United States v. Curtiss-Wright Exp. Corp.*, 299 U.S. 304 (1936), upheld the constitutionality of a joint resolution delegating authority to the President to impose an embargo on arms sales to Bolivia and Paraguay, which were embroiled in the Chaco War. *Id.* at 312–13, 329. Although the discretion Congress conferred upon the President would today seem utterly uncontroversial as the nondelegation doctrine is now interpreted, the Court, speaking through Justice Sutherland, used the case to advance novel propositions of foreign affairs law. *See* David M. Levitan, *The Foreign Relations Power: An Analysis of Mr. Justice Sutherland's Theory*, 55 YALE L.J. 467, 472 (1946) (describing Justice Sutherland's position as "radical"). The Court asserted, first, that the federal government derives its foreign affairs powers not from any constitutional grant, but rather from the fact of national sovereignty and, second, that participation in the foreign affairs powers of the United States is essentially limited to the President. *Curtiss-Wright*, 299 U.S. at 318–19. The Court upheld the delegation, stating: "[W]e are here dealing not alone with an authority vested in the President by an exertion of legislative power, but with such an authority plus the very delicate, plenary and exclusive power of the President as the sole organ of the federal government in the field of international relations." *Id.* at 319–20. The Court's views of the foreign relations power's sources and of presidential exclusivity have both been subject to substantial criticism. *See, e.g.*, Louis Fisher, *Presidential Inherent Power: The "Sole Organ" Doctrine*, 37 PRESIDENTIAL STUD. Q. 139, 149 (2007) (explaining that much of the scholarly treatment of *Curtiss-Wright* has been "highly critical"); Levitan, *supra*, at 493 (calling Justice Sutherland's theory "extreme" and "an unfortunate departure from the long accepted and cherished notions as to the nature of the American system").

75. GRIFFIN, *supra* note 6, at 64.

76. *See id.* at 7 (commenting that almost all presidents since Truman have claimed a unilateral power to wage war, supported by an "eminently defensible policy rationale").

77. National Security Act of 1947, Pub. L. No. 80-253, § 102, 61 Stat. 495, 497–99 (codified as amended at 50 U.S.C. §§ 403–403x (2006 & Supp. V 2012)).

78. GRIFFIN, *supra* note 6, at 101.

79. National Security Act of 1947 § 101.

80. GRIFFIN, *supra* note 6, at 102.

In each of these post-war developments, Professor Griffin discerns an important lesson often missed by critics of unprecedented Executive power, namely, that the Executive Branch did not unilaterally expand itself. The presidency may have become "imperial" in terms of its asserted prerogatives and the accompanying threat to civil liberties,[81] but it was not just presidents striking out on their own who created the modern national security state. Their capacities were undergirded by legislation that reflected a set of policy objectives and ambitions for their realization that had widespread support in both Congress and the general electorate.[82]

New governmental capacities in tandem with both a widespread assessment of the Communist danger and an equally widespread conviction as to the proper national strategy in response to that danger laid the groundwork for President Truman's unilateral commitment of U.S. troops to a "real war" in Korea. "Truman's decision," Professor Griffin writes, "was a departure from the Constitution and the American constitutional tradition."[83] Yet there was no legal process available to enforce the original legal understanding. Popular support for the war was overwhelming at first, and the President was determined "to lay down a marker and establish a 'precedent' for the exercise of presidential war powers in the new circumstances of the Cold War."[84]

Professor Griffin identifies three key constitutional ideas that came to be central to the post-World War II constitutional order. The first, and probably the most widely shared, is the idea of presidential supremacy in foreign policy. On defense and national security strategy, "Congress is effectively a junior partner, subordinate to the president."[85] The second is "the erasure of a clear distinction between wartime and peacetime."[86] Acceptance of the idea that the Cold War was an ongoing "real" war, despite formal ends to both World War II and the Korean War, implied the legitimacy of tactics, including covert war, that might have otherwise seemed "repugnant."[87] The third element is the assertion of unilateral presidential authority to initiate war.[88] No president since Truman has committed U.S. troops to combat on the scale of Korea without some form of specific congressional authorization.[89] But, with the exception of

81. *Id.* at 264.

82. *See id.* at 65–69 (detailing how the congressional support for the Marshall Plan and other foreign policy objectives, driven by fear of the Cold War, created foreign policy power in the Executive).

83. *Id.* at 71.

84. *Id.* at 73.

85. *Id.* at 96.

86. *Id.*

87. *Id.*

88. *Id.*

89. *Id.*

President Eisenhower—who disavowed the existence of unilateral authority to aid the French at Dienbienphu in 1954[90]—no president after Truman has described congressional authority to deploy U.S. troops in foreign combat as a constitutional necessity.[91] Post-Eisenhower presidents seeking congressional authority for various forms of military commitment have uniformly asserted that they did not legally need that approval in order to deploy our armed forces constitutionally.[92]

Professor Griffin's core concern regarding this change in the constitutional order is not that presidents may engage in unconstitutional wars. He is explicitly reluctant to view presidential initiatives as "legal" or "illegal" if undertaken in contexts largely unguided by judicial decision making.[93] Rather, his critical anxiety is about "the policymaking process for war"[94]—namely, that the new constitutional order, unlike the order it sought to displace, totally obviates what he calls the "cycle of accountability" that the old order implied for the sharing of war-making responsibility:

> The cycle occurs when conflict and cooperation between the branches with respect to an area of policy like war powers are repeated across time. Each branch knows it will be judged by the other and by the people. Each branch thus feels the weight of responsibility and decision. . . . Each cycle increases the chance that policy the next time around will be formulated against the backdrop of "lessons of history."[95]

As Professor Griffin documents, this cycle of accountability "has not properly operated with respect to foreign affairs after 1945."[96]

What Professor Griffin identifies as missing in the political branches' interactions is the practice of mutual testing and deliberation over war, "an ongoing institutional practice in which both branches are held accountable."[97] Under old order assumptions, a president intent on military action would need to persuade Congress of its merits, an exercise that would invite reflective discussion and require the Executive Branch to sharpen its thinking on key questions. A "legislative process to encourage deliberation" would in turn promote "war planning" and a clearer

---

90. *Id.* at 96, 105.

91. *Id.* at 96.

92. *Id.* at 33; *see also, e.g.*, Press Release, White House, Statement by the President on Syria (Aug. 31, 2013), *available at* http://www.whitehouse.gov/the-press-office/2013/08/31/statement-president-syria ("I believe I have the authority to carry out this military action without specific congressional authorization . . . .").

93. *See* GRIFFIN, *supra* note 6, at 301 n.159.

94. *Id.* at 4.

95. *Id.* at 18.

96. *Id.* at 5.

97. *Id.*

"discussion of war aims in the executive branch."[98] What we would have, at least were the system to work effectively, would be exactly what Professor Zeisberg describes as the two branches' bringing "their special institutional capacities to bear on the problem of interpreting the Constitution's substantive standards about war."[99]

Because the cycle of accountability is, as Professor Griffin sees it, the key institutional arrangement motivating the Constitution's allocation of war powers, the lesson of the Korean War for Congress was just as destructive as what the Executive Branch took from the experience.[100] Specifically, Congress learned that its acquiescence in presidential claims of inherent military authority would now give the Legislative Branch a constitutional excuse for evading political responsibility for war.[101] Should a military adventure prove popular, Congress's votes to fund the war would earn the public's political approval. Should opinion turn negative—as it did in Korea—a quiescent Congress could pretend it was merely a bystander. The absence of explicit legislative approval would enable members of Congress to distance themselves from war making as if the now unpopular initiative were wholly the President's.

The United States has no doubt witnessed time and time again the failures of war planning and military policy making that breaking the cycle of accountability portends. Professor Zeisberg's account of the Polk Administration's duplicity with regard to the Mexican War amply demonstrates that, even under "original order"[102] assumptions, the Executive Branch was fully able to undermine the interbranch deliberative process.[103] In more recent decades, despite the enactment of explicit authority for military action in both Vietnam and Iraq, presidents managed to so orchestrate the legislative process as to short-circuit real accountability.[104] Presidential (and perhaps even congressional) conviction as to the superfluity of legislative approval provided an ideological excuse for fecklessness. For this reason, it is easy to spot a direct link between the logic of unilateral presidentialism and the pathologies of Executive Branch decision making that proved so disastrous in both Southeast Asia and

98. *Id.* at 75.

99. ZEISBERG, *supra* note 7, at 18–19.

100. GRIFFIN, *supra* note 6, at 74–77.

101. *Id.* at 74.

102. Professor Griffin refers to the "original order" as the constitutional order of war powers envisioned by the Framers. *Id.* at 53.

103. *See* ZEISBERG, *supra* note 7, at 78–83 (discussing how Polk misrepresented his position to Congress with regard to the Mexican War).

104. *See* GRIFFIN, *supra* note 6, at 123, 230–31 (discussing, first, how the Tonkin Resolution passed Congress with great applause but that obtaining legal support was not the primary goal of the resolution (it was driven by political motivations); and second, how the Bush Administration "had everything to gain" by asking Congress for advice about the war aims in Afghanistan and Iraq, but failed to do so).

Iraq.[105] Even if legal, these wars were conducted with a dismal caliber of Executive Branch decision making that was too little challenged in the Legislative Branch. Pre-World War II constitutional understandings did not guarantee responsible deliberation either within or between the political branches, but the post-World War II constitutional order has largely excused Congress from testing Executive Branch arguments over war aims and plans with sufficient rigor to strengthen presidential decision making.

III. Jettisoning Law Too Hastily?

The analyses of both Professors Zeisberg and Griffin are buttressed by illuminating historical treatments of wartime decision making. The interior chapters of the Zeisberg book offer instructive comparisons of institutional performance, between (1) the run-up to the Mexican War and the run-up to World War II, (2) congressional deliberations surrounding the so-called Roosevelt Corollary to the Monroe Doctrine and congressional participation in constructing the Cold War security order, (3) Executive decision making regarding the Cuban missile crisis and presidential decision making on the bombing of Cambodia, and (4) the Senate's investigation of the connection between business conduct and U.S. entry into World War I and the Iran–Contra investigation.[106] These comparisons amply show the usefulness of Professor Zeisberg's normative criteria for assessing the elected branches' behavior with regard to the regulation and making of war.

For his part, Professor Griffin not only complicates the "imperial presidency" narrative by placing Congress firmly in the picture, but he also clearly shows how the post-World War II constitutional order played out with negative institutional consequences in both Vietnam and in the post-Vietnam period, as well as with regard to U.S. war policy after 9/11.[107] His book, viewed in tandem with Professor Zeisberg's, enacts an intriguing role reversal. The political scientist, Professor Zeisberg, connects her historical discussion more explicitly to a normative scheme for assessing the elected branches' constitutional authority. The lawyer, Professor Griffin, points more directly to the necessity for institutional change to reinvigorate the cycle of accountability he links to the original constitutional order. What directly unites the two books, however, is the authors' shared conviction

---

105. *See* PETER M. SHANE, MADISON'S NIGHTMARE: HOW EXECUTIVE POWER THREATENS AMERICAN DEMOCRACY 56–81 (2009) (using the disastrous effects of the Vietnam and Iraq Wars to illustrate how unilateral presidentialism promotes bad military and policy decision making).

106. *See* ZEISBERG, *supra* note 7, at 54–91 (comparing the run-up to the Mexican war and the run-up to WWI); *id.* at 92–145 (comparing the Roosevelt Corollary and Congress's participation in the Cold War security order); *id.* at 146–83 (comparing the Cuban missile crisis and the bombing of Cambodia); *id.* at 184–221 (comparing WWI and the Iran–Contra investigation).

107. *See* GRIFFIN, *supra* note 6, at 120–52 (discussing the post-WWII constitutional order and its negative implications); *id.* at 153–93 (discussing the post-Vietnam period and its negative implications); *id.* at 194–235 (discussing the same after 9/11).

that we should be less concerned with whether a presidential war initiative is or is not legal than we are with the quality of institutional interaction regarding war making—a regime of interaction intended to be catalyzed by the Constitution's structural design and allocations of war-making power.

The authors' shared conviction, however, runs the risk of pushing law out of the institutional picture too hastily. Some basic issues of war making are never going to be judicialized; courts are unlikely ever to pronounce particular wars legal or illegal.[108] Nor would we want or expect the elected branches, in deciding about war, to use legality as the sole or even predominant criterion for their decision making. The Iraq War, for example, was constitutional but in many respects disastrous. But having the branches pursue their interaction in a more conspicuously legal frame would not only be appropriate as a matter of principle, but it could also help advance what I take to be each author's animating project, namely, revitalizing the elected branches' deliberative interactions regarding war making.

Consider the theory side first. It is true enough, of course, that judicial intervention in decisions about how and whether to make war is all but nonexistent. It is misleading, however, to call the war powers domain of the Constitution "nonlegalized."[109] War-related decision making is deeply legalized in four conspicuous ways. First, the Constitution, which assigns the elected branches different roles with regard to war making,[110] is law. Professor Griffin himself puts the point precisely: "Even in the sphere of the nonlegalized Constitution, the document is still supreme law. It is an available text providing a normative frame of reference against which shifting political events can be evaluated."[111] Second, as part of Executive Branch decision making on war, government lawyers labor seriously (for the most part) over the precise question of whether the President's intended military operation is legal under both domestic and international law. Having served even briefly as a Justice Department lawyer, I am confident that the twenty-three post-1950 Executive Branch legal opinions that Professor Griffin cites in his appendix[112] are but a miniscule slice of government lawyers' war-related analytic work; the list of published opinions fails to expose the undoubted occasional, but invariably confidential, advice that results in military initiatives not going forward, or not going forward as planned. Third, Congress regulates war making.

---

108. The nearest counterexample so far is *The Prize Cases*, 67 U.S. (2 Black) 635 (1862), which held that President Lincoln's blockade of southern ports after the attack on Fort Sumter, and the subsequent capture of ships attempting to run the blockade, were lawful exercises of the President's constitutionally based unilateral war powers. *Id.* at 669–71.

109. GRIFFIN, *supra* note 6, at 73.

110. U.S. CONST. art. I, §§8, 10; *id.* art. II, § 2.

111. GRIFFIN, *supra* note 6, at 73.

112. *Id.* app.

Wholly apart from the War Powers Resolution,[113] Congress has enacted dozens of statutory authorities that are triggered only by declarations of war or presidential determinations that war exists.[114] Covert operations, which Professor Griffin properly identifies as central to the operation of the modern constitutional order, are governed by statute.[115] Finally, even if courts are unlikely ever to determine categorically a war's legality, the availability of courts to enforce both domestic and international rules regarding the conduct of war is potentially of great consequence—as Presidents Truman and George W. Bush both learned.[116] In all these senses, the domain of governmental war making is shot through and through with law.

Likewise, Professor Zeisberg's attempt to create a scheme of constitutional authority that emphasizes all but exclusively the quality of the branches' political interaction does not—and likely cannot—completely

---

113. 50 U.S.C. §§ 1541–1548 (2006 & Supp. V 2012).

114. *See* JENNIFER K. ELSEA & RICHARD F. GRIMMETT, CONG. RESEARCH SERV., RL31133, DECLARATIONS OF WAR AND AUTHORIZATIONS FOR THE USE OF MILITARY FORCE: HISTORICAL BACKGROUND AND LEGAL IMPLICATIONS 27–34 (2011).

115. *See* GRIFFIN, *supra* note 6, at 121 ("[I]n the post-1945 constitutional order presidents had unquestioned authority with respect to the use of nuclear weapons and covert military action . . . ."); *see, e.g.*, 50 U.S.C. § 413b (2006 & Supp. V 2012) (governing presidential approval and reporting for covert actions).

116. *See* Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 582, 587 (1952) (refusing to uphold President Truman's Executive Order on the grounds that it was a lawful exercise of the President's military power); Boumediene v. Bush, 553 U.S. 723, 732–33 (2008) (holding that prisoners in Guantanamo, detained under the authorization of President Bush, had a right to petition for a writ of habeas corpus); Hamdan v. Rumsfeld, 548 U.S. 557, 567 (2006) (referring again to Guantanamo and holding that "the military commission convened to try Hamdan lacks power to proceed because its structure and procedures violate" the Uniform Code of Military Justice and the Geneva Convention); Hamdi v. Rumsfeld, 542 U.S. 507, 509 (2004) (referring again to Guantanamo and holding that "due process demands that a citizen held in the United States as an enemy combatant be given a meaningful opportunity to contest the factual basis for that detention before a neutral decisionmaker"); Rasul v. Bush, 542 U.S. 466, 484 (2004) (holding that Guantanamo detainees have a right to petition for habeas corpus). **[AU1]** Professor Griffin writes: "Surprisingly, until the detainee cases that followed 9/11, Justice Jackson's remarkable concurrence was not thought by executive branch attorneys even to be relevant to presidential decisionmaking concerning war." GRIFFIN, *supra* note 6, at 92. The statement is incorrect, even as to published legal advice. For example, the Assistant Attorney General, in an opinion about deploying armed forces in Haiti, said:

> [T]his is not, for constitutional purposes, a situation in which the President has "take[n] measures incompatible with the expressed or implied will of Congress," [*Youngstown Sheet & Tube Co.*,] 343 U.S. at 637 (Jackson, J., concurring). Rather, it is either a case in which the President has acted "pursuant to an . . . implied authorization of Congress," so that "his authority is at its maximum," *id.* at 635, or at least a case in which he may "rely upon his own independent powers" in a matter where Congress has "enable[d], if not invite[d], measures on independent presidential responsibility." *Id.* at 637.

Letter from Walter Dellinger, Assistant Attorney Gen., to Senators Robert Dole, Alan K. Simpson, Strom Thurmond & William S. Cohen (Sept. 27, 1994), *available at* http://www.justice.gov/olc/haiti.htm.

suffice as a mode of constitutional analysis. This is easiest to see if one tries to identify the operational or even rhetorical significance of criticizing government action as lacking constitutional authority, if one does not mean to challenge legality.

For example, Professor Zeisberg recounts that, when she discusses her book project, "some mention the US entry into the Iraq War as a moment of constitutional deficiency"[117]—an intuition bolstered, no doubt, by the Executive Branch's flailing in describing its war aims, by the failures of our intelligence agencies, by the Administration's inadequate war planning, and by Congress's hasty and ill-informed deliberations.[118] It is not clear, however, what is added to the well-deserved indictment of the elected branches' abysmal institutional performance by describing it as a moment of "impaired constitutional authority"[119] or "constitutional deficiency,"[120] as opposed to just "abysmal institutional performance." To the extent congressional authorization is a legal prerequisite to war making, no scholar has seriously argued that the 2003 invasion was unlawful as a domestic matter.[121] Nor, to the extent that either Congress or the Executive Branch feels sufficiently chastened by the Iraq experience to do its job better the next time around does the constitutional-deficiency label seem to add to its already stark institutional incentives. Following Professor Zeisberg, one could accuse a President of war making that lacks constitutional authority without implying any transgression of legal limits; one could charge Congress with lacking constitutional authority in its deliberations, even if Congress has not failed to meet any legal obligation. But using a rhetoric of "constitutional authority" that is divorced from law in these ways feels a bit like saying that something is not just "bad," but "very bad."[122]

---

117. ZEISBERG, *supra* note 7, at 47.

118. SHANE, *supra* note 105, at 68–76. Excellent full-length treatments of this subject include THOMAS E. RICKS, FIASCO: THE AMERICAN MILITARY ADVENTURE IN IRAQ (2006), and MICHAEL R. GORDON & BERNARD E. TRAINOR, COBRA II: THE INSIDE STORY OF THE INVASION AND OCCUPATION OF IRAQ (2006).

119. ZEISBERG, *supra* note 7, at 32.

120. *Id.* at 47.

121. Whether the U.S. invasion of Iraq violated international law remains a subject of dispute. *Compare* John Yoo, *International Law and the War in Iraq*, 97 AM. J. INT'L L. 563 (2003) (defending legality of invasion), *with* Donald K. Anton, International Law and the 2003 Invasion of Iraq Revisited 12–13 (Apr. 30, 2013) (unpublished manuscript), *available at* http://ssrn.com/abstract=2258842 ("In addition to the Arab League, the Non-Aligned Movement of 116 and several other states wrote to the Security Council to put on record their view that the use of force in Iraq by the 'coalition of the willing' was in violation of the UN Charter.").

122. To be fair, Professor Zeisberg does not position herself as a complete legality rejectionist; she states that her method "is not an antilegal method so much as one that transcends and embraces the 'legal.'" ZEISBERG, *supra* note 7, at 223. For a war powers theory to "embrace the 'legal,'" however, that theory has to give law a more explicit role than appears in Zeisberg's account.

Given these authors' common desire—a yearning I share—to see the elected branches make their war-making decisions more responsibly, the compelling task for constitutional theorists is this: Identify the stance towards law that government actors should have in order to maximize the likelihood of activating what Professor Griffin calls the cycle of accountability. Rather than remove legality from our frame of reference in thinking about war powers, we should ask how law might best be deployed to support optimal institutional interaction.

In this regard, I would argue that instead of urging decision makers to regard themselves as part of a "nonlegalized" constitutional domain, we should want government actors habitually to understand their official behaviors with regard to war making as acts of legal obligation or legally delimited discretion.[123] In deciding what it is appropriate to do in their official capacities, they should weigh legal reference points in determining right and wrong, suitable and unsuitable. These reference points include the Constitution, statutes, judicial decisions, and legal opinions previously issued by attorneys general and the Office of Legal Counsel, as well as customary methods of legal interpretation by which officials routinely describe, justify, and thus normatively understand their governmental acts. They should do so even when no formal sanction exists to punish inattention to the relevant reference points. They should regard legal justification as an important factor even when other policy-driven or egocentric justifications would prompt the same behavior. I would argue, in short, that putting the law back into "law and politics" is a sensible move towards improving the branches' political performance.

To raise law's profile in the war powers domain may seem paradoxical given the emphatic reluctance of judges to say when wars are lawful. But, precisely because courts are unlikely to resolve key constitutional disagreements between Congress and the Executive, lawyers for both branches are positioned to take on a particularly responsible role. They can and should refrain from arguing as if they are merely advocates preparing briefs for a hypothetical third-party referee. Lawyers for both Congress and the President should hold themselves accountable to play independent, quasi-adjudicative roles in the elaboration of war-making law on behalf of their respective branches. Despite their likely disagreements, it is that common stance that is most likely to temper the antagonism of the elected branches with the public interest in workable governance. That is because no plausible congressional or Executive Branch war powers legal doctrine can completely ignore the capacities, interests, and prerogatives of the other

123. The stance towards law urged here with respect to interbranch war powers negotiation closely tracks my recommendations regarding interbranch executive privilege negotiations, in Shane, *Legal Disagreement*, *supra* note 16, at 484–501.

branch.[124] A quasi-adjudicative attitude is thus more conducive to working out over time a conscientious balancing of conflicting, legally relevant interests. Official understanding, explanation, and justification of government conduct is likely to be most responsible, and thus most consistent with workable governance, if the officials involved—and their lawyers—regard themselves as making, not just following, constitutional law.

Because the branches are likely to disagree on key principles,[125] even if each interprets the Constitution in professionally responsible ways, doctrinal analysis is only half the picture. Government lawyers and their clients need to couple a commitment to conscientious legal analysis with a second institutional commitment—an ethos of negotiation that steers officials, while actually bargaining, away from the vindication of doctrinal principle for its own sake and toward the reconciliation of the competing institutional interests that each branch recognizes.[126] Congress, that is, may always insist, in principle, that legislative authorization is prerequisite to lawful war making.[127] Presidents may continue to insist on unilateral power in theory.[128] The branches can "agree to disagree" in principle, however, so long as Congress is willing in practice to concede that some of its authorization may be implicit or indirect, and the President, in practice, effectively concedes that his initiatives are subject to congressional regulation.

The regime I have in mind of principled lawyering and pragmatic negotiation is perfectly consistent with Professor Zeisberg's criteria for the elected branches' constitutional authority. Rather than demand that the branches reach a settled legal consensus, I would urge each to come to a responsible, independent interpretation as to the allocation of war-making authority under the Constitution. Such lawyering would accord with the

---

124. *See* United States v. AT&T Co., 567 F.2d 121, 127 (D.C. Cir. 1977) (stating that the Framers intentionally abstained from defining governmental powers in minute detail with the expectation that the branches would compromise with each other and seek "optimal accommodation" when a conflict in scope of constitutional authority arose).

125. *See, e.g.*, Kucinich v. Obama, 821 F. Supp. 2d 110, 112 (D.D.C. 2011) (examining a conflict between the President's use of military action in Libya as an exercise of his "authority to conduct U.S. foreign relations" and select Congress members' views that this action violated the War Powers Clause).

126. Shane, *Legal Disagreement*, *supra* note 16, at 501; *see also* Peter M. Shane, *Negotiating for Knowledge: Administrative Responses to Congressional Demands for Information*, 44 ADMIN. L. REV. 197, 226–34 (1992) [hereinafter Shane, *Negotiating for Knowledge*] (suggesting one framework for reforming tensions between the Legislative and Executive Branches in the information-sharing context).

127. *See* U.S. CONST. art. 1, § 8, cl. 11 (giving Congress the authority to declare war).

128. John C. Yoo, *Kosovo, War Powers, and the Multilateral Future*, 148 U. PA. L. REV. 1673, 1688–95 (2000) (outlining the theoretical argument for and benefits of unilateral presidential war-making power).

Zeisberg norm of independent judgment.[129] To the extent the branches disagree on first principles, I would want each to negotiate in a manner respectful of what the other branch takes to be its core governance interests. This accords with Professor Zeisberg's prescription that the branches "use their powers in ways that are more rather than less responsive to the positions of their rivals."[130] Finally, in negotiating over how or whether to proceed with any war-making project, I would ask that the negotiation take on the character of problem solving, rather than doctrinal vindication, which would mean that each branch would bring to the table its strongest evidence-based arguments for proceeding in one way rather than another. This would fulfill Professor Zeisberg's norm of connecting each branch's exercise of power to its particular governance strengths. A mutual commitment to the kind of legally framed, problem-solving negotiation regime I am advocating would do much to help restore the cycle of accountability central to Professor Griffin's analysis. Framing interbranch negotiation as enacting a governing regime of law could function not as a distraction from politics, but as a rubric to motivate more responsible politics.

There is, it should be said, one critical way in which the legal stance I am advocating departs markedly from the constitutional approaches of both Professors Zeisberg and Griffin. Although both writers seek a form of institutional interaction in which the relevant actors learn lessons from history, each resists the idea that a recognizable war powers law might be elaborated over time between the elected branches if they begin to treat historical experience as a source of something like legal precedent.[131]

---

129. ZEISBERG, *supra* note 7, at 31.

130. *Id.* at 38 (emphasis omitted).

131. Both Professors Zeisberg and Griffin write critically of practice-based reasoning about constitutional rules, but their arguments should not be carried too far. Professor Zeisberg writes: "No plausible general theory of constitutional authority tells us that long-standing practice on its own is authoritative." *Id.* at 18. Arguments based on practice, however, are typically offered within a framework argument based on constitutional text and purposes. Practice "on its own" is rarely the issue.

In a similar vein, Professor Griffin observes: "Especially on the presidentialist side, participants [in these debates] tend to avoid the uncomfortable truth that the rule providing Congress the responsibility for authorizing war is supreme law and hence legally obligatory. Subsequent practice could not count against or nullify the rule without placing the entire Constitution in question." GRIFFIN, *supra* note 6, at 48 (footnote omitted).

Again, this analysis is persuasive to the extent practice is offered as the exclusive basis for reading the Constitution to vest the President with unilateral power to initiate full-scale "war"—a power recent presidents have claimed rhetorically, but none other than Truman has acted upon. What is often debated, however, is what counts as "war" under Article I and the level of explicitness with which Congress is obliged to authorize military engagement. Subsequent practice might well be thought helpful in constructing rules with regard to such questions, an interpretive proposition that is different from changing "rules in the Constitution," *id.*, and which does not at all place the Constitution in question.

Professor Zeisberg is emphatic on this point. She states that one way her "relational conception can be understood as 'political' relates to the limited place it gives to precedential reasoning."[132] She acknowledges that reference to precedent may assist officials in making good judgments and help assure the public that they are behaving reasonably.[133] But Professor Zeisberg argues that the priority her theory gives "to the intersection between the constitutional capacities of the branches and the security context of the moment calls for decidedly presentist forms of justification. The relational conception prioritizes good judgment in the particular context over and above consistency across cases."[134] To my mind, however, this call for "prioritization" seems like a bit of a red herring. A norm of consistency across historical cases will be predictably weaker in any decision-making context where politics plays a huge role. That is no reason to deny that the systematic reevaluation of historical examples can, over time, yield mutual interbranch understandings that take on the form of norms. International law is built substantially on the normative force of practice; there is no reason in principle why separation of powers law cannot follow suit.[135]

Professor Griffin's discontent with practice-based arguments may be more contingent. The practice-based arguments offered for unilateral presidential war-making power have, as he argues, been made sloppily and typically without any criteria for which interventions count and which do not.[136] As a result, he asserts, the practice-based argument for unilateral presidential war making "has long been one of the most disreputable arguments ever made within separation of powers jurisprudence."[137] Indeed, some of the precedents for presidential unilateralism on which modern presidents rely now seem linked to an imperialist program of the late nineteenth and early twentieth centuries that no president would now want to be associated with. As Professor Griffin writes: "It would appear difficult to base a *contemporary* argument for presidential intervention on obsolete foreign policies that later presidents officially repudiated."[138] The argument Professor Griffin makes, however, is precisely the kind of *lawyerly* dispute over precedent that the legalized negotiation regime I

132. ZEISBERG, *supra* note 7, at 250.

133. *Id.* at 251.

134. *Id.*

135. *See* Christopher A. Ford, *War Powers as We Live Them: Congressional-Executive Bargaining Under the Shadow of the War Powers Resolution*, 11 J.L. & POL. 609, 613–20 (1995) (arguing that the political branches' post-War Powers Resolution behavior had created an "effective law" of war powers); *see also* Shane, *Lessons*, *supra* note 16, at 1298–304 (suggesting that "custom [can be] a source of law in separation of powers disputes" and treating the behavior of institutional actors under the War Powers Resolution as an example).

136. GRIFFIN, *supra* note 6, at 78.

137. *Id.*

138. *Id.* at 84.

advocate would welcome. It should, indeed, be odious to cite U.S. pro-colonial interventions in China[139] or repeated incursions into Central and South America[140] as legal support for presidential war making, just as it would be to write a 2013 equal protection argument based on *Korematsu v. United States*[141] or *Plessy v. Ferguson*.[142] But that sounds like a productive and law-inflected argument for the political branches to have.

Professor Griffin unfortunately limits the deserved institutional reach of his own analysis by marginalizing debates over the legality of military deployments short of large-scale war. As noted above, he thinks judgments of legality are not useful in the absence of judicial precedent.[143] He further

> maintain[s] that the [war powers] debate should be centrally concerned with the ability of the executive branch to initiate war, "real" wars, major wars, rather than under what circumstances it can use military force to rescue citizens, intimidate the nation's enemies, and cooperate with allies and international organizations in humanitarian endeavors.[144]

Professor Griffin is correct, of course, that major wars exact extraordinary sacrifices and threaten the greatest harm to constitutional values.[145] But the fact remains that, apart from Korea—and notwithstanding various presidents' announced convictions—no real war has proceeded since Korea without some form of express congressional authority.[146] The greatest harm wrought by presidential claims of unilateral authority to make real war is likely to be the spillover effect on lesser military engagements. It is hard to see how a cycle of accountability can be resurrected for military deployments, large or small, without resort to some theory of constitutional limits and obligations. Limits-and-obligations talk is the stuff of law.

IV. Can a Legally Informed Interbranch Negotiation Practice Be Institutionalized?

Professors Zeisberg and Griffin are persuasive that U.S. constitutionalism would be better served by a more reliably accountable and

---

139. *See, e.g.*, Jerome Alan Cohen, *China and Intervention: Theory and Practice*, 121 U. PA. L. REV. 471, 476 (1973) (discussing the nineteenth-century opening of China by western military force).

140. *See supra* note 73.

141. 323 U.S. 214, 217–18 (1944) (effectively upholding the internment of Japanese-Americans during World War II).

142. 163 U.S. 537, 550–51 (1896) (validating "separate, but equal" as the relevant standard in judging racially segregated public transportation facilities).

143. *See supra* text accompanying note 92.

144. GRIFFIN, *supra* note 6, at 30 (emphasis omitted).

145. *See id.* at 6.

146. *See supra* text accompanying note 89.

deliberative interbranch relationship regarding war making. Whether described as a system of deeper constitutional authority, a new constitutional order, or the institutionalization of legally framed interbranch negotiations, an obvious question looms: Could any such system be brought about?

Professor Griffin's answer is not hopeful. He writes:

> What Congress should be asking for today is for the executive branch to consistently engage with it on matters of foreign policy and national security strategy . . . . [D]oing this in a meaningful way will require Congress to change its structure, much as the executive branch was reorganized after 1945.[147]

The problem, that is, with our overreaching presidency does not lie exclusively with the ambitions of the Executive Branch. A profound obstacle to restarting a reliable cycle of accountability is that Congress is not so structured as to be an effective, coequal partner in foreign policy and national security deliberations. Although Professor Griffin does not deal with structural reform proposals in detail, the thrust of his thinking is clear: Congress needs to centralize authority in its leadership to deal with the Executive Branch in military affairs so that the President has credible, experienced consultation partners, who would be able—as the occasion calls for it—to mobilize their colleagues to support the President's national security plans going forward. Congress would also have to have processes in place adequate "to test the executive branch's claims with respect to war and foreign affairs in a way that works to the advantage of both branches and the nation as a whole."[148]

Intriguingly, a promising first draft of a blueprint for reform could be a redraft of the War Powers Resolution introduced in 1995 by then-Senator Joseph Biden.[149] Among other things, his so-called "Use of Force Act"[150] would have set forth in far greater detail (albeit with more deference to the Executive than exists in the War Powers Resolution) Congress's

147. GRIFFIN, *supra* note 6, at 251.

148. *Id.* at 257. In the 1980s, the American Bar Association endorsed a set of measures for the more constructive resolution of interbranch disputes over access to Executive Branch information, notwithstanding profound differences between Congress and the President regarding the contours of executive privilege and its availability against Congress. Shane, *Negotiating for Knowledge*, *supra* note 126, at 231. Its suggestions, and a similar recommendation from the Administrative Conference of the United States, Recommendations of the Administrative Conference Regarding Administrative Practice and Procedure, 55 Fed. Reg. 53,269 (Dec. 28, 1990) (to be codified at 1 C.F.R. pts. 305 & 310), were based on this author's study advocating the kind of legally framed interbranch negotiation regime for information disputes, Shane, *Legal Disagreement*, *supra* note 16, at 465–66, that I have also advocated for war making. The ABA specifically asked Congress to study the possible wisdom of creating a central body in Congress with continuing responsibility for executive privilege negotiations. Shane, *Negotiating for Knowledge*, *supra* note 126, at 234.

149. *See* S. 564, 104th Cong. (1995).

150. *Id.* § 1.

understanding of the permissible reach of presidential military authority.[151] It would have also given the President the benefit of a statutory "Congressional Leadership Group" to facilitate interbranch consultation.[152] A cycle of accountability could be yet further strengthened by creating in Congress "a single, principal point of oversight and review"[153] for decisions about war. This could be either a joint committee or a single committee with jurisdiction in each house.

Unfortunately, nothing in Congress's track record sustains optimism for structural reform. Individual members may prize reelection even more than policy impact,[154] but it seems a safe bet that most would prefer some personal committee roles that would assure a measure of visibility and the appearance of influence for each and every member. That reality and the ever-more-decentralized nature of political fund-raising[155] have exerted a persistent centrifugal force on Congress that works to resist centralization initiatives. The kinds of reform needed for war deliberations resemble the kinds of reform that have been repeatedly urged for the system of congressional intelligence oversight. Seven of twelve intelligence and terrorism studies between 1991 and 2001 recommended congressional restructuring; Congress did nothing.[156] The 9/11 Commission recommended that Congress create "a single, principal point of oversight and review for homeland security."[157] It has not happened.[158]

As I have argued elsewhere, however:

> Remarkably, . . . it is possible in the current political moment to imagine a confluence of political forces that could actually unite around a program of war powers reform. President Obama's unprecedented turn to Congress for explicit authority to stage a

---

151. *See id.* § 101 (enumerating the specific scenarios in which the President could unilaterally authorize the use of military force abroad and the proper principles according to which such force could be employed).

152. *Id.* § 102.

153. This is the phrase the 9/11 Commission used in recommending structural change in Congress for oversight of homeland security. NAT'L COMM'N ON TERRORIST ATTACKS UPON THE U.S., THE 9/11 COMMISSION REPORT 421 (2004) [hereinafter 9/11 COMMISSION REPORT].

154. *See* Victoria Nourse, *Misunderstanding Congress: Statutory Interpretation, the Supermajoritarian Difficulty, and the Separation of Powers*, 99 GEO. L.J. 1119, 1126 (2011) (noting the consensus among political scientists that members of Congress care intensely about reelection).

155. *See* Katharine Q. Seelye & Leslie Wayne, *The Web Takes Ron Paul for a Ride*, N.Y. TIMES, Nov. 11, 2007, http://www.nytimes.com/2007/11/11/us/politics/11paul.html?pagewanted=all (discussing with a visiting Duke law professor who ran Howard Dean's 2004 presidential campaign how the internet has led to decentralization of political funding).

156. AMY B. ZEGART, SPYING BLIND 154 (2007).

157. 9/11 COMMISSION REPORT, *supra* note 153.

158. NAT'L PREPAREDNESS GRP., BIPARTISAN POLICY CTR., TENTH ANNIVERSARY REPORT CARD: THE STATUS OF THE 9/11 COMMISSION RECOMMENDATIONS 16 (2011), *available at* http://bipartisanpolicy.org/sites/default/files/CommissionRecommendations.pdf.

> limited strike operation against Syria has revealed a complex political landscape with regard to war-making.
>
> The Democrats are divided primarily among the "humanitarian interventionists" and "presidentialism skeptics." That is because some liberals believe U.S. military interventions can be effective tools for forestalling international atrocities, while others are wary of the utility of U.S. military interventions and, perhaps more profoundly, the breadth of presidential claims of authority to deploy U.S. military forces unilaterally.
>
> The Republicans are divided between "military hawks" and "neo-isolationists." The former are conservatives persuaded that radical Islamists have taken the place of Communists as a persistent, ubiquitous global enemy and that nearly unfettered presidential military discretion is essential to defeat their cause; the latter, like presidentialism skeptics, doubt the utility of U.S. military interventions and are inclined to resist international deployments because of their fiscal, as well as geopolitical consequences.

A legislative project to reinvigorate the cycle of accountability could potentially unite the neo-isolationists and presidentialism skeptics. It is noteworthy also that the draft Use of Force Act[159] would have gone beyond the War Powers Resolution in legitimating presidential military initiative. It would authorize the President to use force, in conformity with the Act, to forestall imminent acts of terrorism[160] and to protect internationally recognized rights of innocent and free passage in the air and on the seas.[161]

> Providing explicit Congressional sanction for presidential action aimed at such purposes might win the support of both the humanitarian interventionists and perhaps even some military hawks.[162]

In any event, what is clear is that the kinds of institutional performance that Professors Zeisberg and Griffin both crave depend on reforms within Congress that will be motivated, if at all, only by external political pressures. What is necessary is that those pressures coalesce to support the institutionalization of new structures and processes that will enable the cycle of accountability these authors describe. President Obama's handling

---

159. *See* text accompanying notes 148–51.

160. The requirement of "imminence" would no doubt be of concern should the Use of Force Act or something like it ever be reintroduced into Congress. My point is not, however, to endorse any particular wording but simply to point out the possibility of elaborating permissible grounds for presidential military deployments in a manner more thoroughgoing than the War Powers Resolution.

161. S. 564, 104th Cong. (1995) § 101 (a).

162. Peter M. Shane, *Using the Syria Debate to Launch War Powers Reform*, HUFFINGTON POST (Sept. 9, 2013, 9:57 AM), http://www.huffingtonpost.com/peter-m-shane/using-the-syria-debate-to_b_3881124.html.

of Syria would seem to support this point. Americans are unlikely to learn any time soon what precise mix of decision-making factors led President Obama to invite Congress's participation in deciding on a military strike. One of those factors was quite probably the excuse his invitation provided for delaying a strike whose prospect was not proving popular either with international allies or with the American public[163]—an entirely political consideration. But it would be just as surprising if the President were not focused on the domestic legal legitimacy that a legislative authorization to use military force would confer. Indeed, without that legal framing, it seems a fair guess that even the political utility of going to Congress would have been greatly reduced. So the question is how to get presidents to make this sort of legally influenced judgment more often.

The crux of my friendly amendment to the Zeisberg and Griffin analyses is simply this: I believe each understates the potential role of law in catalyzing those structures and processes that could lead to more effective interbranch deliberation. Even if legality does not—and perhaps should not—be the predominant factor in war-making decisions, what Professor Griffin calls "the nonlegalized Constitution" can still be helpfully infused by the values, attitudes, and analytic approaches associated with conscientious lawyering.

163. *See* Andrew Dugan, *U.S. Support for Action in Syria is Low vs. Past Conflicts*, GALLUP POL. (Sept. 6, 2013), http://www.gallup.com/poll/164282/support-syria-action-lower-past-conflicts.aspx (finding only 36% of Americans favored military action in Syria compared with 59% in Iraq, 82% in Afghanistan, and 62% in the Persian Gulf).

* * *

Notes

Buyers Without Remorse: Ending the Discriminatory Enforcement of Prostitution Laws*

During the Progressive Era, America seemed to wake up to the real threat of the "Social Evil." Prostitutes, who had hitherto been cast as unfortunate and naïve women who allowed themselves to be seduced and ruined, were now seen as dangerous carriers of frightening and incurable disease. The Federal Government reacted by passing the Mann Act in 1910. Within 15 years prostitution had been criminalized in every state.

Criminalization, however, only ever really affected the sellers of sex. The demand side of commercial sex—comprised of men who were given the common, judgment-free, and anonymous-sounding appellation "john"—continued to buy sex with near impunity. Over the course of the twentieth century, police departments perfected methods of finding and arresting prostitutes, including the use of street sweeps and male decoys. Few women who were charged with prostitution challenged these methods. The few who did came to court armed with statistics showing pervasive discriminatory enforcement of prostitution laws against prostitutes and even police testimony admitting the same. However, these women overwhelmingly saw their defenses thrown out. While a small and modestly growing number of enlightened judges have dismissed cases against women charged with prostitution on the grounds of discriminatory enforcement, the problem remains. According to recent FBI statistics, roughly two women are arrested for prostitution for every one man.

This Note urges more courts to recognize discriminatory enforcement as a defense to a prostitution charge when a defendant produces either statistical or testimonial evidence that supports the defense. This will necessitate three important changes in how courts currently assess discriminatory enforcement claims. First, courts will need to recognize that prostitutes and johns are similarly situated. By doing so, they will not be able to ignore statistics that show a large disparity in arrest rates between prostitutes and johns. Second, courts will need to lower the burden of proof for proving discriminatory intent. Third, courts will need to closely scrutinize the traditional justifications for

---

* I am grateful to Professor Cary Franklin for providing insight and encouragement long after her seminar had ended. I would also like to thank the Volume 92 Notes editors, Ryan Meltzer, Dina McKenney, and Brent Rubin for their hard work on this Note and for making my experience in the Notes Office so memorable. Finally I would like to thank my family—Mom, Molly, and Cassie—for their unwavering support and love.

selective enforcement, reflexively offered by police departments and prosecutors, and reject those excuses that are tainted by sexism. In so doing, courts will force police departments to treat prostitution as a crime inherently involving supply and demand and enforce antiprostitution statutes, laws prohibiting an ancient crime, in a modern manner.

I. Introduction

In 1920, an unusually enlightened New York judge wrote, "The court is aware that it has been the custom heretofore followed to arrest the women and let the men go; but the time has come when the custom cannot longer be permitted to continue."[1] The set of facts that elicited this particular judge's frustration would have been familiar to many judges who took on criminal cases. Two police officers, under the guise of delivering a telegram, pushed their way into a small apartment.[2] Inside they found two men and two women in a state of undress.[3] After the men admitted to having paid the women for "a good time," the police officers ordered the women to give the men a refund.[4] One officer then took the men aside, telling them, "I am going to arrest these women, and if you are interested in them, why you can appear as a witness in their behalf."[5] After both men politely declined, they were released and the women were tried and convicted of vagrancy.[6] This exchange, which elicited the frustration of the judge who reversed one of the women's convictions, was already typical by the early twentieth century.[7] Despite this judge's resolution that "the custom cannot longer be permitted to continue,"[8] continue it did, repeated in hotel rooms, cars stopped in parking lots, and darkened clubs.

This Note takes the position that enforcement of prostitution laws against the sellers of sex disproportionately harms women and that courts should act to end this practice by recognizing the defense of discriminatory enforcement of prostitution laws. Part II contains a short note on terminology. Part III provides the historical backdrop against which the current enforcement of prostitution laws is set. Part IV argues that the discriminatory enforcement of prostitution laws is an ongoing problem and provides statistical evidence to support this claim. Part V outlines the background and elements of the discriminatory-enforcement defense. Part VI urges courts to recognize this defense and to make it easier for defendants to prove by acknowledging that prostitutes and johns are

1. People v. Edwards, 180 N.Y.S. 631, 635 (Ct. Gen. Sess. 1920).
2. *Id.* at 632.
3. *Id.*
4. *Id.* at 632–33 (according to the police officer's testimony, the sum was $5 each).
5. *Id.* at 634.
6. *Id.* at 632, 634.
7. *See infra* Part III.
8. *Edwards*, 180 N.Y.S. at 635.

similarly situated, lowering the standard of proof needed to establish the defense, and discarding the traditional, sexist justifications for the continuation of discriminatory policies. Part VII briefly concludes.

II. A Note on Terminology

This Note focuses on female prostitutes and refers to them simply as "prostitutes" for two reasons.[9] First, this Note confines its analysis to prostitution as it currently exists. Evidence suggests that female prostitutes still greatly outnumber male prostitutes in all areas of sex work, including street prostitution, although exact numbers are unavailable.[10] Furthermore, female street prostitutes are also targeted for arrest in greater numbers, partly due to male police officers who pose as prospective johns.[11]

Second, both the historical and popular conceptions of prostitution are that it is a job done by women. The tie between the image of the prostitute and the idea of womanhood is ancient and enduring. Indeed, in *Ex parte Carey*,[12] a Progressive Era court expressly held that prostitution cannot be committed by a man.[13] Although society's understanding of prostitution has since evolved—courts generally admit that prostitution is something that can be done by a man[14]—prostitution still retains its traditional image as a woman's profession.[15]

---

9. This Note recognizes that male prostitution is a real phenomenon that has received growing recognition in recent years. It also recognizes that the ratio of male to female prostitutes cannot be estimated with any great accuracy and will almost certainly change over time. This Note also takes the position that male prostitution gives rise to unique issues that merit their own treatment.

10. *See* Jacqueline Cooke & Melissa L. Sontag, *Prostitution*, 6 GEO. J. GENDER & L. 459, 470 (2005) (citing an estimate that male prostitutes comprise at most one-third of prostitutes in urban centers and admitting that the actual "ratio of female to male prostitutes is unclear"); Catharine A. MacKinnon, *Trafficking, Prostitution, and Inequality*, 46 HARV. C.R.-C.L. L. REV. 271, 291 & n.70 (2011) (providing examples of publications that state that the majority of prostitutes are women but give different estimates of the ratio of female to male prostitutes); Charles H. Whitebread, *Freeing Ourselves from the Prohibition Idea in the Twenty-First Century*, 33 SUFFOLK U. L. REV. 235, 240 (2000) (observing that although female prostitution is the most common form of prostitution, male prostitution is increasing).

11. *See infra* Part IV.

12. 207 P. 271 (Cal. Dist. Ct. App. 1922).

13. *Id.* at 275 ("The words 'soliciting for prostitution' have a well-understood and distinct meaning. They are held to mean the act of a fallen woman in hailing passers-by . . . . They are so understood by police officers and all others who are called upon to labor with this class of people.").

14. *See, e.g.*, People v. Superior Court of Alameda Cnty., 562 P.2d 1315, 1322 (Cal. 1977); Blake v. State, 344 A.2d 260, 262 (Del. Super. Ct. 1975); State v. Gaither, 224 S.E.2d 378, 379 (Ga. 1976).

15. Feminist scholar Evelina Giobbe comments on the strength of the tie between prostitutes and womanhood: "The prostitute symbolizes the value of women in society. She is paradigmatic of women's social, sexual, and economic subordination in that her status is the basic unit by which all women's value is measured and to which all women can be reduced." CATHARINE A. MACKINNON, WOMEN'S LIVES, MEN'S LAWS 160 (2005).

Thus, the assumption that most prostitutes are women more or less accurately reflects reality, for better or for worse. However, many police departments, prosecutors, and courts make the additional assumption that prostitutes alone merit punishment. This assumption contributes to the perpetuation of police practices that result in the discriminatory enforcement of prostitution laws.

III. History of Prostitution in the United States

The history of prostitution in the United States is marked by swings between aggressive antiprostitution campaigns and grudging tolerance of its existence.[16] Throughout this schizophrenic history, three things have remained constant. First, the American people have almost universally refused to legalize and regulate prostitution.[17] Second, state prostitution laws have specifically targeted the sellers of sex for criminal sanctions.[18] Third, the American people have created stories about prostitutes that fit the exigencies of the time. Some of these stories matured into stereotypes, which in turn became the rationales underlying prostitution laws and their selective enforcement.[19] Part III of this Note will examine some of these prostitution narratives and how they have impacted the development of prostitution laws in the United States.

A. Prostitution in Early America: The Age of Relative Tolerance

In early America, prostitution was not a distinct criminal offense, although prostitutes themselves were "considered a disgrace."[20] In colonial America, men and women who committed chastity offenses were punished

16. BARBARA MEIL HOBSON, UNEASY VIRTUE: THE POLITICS OF PROSTITUTION AND THE AMERICAN REFORM TRADITION 4 (1987).

17. *Id.*

18. *See id.* at 5 (observing that "[h]istorically, the main goal of the feminist attack on prostitution policy has been to achieve the equal application of the laws" and explaining that while "[f]eminists have sought stricter enforcement of laws against keepers and pimps, and most important, criminal penalties for men who buy prostitutes' services[,] . . . even when such laws were enacted they were never enforced").

19. For a discussion of the historical roots of discriminatory criminalization of prostitution, see Julie Lefler, Note, *Shining the Spotlight on Johns: Moving Toward Equal Treatment of Male Customers and Female Prostitutes*, 10 HASTINGS WOMEN'S L.J. 11, 12–16 (1999). As Lefler observes:

> Much of the differential treatment of prostitutes and johns in the United States today can be traced to the sexual double standard present throughout this country's history. America's past is fraught with sympathy and excuses for the sexual appetites of men, yet condemnation of women for essentially the same behavior.

Id. at 12.

20. Kate DeCou, *U.S. Social Policy on Prostitution: Whose Welfare Is Served?*, 24 NEW ENG. J. ON CRIM. & CIV. CONFINEMENT 427, 430 (1998).

under lewdness or nightwalking statutes.[21] Prosecutors and police enjoyed wide discretion, and they used this discretion to protect johns while punishing prostitutes. As a consequence, johns were never charged in equal proportion to prostitutes, and even when men were charged as customers, they were convicted at lower rates and received lighter sentences, usually fines, while women were imprisoned.[22]

Beginning in the 1830s, discourse on prostitution was dominated by the "fallen woman" narrative, in which an innocent girl from a humble background was seduced, ruined, and abandoned by a rich and powerful man.[23] This narrative was founded in Victorian logic, which envisioned a strict dichotomy between women who were "thoroughly depraved" and women who were "thoroughly virtuous."[24] A thoroughly virtuous woman's morality was governed by finely tuned intuition that, once set askew, could never be set right.[25] A passage by the Victorian writer William Acton paints a vivid picture of the fallen woman:

> She is a woman with half the woman gone, and that half containing all that elevates her nature, leaving her a mere instrument of impurity; degraded and fallen she extracts from the sin of others the means of living, corrupt and dependent on corruption, and therefore interested directly in the increase in immorality—a social pest carrying contamination and foulness of every quarter.[26]

Acton's association, shared by many others, of prostitutes with filth, disease, and social contamination would prove enduring and influential. However, during the 1800s prostitution itself was still not classified as a

---

21. HOBSON, *supra* note 16, at 32 (explaining that these laws effectively punished status rather than specific acts and required no proof of solicitation or sale).

22. *Id.* at 34 (citing statistics in which "22.4 percent of male chastity offenders (not including brothelkeepers) were sent to the house of correction in Boston, compared with 71 percent of women"). Indeed, an 1826 court docket from Boston suggests that courts were already concerned with protecting the privacy of gentlemen who used prostitutes. In two cases, the docket book does not list the names of the male defendants who were found guilty of chastity offenses and fined. *Id.* Other punishments included "flogging or embarrassment in public squares." DeCou, *supra* note 20.

23. *See* HOBSON, *supra* note 16, at 56 (explaining that these stories were widely circulated in journals like the *Friend of Virtue*).

24. *Id.* at 111.

25. *Id.* at 110.

26. *Id.* at 111 (quoting WILLIAM ACTON, PROSTITUTION CONSIDERED IN ITS MORAL, SOCIAL, AND SANITARY ASPECTS IN LONDON AND OTHER LARGE CITIES AND GARRISON TOWNS WITH PROPOSALS FOR THE CONTROL AND PREVENTION OF ITS ATTENDANT EVILS 166 (photo. reprint 1972) (2d ed. 1870)).

criminal offense.[27] Antiprostitution fervor took the form of unofficial harassment and "whorehouse riots," violent outbursts directed at brothels.[28]

Beginning in the 1800s, female reform societies also tried, with varying success, to shape the national discourse on prostitution. Women were active in moral-reform movements in the nineteenth century, which worked to make men share the blame for fallen women's loss of virtue, sometimes even naming and shaming high-status men.[29] Many groups of feminists during the Progressive Era advocated for an end to prostitution as a step toward the emancipation of women.[30] These feminists spoke of an "equal moral obligation," shared by men and women, regarding prostitution.[31] They viewed prostitution as one manifestation of the larger problem of male domination of political, economic, and social institutions.[32] Their ideas are shared by many feminists today.[33]

B. Prostitution During the Progressive Era: Criminalization

Reformers in the Progressive Era tweaked the fallen woman narrative to create the metaphor of "white slavery." Like the fallen woman, the victim of white slavery was a young, innocent girl who was tricked or physically coerced into prostitution.[34] White slavery was not the only cause of anxiety for Progressive Era Americans. Rapid urbanization and the expansion of capitalism resulted in a sense of a loss of control over young women and the accompanying fear that more young women would turn to prostitution.[35]

---

27. *See id.* at 32 ("Nightwalking, the offense most prostitutes were charged with, had been part of the general law . . . . No mention of solicitation by prostitutes appeared in the text. Prostitution in effect represented a deviant status and not a specific act." (footnote omitted)).

28. RUTH ROSEN, THE LOST SISTERHOOD: PROSTITUTION IN AMERICA, 1900–1918, at 4 (Johns Hopkins Paperbacks ed. 1983). Although after the Civil War prostitutes benefited from a period of relative tolerance, even this tolerance had its limits. St. Louis's attempt to legalize and regulate prostitution prompted widespread public outrage, which led to the defeat of the city's "social evil ordinance" in 1874. HOBSON, *supra* note 16, at 147.

29. HOBSON, *supra* note 16, at 51–52.

30. *Id.* at 150.

31. *Id.* at 151–52.

32. *Id.* at 152.

33. *See, e.g.*, Alexandra Bongard Stremler, Essay, *Sex for Money and the Morning After: Listening to Women and the Feminist Voice in Prostitution Discourse*, 7 U. FLA. J.L. & PUB. POL'Y 189, 191–92 (1994–1995) (situating the marginalization of feminist perspectives in the modern prostitution debate within the broader tendency to relegate feminist jurisprudence to the sidelines in favor of "universal" traditional male opinion).

34. *See* HOBSON, *supra* note 16, at 142 (describing typical white slave narratives featuring "men with hypodermic needles waiting to drug and abduct their prey in darkened movie theaters or subways"). In addition, the white slavery metaphor reflected growing anxiety with the rise of the pimp system and organized human trafficking. *Id.* at 142–43.

35. Ann M. Lucas, *Race, Class, Gender, and Deviancy: The Criminalization of Prostitution*, 10 BERKELEY WOMEN'S L.J. 47, 51–52 (1995).

According to the fallen woman narrative, which still held sway during the Progressive Era, once a woman had turned to prostitution, she had not only permanently corrupted herself but had also become a source of contamination, a conduit for the spread of venereal disease.[36] The outbreak of World War I brought this concern to the forefront and the government reacted by instituting a policy of venereal-disease control that mandated the testing of suspected prostitutes.[37]

These sources of anxiety caused the American public to unite in an effort to stamp out the "Social Evil."[38] In 1910, Congress passed the Mann Act (also known as the White-Slave Traffic Act),[39] which criminalized, *inter alia*, the interstate transportation of women for the purpose of prostitution.[40] The Mann Act was followed by the Standard Vice Repression Law of 1919, which effectively criminalized all prostitution.[41]

C. *Prostitution Today: State Prostitution Laws*

While the passage of the Mann Act and the Standard Vice Repression Law marked an important development in federal antiprostitution law, the power to prohibit prostitution remained with the states.[42] By 1925 every state had passed some form of antiprostitution law.[43] Some of these states continued the tradition of primarily targeting prostitutes for punishment through facially discriminatory statutes that defined prostitution as a crime committed by women.[44] Attitudes about prostitution were slow to change;[45]

---

36. *See* HOBSON, *supra* note 16, at 110 (discussing the fallen-woman paradigm); Lucas, *supra* note 35, at 54–55 ("Progressives worried not only about . . . declining morality, but about the spread of disease as well.").

37. HOBSON, *supra* note 16, at 170.

38. ROSEN, *supra* note 28, at 38.

39. White-Slave Traffic (Mann) Act, ch. 395, 36 Stat. 825 (1910) (codified as amended at 18 U.S.C. §§ 2421–2424 (2006)).

40. *Id.* In 1986, Congress amended the Mann Act to prohibit the interstate transportation of individuals, thereby expanding its protection to adult men. Child Sexual Abuse and Pornography Act of 1986, Pub. L. No. 99-628, § 5(b)(1), (c), 100 Stat. 3510, 3511 (codified as amended at 18 U.S.C. §§ 2421–2424).

41. DeCou, *supra* note 20, at 432.

42. *See* Hoke v. United States, 227 U.S. 308, 317, 321, 323 (1913) (upholding the constitutionality of the Mann Act while noting that "[t]here is unquestionably a control in the States over the morals of their citizens, and, it may be admitted, it extends to making prostitution a crime"). States may regulate commercial sex under their police power. *Id.* at 321.

43. Whitebread, *supra* note 10, at 243.

44. *See, e.g.*, State v. Devall, 302 So. 2d 909, 910, 913 (La. 1974) (reversing a trial judgment that sustained a motion to quash Louisiana's prostitution statute on equal-protection grounds and observing that "[a]ccording to the terms of the statute, the crime it reprobates can only be committed by a woman"). While all of these early state laws targeted prostitutes, their methods differed. Some states criminalized solicitation, others criminalized the sex act, and still others created a class of "common nightwalkers," comprised of people with a history of prostitution. DeCou, *supra* note 20, at 433. This lack of uniformity carried on into the early 1970s. *Id.* at 434 (indicating that in 1973, forty-four states criminalized solicitation, thirty-eight criminalized

by 1973 at least seven states still had laws that facially discriminated against women by criminalizing only the sale of sex by females.[46] Today, most prostitution statutes define the crime of prostitution in gender-neutral language.[47]

Additionally, many state statutes did not criminalize the act of patronizing a prostitute until the latter part of the twentieth century.[48] Today, the act of patronizing a prostitute is not universally criminalized. Instead, some state statutes punish the prostitute but not the john, some punish both but impose harsher penalties on the prostitute, and some punish the prostitute and the john equally.[49]

State courts differed in their interpretations of state statutes, some holding that only women can be prostitutes and others holding that both genders are capable of committing the crime of prostitution. Early cases tend to reflect the Progressive Era view that prostitution was a crime that could not be committed by a man.[50] This view proved enduring; as late as 1972 the Indiana Supreme Court upheld the constitutionality of one of these

---

commercial sex acts, and thirteen criminalized the status of being a prostitute) (citing RICHARD SYMANSKI, THE IMMORAL LANDSCAPE: FEMALE PROSTITUTION IN WESTERN SOCIETIES 86 (1981)).

45. In 1977, a New York family court noted that, although New York amended its penal law to define the crime of prostitution in gender-neutral language and criminalized the act of patronizing a prostitute in the 1960s, the "historical sex bias" of New York prostitution laws had endured. *In re* P., 400 N.Y.S.2d 455, 460 (N.Y. Fam. Ct. 1977).

46. DeCou, *supra* note 20, at 434 n.70 (listing Alaska, Indiana, Louisiana, North Dakota, Utah, Wisconsin, and Wyoming).

47. *See* Lefler, *supra* note 19, at 19 (observing in her note, published in 1999, that "most statutes are now gender neutral or at least provide some punishment for johns"). For examples of state prostitution statutes that employ gender-neutral language, see CAL. PENAL CODE § 647(b) (West Supp. 2014), stating, "A person agrees to engage in an act of prostitution when, with specific intent to so engage, he or she manifests an acceptance of an offer or solicitation to so engage"; FLA. STAT. ANN. § 796.07(1)(a) (West Supp. 2014), defining prostitution as "the giving or receiving of the body for sexual activity for hire"; N.Y. PENAL LAW § 230.00 (McKinney 2008), declaring that "a person is guilty of prostitution when such person engages or agrees or offers to engage in sexual conduct with another person in return for a fee"; and VT. STAT. ANN. tit. 13, § 2631 (2009), defining prostitution as "the offering or receiving of the body for sexual intercourse for hire." The Mann Act itself was amended in 1986 to incorporate gender-neutral language. *See* Child Sexual Abuse and Pornography Act of 1986, Pub. L. No. 99-628, § 5(b)(1), (c), 100 Stat. 3510, 3511 (codified as amended at 18 U.S.C. §§ 2421–2424 (2006)) (replacing the terms "female" and "woman or girl" with "individual").

48. Lefler, *supra* note 19, at 16 & n.45.

49. *Id.* at 16–17.

50. *See, e.g.*, *Ex parte* Carey, 207 P. 271, 275 (Cal. Dist. Ct. App. 1922) ("But a man can no more commit the offense of soliciting for prostitution than that of carrying on the business of prostitution. . . . The words 'soliciting for prostitution' have a well-understood and distinct meaning. They are held to mean the act of a fallen woman in hailing passers-by . . . ."); State v. Gardner, 156 N.W. 747, 749 (Iowa 1916) ("[T]he statute in question does not contemplate that a man can be a prostitute or can practice prostitution . . . .").

facially discriminatory statutes.[51] Two dissenting judges in the South African case *State v. Jordan*,[52] decided in 2002, articulate the enduring and widespread belief that prostitutes alone bear the criminal stain of exchanging sex for money:

> The female prostitute has been the social outcast, the male patron has been accepted or ignored. She is visible and denounced, her existence tainted by her activity. He is faceless, a mere ingredient in her offence rather than a criminal in his own right, who returns to respectability after the encounter.... Thus, a man visiting a prostitute is not considered by many to have acted in a morally reprehensible fashion. A woman who is a prostitute is considered by most to be beyond the pale. The difference in social stigma tracks a pattern of applying different standards to the sexuality of men and women.[53]

However, as the twentieth century progressed some state courts adopted a more equitable view. Some of these courts excised from their prostitution statutes discriminatory language that defined prostitution as a crime that can only be committed by a woman.[54] At least one other court held a facially discriminatory statute to be unconstitutionally vague and overbroad.[55] Unfortunately, the trend toward more inclusive prostitution statutes has been undercut by the manner in which these statutes are enforced, which ensures that prostitutes are still arrested at higher rates than johns.

IV. The Problem of Discriminatory Enforcement

The progress made by moving from facially discriminatory prostitution statutes to those that employ gender-neutral language is undercut by the discriminatory enforcement of the law by police and prosecutors. Statistics published by the Federal Bureau of Investigation (FBI) reveal that women are arrested at roughly twice the rate as men for

---

51. Wilson v. State, 278 N.E.2d 569, 570–71 (Ind. 1972). The statute in question defined a prostitute as "[a]ny female who frequents or lives in a house or houses of ill fame, knowing the same to be a house of ill fame, or who commits or offers to commit one or more acts of sexual intercourse or sodomy for hire." *Id.* at 571 (DeBruler, J., dissenting) (quoting IND. CODE § 35-30-1-1 (1971)).

52. 2002 (6) SA 642 (CC) (S. Afr.).

53. *Id.* at 667–68 para. 64 (O'Regan & Sachs, JJ., dissenting).

54. *See, e.g.*, Plas v. State, 598 P.2d 966, 967–69 (Alaska 1979) (striking the phrase "by a female" from an Alaska statute that defined prostitution as "the giving or receiving of the body by a female for sexual intercourse for hire").

55. *See, e.g.*, Holloway v. City of Birmingham, 317 So. 2d 535, 536, 540 (Ala. Crim. App. 1975) (finding a Birmingham ordinance, which stated, "No female shall prostitute herself or use any indecent or lascivious language, gestures or behavior to induce any other person to illicit sexual intercourse," to be vague and overbroad).

the crime of prostitution and commercialized vice.[56] In 2012,[57] women comprised 67.7% of those arrested for prostitution and commercialized vice.[58] This female-to-male ratio of arrestees appears to be typical; each year from 2004 until 2011 women represented between 64.2% and 69.6% of arrestees while men represented between 30.4% and 35.8%.[59] The ratio of two female arrestees for every one male arrestee reflects only a 10%–15% increase in arrest rates for males relative to females from 1970.[60]

The ratio of female to male arrestees also varies from state to state and from year to year. Some states have achieved a rough parity between male and female arrestees in some years[61] and other states have displayed

---

56. The FBI defines the offense of prostitution and commercialized vice as follows:

> The unlawful promotion of or participation in sexual activities for profit, including attempts. To solicit customers or transport persons for prostitution purposes; to own, manage, or operate a dwelling or other establishment for the purpose of providing a place where prostitution is performed; or to otherwise assist or promote prostitution.

FED. BUREAU OF INVESTIGATION, CRIME IN THE UNITED STATES 2012: OFFENSE DEFINITIONS (2013).

57. The last year for which data was available at the time of publication.

58. *Crime in the United States 2012: Table 42*, FED. BUREAU OF INVESTIGATION, http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/tables/42tabledatadecoverviewpdf.

59. *See id.* (listing female arrestees at 67.7% of total arrestees and male arrestees at 32.3% of total arrestees); *Crime in the United States 2011: Table 42*, FED. BUREAU OF INVESTIGATION, http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/tables/table-42 (68.8% female and 31.2% male); *Crime in the United States 2010: Table 42*, FED. BUREAU OF INVESTIGATION, http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/tables/10tbl42.xls (68.7% female and 31.3% male); *Crime in the United States 2009: Table 42*, FED. BUREAU OF INVESTIGATION (Sept. 2010), http://www2.fbi.gov/ucr/cius2009/data/table_42.html (69.6% female and 30.4% male); *Crime in the United States 2008: Table 42*, FED. BUREAU OF INVESTIGATION (Sept. 2009), http://www2.fbi.gov/ucr/cius2008/data/table_42.html (69.4% female and 30.6% male); *Crime in the United States 2007: Table 42*, FED. BUREAU OF INVESTIGATION (Sept. 2008), http://www2.fbi.gov/ucr/cius2007/data/table_42.html (68.1% female and 31.9% male); *Crime in the United States 2006: Table 42*, FED. BUREAU OF INVESTIGATION (Sept. 2007), http://www2.fbi.gov/ucr/cius2006/data/table_42.html (64.2% female and 35.8% male); *Crime in the United States 2005: Table 42*, FED. BUREAU OF INVESTIGATION (Sept. 2006), http://www2.fbi.gov/ucr/05cius/data/table_42.html (66.4% female and 33.6% male); *Crime in the United States 2004: Table 42*, FED. BUREAU OF INVESTIGATION (Feb. 2006), http://www2.fbi.gov/ucr/cius_04/persons_arrested/table_38-43.html#table42 (69.1% female and 30.9% male).

60. In 1970, 20.7% of arrestees were men. Lefler, *supra* note 19, at 19–20.

61. For example, the state of Wisconsin arrested more men than women for prostitution in 2011 and 2010. *See* WIS. OFFICE OF JUSTICE ASSISTANCE, ARRESTS IN WISCONSIN 2011, at 309 tbl.19 (2012), *available at* https://wilenet.org/html/justice-programs/programs/justice-stats/library/crime-and-arrest/2011-arrests-in-wisconsin.pdf (comparing 238 men arrested to 231 women arrested); WIS. OFFICE OF JUSTICE ASSISTANCE, ARRESTS IN WISCONSIN 2010, at 308 tbl.19 (2011), *available at* https://wilenet.org/html/justice-programs/programs/justice-stats/library/crime-and-arrest/2010-arrests-in-wisconsin.pdf (comparing 203 men arrested to 177 women arrested). However, in both 2008 and 2009 more women than men were arrested in the same state. *See* WIS. OFFICE OF JUSTICE ASSISTANCE, ARRESTS IN WISCONSIN 2009, at 276 tbl.19 (2010), *available at* https://wilenet.org/html/justice-programs/programs/justice-stats/library/crime-and-arrest/2009-arrests-in-wisconsin.pdf (comparing 228 women arrested to 148 men arrested); WIS. OFFICE OF

inequitable arrest rates for many years in a row.[62] Statistics cited by defendants on trial for prostitution show similar patterns of unequal enforcement. In *City of Yakima v. Johnson*,[63] the defendant offered statistical evidence showing that, for ten years, almost every person arrested for violating an ordinance prohibiting prostitution was a woman.[64] Similarly, in *In re Elizabeth G.*,[65] the defendant offered evidence that the percentage of people arrested for violating a solicitation statute who were men ranged between 1.8% (1974) and 27.3% (1975).[66] Lastly, one court in New York noted certain statistics showing that "the overwhelming number of people arrested for prostitution-related offenses in Buffalo are female (the number is at least 70% and probably closer to 90%)."[67]

Some police officers have provided surprisingly candid testimony that sheds light on one cause of the discrepancy between the numbers of male and female arrestees. In one case, a police detective testified that it was department policy not to arrest the man and that "[t]here hasn't been a male arrested . . . since we've been working on the prostitution area."[68] The detective further testified that even when a john was present and subject to arrest, such as when police stop a john's car after the prostitute gets in, it was the "general policy that you don't arrest the male."[69] The detective gave two reasons for the policy: the women were arrested because complaints from people in the area related "mainly [to] the girls" and because the women were known to the police.[70]

Another explanation for high female arrest rates is the openly acknowledged police strategy of using male decoys.[71] In a typical scenario,

---

JUSTICE ASSISTANCE, ARRESTS IN WISCONSIN 2008, at 270 tbl.19 (2009), *available at* https://wilenet.org/html/justice-programs/programs/justice-stats/library/crime-and-arrest/2008-arrests-in-wisconsin.pdf (comparing 430 women arrested to 257 men arrested).

62. For example, in California women comprised 67.4% of total arrestees in 2012, 72.3% of total arrestees in 2011, and 70.9% of total arrestees in 2010. *See* KAMALA D. HARRIS, CAL. DEP'T OF JUSTICE, CRIME IN CALIFORNIA 2012, at 42 tbl.34 (2013), *available at* http://oag.ca.gov/sites/all/files/agweb/pdfs/cjsc/publications/candd/cd12/cd12.pdf?; KAMALA D. HARRIS, CAL. DEP'T OF JUSTICE, CRIME IN CALIFORNIA 2011, at 42 tbl.34 (2012), *available at* http://oag.ca.gov/sites/all/files/agweb/pdfs/cjsc/publications/candd/cd11/cd11.pdf?; KAMALA D. HARRIS, CAL. DEP'T OF JUSTICE, CRIME IN CALIFORNIA 2010, at 42 tbl.34 (2011), *available at* http://oag.ca.gov/sites/all/files/agweb/pdfs/cjsc/publications/candd/cd10/preface.pdf?.

63. 553 P.2d 1104 (Wash. Ct. App. 1976).

64. *Id.* at 1106 & n.3.

65. 126 Cal. Rptr. 118 (Cal. Ct. App. 1975).

66. *Id.* at 119–20.

67. People v. Burton, 432 N.Y.S.2d 312, 314 (Buffalo City Ct. 1980).

68. Commonwealth v. An Unnamed Defendant, 492 N.E.2d 1184, 1186 (Mass. App. Ct. 1986) (omission in original).

69. *Id.*

70. *Id.* (alteration in original).

71. The use of male decoys is widespread and can result in many female arrestees. *See* Minouche Kandel, *Whores in Court: Judicial Processing of Prostitutes in the Boston Municipal Court in 1990*, 4 YALE J.L. & FEMINISM 329, 334–35 (1992) (observing that "[t]he vast majority

an undercover police officer will wait until he is approached by a suspected prostitute, allow her to solicit him, and arrest her soon after.[72] By posing as johns, police officers ensure that only prostitutes are arrested. Courts have been generally receptive to police testimony that using officers as decoys is an effective means of combatting prostitution.[73] In contrast, courts have been quick to agree with police departments that the use of female decoys to arrest johns would not be feasible. In *United States v. Wilson*,[74] the court rejected a defense of discriminatory enforcement based in part on uncontroverted police testimony that the use of female decoys was "infeasible" because of the "entrapment aspect."[75]

By according police departments and prosecutors a high rate of deference in allocating scarce resources, many courts turn a blind eye to evidence of discriminatory enforcement. The court in *People v. Burton*[76] concluded that there was no discriminatory enforcement of New York's prostitution law, even though "the overwhelming number of people arrested for prostitution-related offenses in Buffalo are female," because "a comprehensive view satisfies this court that there are good and sufficient reasons . . . to justify the police on concentrating on the arrest of female prostitutes."[77] These reasons included a shortage of manpower and, tellingly, a dearth of women on the police force who could pose as prostitutes because "the police department is guilty of engaging in discriminatory hiring practices of women in the past."[78] Additionally, testimony from officers on the Buffalo Police Department Vice Squad indicated that female decoys were more expensive because they required additional manpower to protect them from prostitutes and from "panicky Johns."[79] The *Burton* court accepted these arguments but noted that they "will not suffice tomorrow when there are many more women available in the police ranks."[80] Other courts have similarly noted that prosecutors[81] and

---

of prostitution arrests are made through the use of police decoys" and noting that "[s]ince most of the police [decoys] are male, those arrested are generally women and a few male prostitutes").

72. *See* City of Minneapolis v. Buschette, 240 N.W.2d 500, 501–02 (Minn. 1976) (summarizing the testimony of officers on the Minneapolis morals squad describing a typical "decoy" operation).

73. *See, e.g.*, *id.* at 501–02, 504 (citing police testimony regarding the department's use of decoys in its "important duty" of prostitution enforcement and disagreeing with the defendant's assertion that "the all-male composition of the morals squad acting as plainclothes decoys or 'tricks' is per se discriminatory").

74. 342 A.2d 27 (D.C. 1975).

75. *Id.* at 29–31. According to an officer who testified at trial, street protocol dictated that the prostitute approach the john and ask if he is "sporting"; correspondingly, if a female decoy were to pose this question to a john, he could raise an entrapment defense. *Id.* at 29.

76. 432 N.Y.S.2d 312 (Buffalo City Ct. 1980).

77. *Id.* at 314–15.

78. *Id.* at 315.

79. *Id.* (internal quotation marks omitted).

80. *Id.*

police departments[82] have great discretion in deciding whom to prosecute. By dismissing evidence of discriminatory enforcement of prostitution statutes, these courts essentially grant police departments permission to continue to arrest more prostitutes than johns.

In conclusion, statistics and police testimony both provide persuasive evidence of pervasive, widespread, and continuing discriminatory enforcement of prostitution statutes. However, judicial deference to the practices of police departments and prosecutors often means that discriminatory practices, such as the exclusive use of male decoys, are held up in court. This is, unfortunately, the state of the law today.

V. The Discriminatory-Enforcement Defense

Before laying out this Note's proposal, a little background is in order. Subpart V(A) briefly describes the origin of the discriminatory-enforcement defense. Subpart V(B) outlines the positive change that would become possible if courts were to recognize a discriminatory-enforcement defense to prostitution charges. Lastly, subpart V(C) describes the three elements of a discriminatory-enforcement defense.

A. Background: Yick Wo v. Hopkins

In the landmark case *Yick Wo v. Hopkins*,[83] the Supreme Court held that the Equal Protection Clause of the Fourteenth Amendment[84] proscribes the discriminatory enforcement of a facially neutral law.[85] In a passage quoted by many courts a century later, the Court stated:

> Though the law itself be fair on its face and impartial in appearance, yet, if it is applied and administered by public authority with an evil eye and an unequal hand, so as practically to make unjust and illegal discriminations between persons in similar circumstances, material to their rights, the denial of equal justice is still within the prohibition of the Constitution.[86]

In summary, *Yick Wo* prohibits state actors from wielding a facially neutral law as if it were discriminatory on its face by selectively enforcing it

---

81. Young v. State, 446 N.E.2d 624, 626 (Ind. Ct. App. 1983); Commonwealth v. King, 372 N.E.2d 196, 205 (Mass. 1977); State v. Johnson, 246 N.W.2d 503, 506 (Wis. 1976).

82. United States v. Wilson, 342 A.2d 27, 30 (D.C. 1975).

83. 118 U.S. 356 (1886).

84. U.S. CONST. amend. XIV, § 1 ("No State shall . . . deny to any person within its jurisdiction the equal protection of the laws."). The Fifth Amendment's Due Process Clause applies these constraints to the federal government. *See* Bolling v. Sharpe, 347 U.S. 497, 499–500 (1954) (observing that "the concepts of equal protection and due process, both stemming from our American ideal of fairness, are not mutually exclusive" and holding that racial segregation in Washington, D.C. public schools violated the Fifth Amendment).

85. *Yick Wo*, 118 U.S. at 373–74.

86. *Id.*

against a certain class of people.[87] Without this prohibition, laws that criminalized gambling, for example, could be enforced in such a way as to only criminalize gambling by African-Americans.[88] Instead, *Yick Wo* instructs that state actors must treat similarly situated people alike.[89]

B. *The Function and Effects of the Defense of Discriminatory Enforcement*

While *Yick Wo* involved the narrower issue of the unequal application of two city ordinances,[90] the Equal Protection Clause of the Fourteenth Amendment applies to "every form of state action, whether legislative, executive, or judicial."[91] In line with this principle, the Supreme Court later indicated that discriminatory enforcement is also a defense to a criminal charge.[92] If the criminal defendant successfully establishes this defense, the court must dismiss the case.[93]

The severe consequence of raising this defense successfully is what makes the remedy so effective.[94] As the Minnesota Supreme Court noted,

---

87. *Id.*

88. This was the situation described in the California case *People v. Harris*, 5 Cal. Rptr. 852 (Cal. App. Dep't Super. Ct. 1960). Julie Lefler raises a similar point in the context of sex-based classifications, stating, "If there is no equality in enforcement it does not matter that laws are gender-neutral." Lefler, *supra* note 19, at 34.

89. For a discussion of who is similarly situated to whom in the context of prostitution cases, see *infra* subpart VI(A).

90. 118 U.S. at 357–58.

91. Comment, *The Right to Nondiscriminatory Enforcement of State Penal Laws*, 61 COLUM. L. REV. 1103, 1105 (1961) [hereinafter *Nondiscriminatory Enforcement*]; *see also Ex Parte* Virginia, 100 U.S. 339, 347 (1880) ("A State acts by its legislative, its executive, or its judicial authorities. . . . The constitutional provision, therefore, must mean that no agency of the State, or of [its] officers or agents . . . , shall deny to any person within its jurisdiction the equal protection of the laws.").

92. *See* Oyler v. Boles, 368 U.S. 448, 456 (1962) (rejecting a selective enforcement defense in a criminal case but leaving open the possibility that such a defense could be used if properly supported); Two Guys from Harrison-Allentown, Inc. v. McGinley, 366 U.S. 582, 583–84, 588–89 (1961) (affirming a lower court decision to deny an injunction restraining the enforcement of a Sunday closing law that was discriminatorily enforced because the plaintiff "may defend against any such proceeding that is actually prosecuted on the ground of unconstitutional discrimination"); *Nondiscriminatory Enforcement*, *supra* note 91, at 1108 (discussing two cases where the Supreme Court has indicated that intentional and deliberate discriminatory state penal enforcement would violate the Equal Protection Clause).

93. *E.g.*, Commonwealth v. An Unnamed Defendant, 492 N.E.2d 1184, 1187–88 (Mass. App. Ct. 1986); State v. McCollum, 464 N.W.2d 44, 52 (Wis. Ct. App. 1990); *cf.* Andrew B. Weissman, *The Discriminatory Application of Penal Laws By State Judicial and Quasi-Judicial Officers: Playing the Shell Game of Rights and Remedies*, 69 NW. U. L. REV. 489, 497 (1974) (explaining that "the fundamental premise that lies at the root of any grant of relief to a culpable person [is] that those clearly guilty should nonetheless be permitted to avoid punishment when the prosecution results from illegal discrimination by law enforcement officials").

94. For sources recognizing the severe nature of this defense, see *City of Minneapolis v. Buschette*, 240 N.W.2d 500, 504 (Minn. 1976) and Weissman, *supra* note 93. As Weissman observes, "Arguments opposing such a grant of relief are simple and to the point: it is highly

"While dismissal may be an extreme remedy, especially when the guilt of certain defendants may seem clear, it is only by this means that the courts will put an end to the arrest and prosecution of persons based on intentional and purposeful discrimination."[95] The threat of having cases thrown out is, indeed, a powerful incentive to change the manner in which an arrest is made. In *United States v. Wilson*, a police lieutenant explained why the department had discontinued the use of female decoys to arrest johns: "[M]ales looking for prostitutes were arrested on the charge of soliciting for prostitution; but *the cases 'were all dismissed* because [of] the entrapment aspect. And [*the detective had*] *never had a conviction of a male subject for soliciting a policewoman that* [*he knew*] *of.*'"[96] It would appear that police officers will not waste their time and energy making arrests when the case against the arrestee will be subsequently thrown out.

That the dismissal of several cases has the potential to make such a noticeable impact on the practices of a police department holds important implications for the defense of discriminatory enforcement. It means that a successful defense has a real likelihood of achieving the larger purpose of acting as "one of the few means the individual citizen has to force public officials to do their job properly."[97] It represents, in effect, the results that are possible when courts work in tandem with the people to send a message to public officials. If courts were to throw out prostitution cases accompanied by evidence of discriminatory enforcement against women, the message to public officials would be clear: the lesson of *Yick Wo* still applies; the laws of the United States must be applied in an equal manner.

The importance of this defense is underscored in prostitution cases because prostitutes as a group—especially the street prostitutes who comprise the majority of defendants who attempt to bring this defense—do not, on average, have much political or economic power.[98] Street prostitutes may be battered by pimps; they may be teenage runaways; they may feel the shame heaped on them by a society that considers them to be

---

unusual to allow a known culprit to go free after an admitted violation of a valid criminal statute." Weissman, *supra* note 93.

95. *Buschette*, 240 N.W.2d at 504.

96. 342 A.2d 27, 29 (D.C. 1975) (emphasis added).

97. People v. Superior Court of Alameda Cnty., 562 P.2d 1315, 1325 (Cal. 1977) (Tobriner, C.J., dissenting) (quoting People v. Gray, 63 Cal. Rptr. 211, 217 (Cal. Ct. App. 1967)); *see also* Weissman, *supra* note 93, at 498 (stating that several courts in favor of judicial relief make an analogy to the exclusionary rule of *Mapp v. Ohio*, 367 U.S. 643 (1961), on the basis that "allowing a defense of discriminatory enforcement is the only practical way to force public officials to undertake their responsibilities in a proper manner").

98. For a description of the many forms of violence suffered by prostitutes working both indoors and on the street, see Melissa Farley et al., *Prostitution and Trafficking in Nine Countries: An Update on Violence and Posttraumatic Stress Disorder*, 2 J. TRAUMA PRAC. 33, 34–37 (2003), and MacKinnon, *supra* note 10, at 282–88. Indeed, Farley et al. describe the alarming prevalence of rape, childhood sexual abuse, verbal abuse, stalking, battering, torture, intimate partner violence, and posttraumatic stress disorder reported by prostituted women. Farley et al., *supra*.

the embodiment of everything vile; they may have been trafficked; they may have lost the ability to insist on a condom, to insist on freedom from violence, or to choose whether to be a prostitute in the first place. But the defense of discriminatory enforcement remains open to them and gives them what may be their only and greatest opportunity to force the police officers who arrest them and the prosecutors who bring their cases into the courts to perform their jobs as representatives of the state with at least a basic level of fairness.[99]

While this Note takes the position that courts should recognize the defense of discriminatory enforcement in prostitution prosecutions, it also recognizes that courts have historically been reluctant to do so. Thus, it may be wise to make a bifurcated effort that focuses on police departments as well as courts. A recent report provided to the United States Department of Justice noted the effectiveness of interventions that focus on reducing the demand for commercial sex.[100] The report cites various demand-centered approaches that have been put into place by police departments, which include, *inter alia*, founding "john schools," instituting reverse stings, publicizing the identities of johns, sending "Dear John" letters, seizing automobiles used to solicit sex, and suspending driver's licenses.[101] These programs show that progressive police departments may be more receptive to focusing on johns. Activists may be able to use the fact that several police departments across the country have adopted this approach to persuade police departments in their area to consider adopting a similar approach.

However, judges must still be committed to the equal enforcement of prostitution laws and to equalizing punishment. In particular, judges should refrain from imposing a prison sentence on a prostitute while sending her

---

99. The low status of prostitution, which translates to a lack of political power, is one way in which discriminatory law enforcement against prostitutes becomes more similar to the situation in *Mapp*. For prostitutes, there very well may not be any other methods to combat discriminatory enforcement, beyond judicial relief, that comport better with our political process.

100. MICHAEL SHIVELY ET AL., ABT ASSOCIATES, INC., A NATIONAL OVERVIEW OF PROSTITUTION AND SEX TRAFFICKING DEMAND REDUCTION EFFORTS, at v (2012), *available at* https://www.ncjrs.gov/pdffiles1/nij/grants/238796.pdf ("Evidence that anti-demand tactics . . . can effectively suppress commercial sex markets is slowly accumulating and is robust in relation to evidence of the effectiveness of other approaches.").

101. *Id.* at 21–24. "Dear John" letters are a method of deterrence in which "[l]etters are sent to addresses of registered car owners, alerting owners that their car was seen in [an] area known for prostitution, and warning them about legal and other consequences of engaging in prostitution." *Id.* at 22. Johns who are arrested for soliciting commercial sex may also be sent to "john schools," education programs that "cover a range of topics designed to persuade or deter men from buying sex." *Id.* at 61. Topics may include the legal and health consequences of engaging in commercial sex and the harm that prostitution inflicts on prostituted women and girls, and discussion may encompass "healthy relationships, anger management, sexual addiction, pimping and pandering, human trafficking, and johns' vulnerability to criminal victimization while engaged in commercial sex." *Id.* at 62.

client off to john school.[102] Doing so would only recall the gross disparities in sentencing between prostitutes and johns that marked the nineteenth and twentieth centuries.[103]

C. *Elements of the Defense*

Many state courts have endorsed the view that discriminatory enforcement can be a defense to a criminal charge for prostitution.[104] While there is some disagreement on when a claim of discriminatory enforcement may be raised, courts generally agree that the claim comprises three elements, all of which must be proven by the defendant.

First, the defendant must show that there was selectivity in enforcement. This normally entails a demonstration of the relevant population of violators—those similarly situated to the defendant—and proof that not all of the violators are prosecuted.[105] Defendants may normally establish this first element by direct evidence that other individual violators were not prosecuted or by statistical evidence indicating few or no prosecutions of other violators of the statute.[106]

Second, the defendant must show that the selectivity in enforcement was intentional and not simply due to mistake or laxity in enforcement.[107] Some courts phrase this element as requiring proof of intentional or purposeful discrimination.[108] This element is in direct tension with the idea

---

102. Indeed, this is what happened in *Salaiscooper v. Eighth Judicial District Court ex rel. County of Clark*, 34 P.3d 509 (Nev. 2001) (per curiam). In this case, the court upheld the constitutionality of the First Offender Program for Men in Las Vegas, a diversion program available to men, and ostensibly women, who were charged with buying sex. *Id.* at 512–13. The State admitted that "the vast majority of sellers of sex are females" and that buyers of sex are "statistically almost always male" but nonetheless argued that the program did not run afoul of the Constitution because it was not long enough or comprehensive enough to rehabilitate prostitutes. *Id.* at 513.

103. *See supra* Part III.

104. *See, e.g.*, People v. Superior Court of Alameda Cnty., 562 P.2d 1315, 1319–23 (Cal. 1977) (recognizing the defense of discriminatory enforcement but finding no constitutional infirmity in the police department's prostitution-arrest policies); State v. McCollum, 464 N.W.2d 44, 51–52 (Wis. Ct. App. 1990) (affirming the dismissal of prostitution charges on the ground of discriminatory enforcement). While many state courts in the 1970s encountered cases in which alleged prostitutes attempted to raise a defense of discriminatory enforcement, the rate at which the defense was raised declined during the 1980s and beyond. This Note attempts to cull representative language from the cases that exist and lay out the issues that were most common among the cases.

105. *See, e.g.*, United States v. Armstrong, 517 U.S. 456, 465 (1996) ("To establish a discriminatory effect in a race case, the claimant must show that similarly situated individuals of a different race were not prosecuted."); Ah Sin v. Wittman, 198 U.S. 500, 507–08 (1905) (requiring the defendant to establish that other violators of the law are not generally prosecuted).

106. Stefan H. Krieger, Comment, *Defense Access to Evidence of Discriminatory Prosecution*, 1974 U. ILL. L.F. 648, 654.

107. Weissman, *supra* note 93, at 502–05.

108. *See, e.g.*, Snowden v. Hughes, 321 U.S. 1, 8 (1944) ("The unlawful administration by state officers of a state statute fair on its face, resulting in its unequal application to those who are

that courts should afford police departments and prosecutors discretion in deciding whom to prosecute.[109]

Third, the defendant must show that the selective enforcement was based on an invidious or unjustifiable standard.[110] While certain criteria for enforcement are clearly invidious, such as race, other criteria, such as sex, are less so.[111] In cases in which the element of invidiousness is uncertain, as often occurs in cases of prostitution, the decision may be based upon the strength of the first two elements.[112]

VI. Barriers to a Finding of Discriminatory Enforcement

While the defendant already bears a heavy burden of proof on the elements of a discriminatory-enforcement defense, courts have set up two additional barriers to proving the defense of discriminatory enforcement. First, some courts take the position that prostitutes and johns are not similarly situated and thus the treatment of one cannot be compared to the treatment of the other. Second, some courts require a high standard of proof to establish discriminatory enforcement. This operates to make proving the second and third elements, which often collapse into each other, more difficult. Third, some courts adopt sexist justifications for the continued discriminatory enforcement of prostitution laws.

---

entitled to be treated alike, is not a denial of equal protection unless there is shown to be present in it an element of intentional or purposeful discrimination."). Many state courts follow this standard. *See, e.g.*, Blake v. State, 344 A.2d 260, 263 (Del. Super. Ct. 1975) (clarifying that, in order to establish "intentional and purposeful discrimination[,] . . . proof that only women have been prosecuted . . . shall not be enough"); State v. Olson, 297 N.W.2d 297, 298 (Minn. 1980) (holding that the defendant did not establish "conscious or purposeful intent to discriminate"); City of Spokane v. Hjort, 569 P.2d 1230, 1231 (Wash. Ct. App. 1977) (holding that the record did not establish "intentional, purposeful, or systematic discrimination against women").

109. Thus, mere failure to enforce the law against others is not enough to establish a claim of discriminatory enforcement. Oyler v. Boles, 368 U.S. 448, 456 (1962).

110. *E.g.*, People v. Superior Court of Alameda Cnty., 562 P.2d 1315, 1320 (Cal. 1977) ("To establish the defense, the defendant must prove: (1) that he has been deliberately singled out for prosecution on the basis of some invidious criterion . . . ." (internal quotation marks omitted)); State v. Devall, 302 So. 2d 909, 912–13 (La. 1974) ("[A]bsent a showing that distinctions involving prostitution are merely pretexts designed to effect an invidious discrimination against the members of one sex . . . , lawmakers are constitutionally free to exclude male prostitution from the coverage of legislation . . . ."); City of Minneapolis v. Buschette, 240 N.W.2d 500, 506 (Minn. 1976) (holding that the defendant had not met her burden of proof that the state's selective prosecution had been invidious).

111. *See* Kenji Yoshino, *Covering*, 111 YALE L.J. 769, 914–15 (2002) ("The Court's grant of intermediate, rather than strict, scrutiny to sex is sometimes justified on the ground that there are 'real differences' between the sexes. Thus, unlike any two racial groups, men and women are deemed to be biologically different in ways that could justify their differential treatment." (footnotes omitted)).

112. *See, e.g.*, *Superior Court of Alameda Cnty.*, 562 P.2d at 1321 (referring to prostitutes in specifically gender-neutral language as "the prostitute, male or female" and "providers" and characterizing police focus on prostitutes as "a profiteer-oriented approach," thereby avoiding the conclusion that police had selectively enforced the law on the basis of the offender's sex, an invidious standard).

Part VI of this Note analyzes each of these hurdles in turn. Subpart VI(A) begins by examining the classification barrier, arguing that prostitutes and johns are indeed similarly situated. Next, subpart VI(B) urges courts to lower the standard of proof for discriminatory enforcement. Finally, subpart VI(C) argues that courts should discard justifications for the discriminatory enforcement of prostitution laws that are tainted with sexism.

A. Prostitutes and Johns Are Similarly Situated

Some courts that have refused to find discriminatory enforcement in particular cases have done so by taking the position that prostitutes and johns are not similarly situated.[113] This Note takes the position that prostitutes and johns are indeed similarly situated and that courts should compare the rates of arrest and prosecution for each in determining whether prostitution laws were enforced with "an unequal hand."[114] However, the process of determining whether two classes of people are similarly situated is a complicated one that deserves further analysis.

The similarly situated inquiry is an answer to the Equal Protection Clause paradox described in Joseph Tussman and Jacobus tenBroek's influential article *The Equal Protection of the Laws*.[115] This paradox arises from the tension between the states' right to enact laws that classify[116] and

---

113. *See infra* notes 137–39 and accompanying text.

114. Yick Wo v. Hopkins, 118 U.S. 356, 373–74 (1886).

115. Joseph Tussman & Jacobus tenBroek, *The Equal Protection of the Laws*, 37 CALIF. L. REV. 341, 344 (1949). The following discussion is indebted to this article, which, although written over sixty years ago, remains well respected and influential. *See* Deborah Hellman, *Two Types of Discrimination: The Familiar and the Forgotten*, 86 CALIF. L. REV. 315, 343 & n.78 (1998) ("The Tussman and tenBroek article, which ranks among the top 20 law review articles in number of times cited, articulates a conception of the Equal Protection Clause and the harm it is intended to proscribe which largely became the standard conception of the Clause and its motivating principle." (footnote omitted)); Gerald Gunther, *The Suprme Court, 1971 TermForeward: In Search of Evolving Doctrine on a Changing Court: A Model for a Newer Equal Protection*, 86 HARV. L. REV. 1, 8 n.28 (1972) (describing *The Equal Protection of the Laws* as the "classic" pre-Warren Court study of the state of equal-protection doctrine). However, Tussman and tenBroek's approach has not been universally accepted. *See* Catharine A. MacKinnon, *Reflections on Sex Equality Under Law*, 100 YALE L.J. 1281, 1295–97 (1991) (arguing that the similarly situated analysis takes white men as its point of comparison and thus does not capture situations in which sex inequality is at its most extreme); Kenneth W. Simons, *Overinclusion and Underinclusion: A New Model*, 36 UCLA L. REV. 447, 459–60 (1989) (describing the defects in the approach advocated by Tussman and tenBroek).

116. *See* Barbier v. Connolly, 113 U.S. 27, 31–32 (1885) ("Regulations [imposed to achieve general benefits under the state's police power] may press with more or less weight upon one than upon another, but they are designed, not to impose unequal or unnecessary restrictions upon any one, but to promote . . . the general good."); Cass R. Sunstein, *Public Values, Private Interests, and the Equal Protection Clause*, 1982 SUP. CT. REV. 127, 129 (arguing that "classifications are the stuff of legislation; and legislation that classifies does not, solely by virtue of that fact, offend any sense of 'equality'").

the Equal Protection Clause's guarantee of equality before the law.[117] Justice White, writing in *City of Cleburne, Texas v. Cleburne Living Center, Inc.*,[118] interpreted this guarantee as "essentially a direction that all persons similarly situated should be treated alike."[119] This guarantee is cabined by the recognition that "[t]he Constitution does not require things which are different in fact or opinion to be treated in law as though they were the same."[120] The compromise is to allow laws that make reasonable classifications, defined as those that "include[] all persons who are similarly situated with respect to the purpose of the law."[121]

Therefore, the central inquiry in the similarly situated analysis is purpose-oriented: whether "the legislative classification bear[s] a close enough relationship to the purpose of the statute."[122] Under intermediate scrutiny, the legislative classification would have to be substantially related to the purpose of the statute.[123] Another way of phrasing this inquiry is to ask if there is such "a substantial difference between men and women that is relevant to the classification—that is, that the reason for burdening members of one sex does not apply with the same force to members of the other sex."[124] Since *Craig v. Boren*[125] recognized that sex-based classifications are subject to heightened scrutiny,[126] at least two courts that have addressed the issue have applied intermediate scrutiny,[127] and at least one court has applied strict scrutiny.[128] Other courts, in refusing to find

---

117. *Yick Wo*, 118 U.S. at 369 ("[T]he equal protection of the laws is a pledge of the protection of equal laws.").

118. 473 U.S. 432 (1985).

119. *Id.* at 439.

120. Tigner v. Texas, 310 U.S. 141, 147 (1940).

121. Tussman & tenBroek, *supra* note 115, at 346.

122. Giovanna Shay, *Similarly Situated*, 18 GEO. MASON L. REV. 581, 588 & n.60 (2011); *see also* Romer v. Evans, 517 U.S. 620, 632 (1996) ("The search for the link between classification and objective gives substance to the Equal Protection Clause; it provides guidance and discipline for the legislature, which is entitled to know what sorts of laws it can pass; and it marks the limits of our own authority.").

123. Shay, *supra* note 122, at 616.

124. William R. Engles, Comment, *The "Substantial Relation" Question in Gender Discrimination Cases*, 52 U. CHI. L. REV. 149, 151 (1985); *see also id.* at 153–54 (describing this approach as the "'similarly-situated' test").

125. 429 U.S. 190 (1976).

126. *See id.* at 197 ("To withstand constitutional challenge, previous cases establish that classifications by gender must serve important governmental objectives and must be substantially related to achievement of those objectives.").

127. *See, e.g.*, State v. Tookes, 699 P.2d 983, 987–88 (Haw. 1985) (holding that Hawaii's prostitution law is gender-neutral and that "[e]ven if the prohibition were deemed to set up a gender-based classification, it would be invalid only if it did not serve important government objectives and was not substantially related to achieving those objectives"); State v. McCollum, 464 N.W.2d 44, 51 (Wis. Ct. App. 1990) ("When faced with a claim of an equal protection violation based on impermissible gender discrimination, courts must apply an intermediate level of scrutiny to the classification system at issue.").

128. *See, e.g.*, Commonwealth v. King, 372 N.E.2d 196, 206 (Mass. 1977).

discriminatory enforcement, are more oblique about what standard they are applying.[129]

While the above analysis concerns laws that classify on their face, the analysis applies with equal force to facially neutral laws that are enforced in a manner that burdens certain classes more than others.[130] In order to prove discriminatory enforcement, a defendant must first identify a control group of similarly situated persons who have not been prosecuted.[131] By doing so, the defendant "isolate[s] the factor allegedly subject to impermissible discrimination" and reduces the chance that other factors account for the discrepancy in arrest and prosecution rates.[132]

The problem that many prostitutes have encountered in establishing a discriminatory-enforcement defense is that courts often choose male prostitutes—and exclude johns—as the control group to be compared against female prostitutes.[133] With the control group so defined, these courts easily conclude that there has not been discriminatory enforcement.[134] In coming to this conclusion, several courts have observed that male and female prostitutes are treated identically.[135] The courts then explain away the fact that female prostitutes are arrested in greater numbers by observing that female prostitutes far outnumber male prostitutes.[136] By

129. *See, e.g.*, People v. Superior Court of Alameda Cnty., 562 P.2d 1315, 1319, 1321 (Cal. 1977) (noting that sex is an arbitrary classification but refusing to compare prostitutes and johns on the basis of sex, instead characterizing prostitutes as "profiteers" and johns as mere "customers"); Young v. State, 446 N.E.2d 624, 624–26 (Ind. Ct. App. 1983) (finding, because the defendant made no "showing of 'bad faith or evil design,'" no grounds for a discriminatory enforcement instruction and mentioning neither intermediate nor rational basis scrutiny (quoting Highland Sales Corp. v. Vance, 186 N.E.2d 682, 689 (Ind. 1962)).

130. *See* Yick Wo v. Hopkins, 118 U.S. 356, 373–74 (1886).

131. United States v. Aguilar, 883 F.2d 662, 705–06 (9th Cir. 1989); *see also* Barbier v. Connolly, 113 U.S. 27, 31–32 (1885) ("Class legislation, discriminating against some and favoring others, is prohibited, but legislation which . . . affects alike all persons similarly situated, is not within the [Fourteenth] [A]mendment.").

132. *Aguilar*, 883 F.2d at 706.

133. *See, e.g.*, *Superior Court of Alameda Cnty.*, 562 P.2d at 1322; Blake v. State, 344 A.2d 260, 262 (Del. Super. Ct. 1975); State v. Gaither, 224 S.E.2d 378, 379 (Ga. 1976); *id.* at 380 (Hall, J., concurring specially).

134. *See, e.g.*, *Gaither*, 224 S.E.2d at 379.

135. *See Superior Court of Alameda Cnty.*, 562 P.2d at 1322 (stating that the "most obvious ground" for the conclusion that the practice of custodially arresting the prostitute while merely citing the customer is not discriminatory "is that male and female prostitutes are treated alike"); *Blake*, 344 A.2d at 262 (finding that a statute criminalizing only the sale of sex was not subject to strict scrutiny because the seller could be either male or female); People v. Burton, 432 N.Y.S.2d 312, 314 (Buffalo City Ct. 1980) (noting the testimony of a police officer and captain that "male prostitution is as vigorously prosecuted by the Buffalo Police Department and the District Attorney's Office as female prostitution").

136. *See, e.g.*, *Gaither*, 224 S.E.2d at 380 (Hall, J., concurring specially) ("Consequently, the clear fact which emerges from this record, that males are rarely if ever prosecuted for prostitution whereas females are prosecuted in great numbers, cannot be proof of discriminatory enforcement when there is no evidence in the record that male prostitutes exist in detectable numbers."); *Burton*, 432 N.Y.S.2d at 314 ("[T]here are fewer arrests of male prostitutes and their customers

defining the control group as male prostitutes, these courts effectively kill the similarly situated analysis before it can really begin.

Instead of following this track of reasoning, courts should ask whether there is a substantial difference between johns—who are overwhelmingly men—and prostitutes—who are overwhelmingly women—that is relevant to the classification.[137] Assuming that the purpose of prostitution laws is to eliminate prostitution, it seems unlikely that there is a "reason for burdening [prostitutes that] does not apply with the same force to [johns]."[138] However, various state courts have adopted a set of justifications for finding that prostitutes and johns are not similarly situated.

The question of whether prostitutes and johns are similarly situated divides courts, even courts within the same state. In one notable example, two New York courts examined statutes criminalizing prostitution and the act of patronizing a prostitute to determine whether prostitutes and johns were similarly situated. The court in *In re Dora P.*[139] concluded that prostitutes and johns are not similarly situated because they each commit a separate and discrete crime.[140] To support its conclusion, the court emphasized that "separate acts" are necessary to commit prostitution and the act of patronizing a prostitute.[141]

One year later, the court in *People v. Nelson*[142] expressly rejected the conclusion of *In re Dora P.* and concluded that "it is reasonable to combine [prostitutes and johns] in the category of 'others similarly situated.'"[143] In coming to this conclusion, the court stressed the similarity of the two crimes, stating that "the only significant difference in the proscribed behavior is that the prostitute sells sex and the patron buys it" and endorsing

---

simply because they are fewer in number than those individuals connected with female prostitution.").

137. Two judges, vigorously dissenting in a South African case, argued that there are only three, unimportant differences between prostitutes and johns:

> Prostitutes and their customers engage in sexual activity, which is one of the constitutive elements of the relationship between men and women in all societies. As partners in sexual intercourse, they both consent to and participate in the action which lies at the heart of the criminal prohibition. There are only three differences between them. The first is that the one pays and the other is paid. The second is that in general the one is female and the other is male. The third is that the one's actions are rendered criminal by [statute] but the other's actions are not. Moreover, the effect of making the prostitute the primary offender directly reinforces a pattern of sexual stereotyping, which is itself in conflict with the principle of gender equality.

State v. Jordan 2002 (6) SA 642 (CC) at 666 para. 60 (O'Regan & Sachs, JJ., dissenting).

138. Engles, *supra* note 124, at 151.

139. 418 N.Y.S.2d 597 (App. Div. 1979).

140. *Id.* at 604.

141. *Id.*

142. 427 N.Y.S.2d 194 (Syracuse City Ct. 1980).

143. *Id.* at 197–98.

a characterization of the two crimes as "reciprocal offenses."[144] It also cited changes made to New York's Penal Law that equalized sanctions for the two crimes.[145] Lastly, it mentioned that "at a time when prostitution was a phase of the former vagrancy statute, the New York courts were divided on whether it embraced both the patron and the prostitute."[146] The court took this as historical evidence that "some courts have considered prostitution and patronage two facets of a single offense."[147]

The logic that led the *Nelson* court to conclude that prostitutes and johns are similarly situated can be applied beyond New York. As in the early cases in New York, courts in California[148] and Iowa[149] questioned whether prostitution statutes could be applied to men. While both courts concluded that prostitution could only be committed by a woman,[150] they both provide evidence that some people made the opposite argument in the early decades of the twentieth century. As the century progressed and attitudes about women evolved, more and more states amended their prostitution statutes, deleting discriminatory language and expressly criminalizing patronization of a prostitute.[151] This trend toward treating the two crimes equally represents a delayed recognition that they are, as the *Nelson* court stated, reciprocal offenses.

The specific fact patterns common to many cases also support the conclusion that prostitutes and clients are similarly situated. In one common scenario, police officers survey an area known to be frequented by prostitutes and their clients.[152] When they see a woman get into a car with a man and drive away, the officers follow the car, stop it, and ask the man what he is doing with the woman.[153] The man then admits that he had been offered sex for a fee and the officers arrest the woman.[154] In other cases, the officer may catch the prostitute and client engaged in the sex act and only arrest the woman.[155] A third type of case presents a slightly different fact pattern. In *State v. McCollum*,[156] four undercover police officers

144. *Id.* at 197.

145. *Id.*

146. *Id.* (citation omitted).

147. *Id.*

148. *Ex parte* Carey, 207 P. 271, 273–74 (Cal. Dist. Ct. App. 1922).

149. State v. Gardner, 156 N.W. 747, 749 (Iowa 1916).

150. *Carey*, 207 P. at 274; *Gardner*, 156 N.W. at 749–50.

151. *See* sources cited *supra* note 47.

152. *See* DeCou, *supra* note 20, at 436 ("The 'street sweep' has been a favorite method of temporarily clearing streets of prostitutes in a particular area.").

153. *E.g.*, People v. Burton, 432 N.Y.S.2d 312, 314 (Buffalo City Ct. 1980); People v. Nelson, 427 N.Y.S.2d 194, 195 (Syracuse City Ct. 1980).

154. *E.g.*, *Burton*, 432 N.Y.S.2d at 314; *Nelson*, 427 N.Y.S.2d at 195.

155. *E.g.*, Commonwealth v. An Unnamed Defendant, 492 N.E.2d 1184, 1186 (Mass. App. Ct. 1986).

156. 464 N.W.2d 44 (Wis. Ct. App. 1990).

attended a private party at a club and witnessed numerous male patrons "fondle[] the women performers and thrust money at them."[157] When the club closed, the officers arrested the female dancers and made no attempt to arrest any of the male patrons who were present.[158] In each of these fact patterns, the woman charged with prostitution and her client are similarly situated: they have agreed to an exchange of money for sex, or they are engaged in a sex act for which the man has paid or agreed to pay. They are committing statutorily similar crimes—indeed, in some states they are violating the same statute at the same time. When courts focus on female prostitutes to the exclusion of johns, they unnecessarily and harmfully restrict the control group, which in turn virtually ensures the failure of a discriminatory-enforcement defense. By looking for discriminatory enforcement in the wrong place, courts dismiss as nonexistent discrimination that is hiding in plain sight.

B. The Standard for Showing Discriminatory Intent Is Unduly High

While courts need to recognize that prostitutes and johns are similarly situated, that is only the first step. Courts that have found selective, but not discriminatory, enforcement have done so by setting the standard of proof for discriminatory enforcement so high that women attempting to raise the defense cannot hope to meet the standard.[159]

In order to successfully raise this defense, a defendant must show that the law has, indeed, been enforced in a discriminatory manner. This can be very difficult, partly because courts disagree on what kinds of activity constitute the type of discrimination that violates the Equal Protection Clause.[160] As Andrew B. Weissman comments:

> All jurisdictions would agree . . . that nonenforcement against others similarly situated is not *of itself* sufficient to taint the prosecution of any particular defendant. . . . Beyond this, even the presence of certain types of intent to discriminate may not be enough to create a defense. The legitimate role that intentional selectivity plays in the administration of criminal laws has never been questioned by the courts.[161]

---

157. *Id.* at 46–47.

158. *Id.* at 47.

159. Lefler also takes aim at the high standards some courts use to determine discrimination, arguing that courts should avoid "setting standards that are nearly impossible to attain" for proving discriminatory intent. Lefler, *supra* note 19, at 26.

160. *See* Weissman, *supra* note 93, at 502 ("In determining the kinds of activity that comprise a violation of the duty of non-discriminatory enforcement, unanimity exists up to a certain point.").

161. *Id.* at 502–03.

In *Oyler v. Boles*,[162] the Supreme Court provided relatively little guidance on the question, implying only that selection "*deliberately* based upon an unjustifiable standard such as race, religion, or other arbitrary classification" would be enough to prove discriminatory enforcement.[163] Afterward, many courts, including those dealing with discriminatory enforcement of prostitution laws, looked to the Supreme Court's decision in *Snowden v. Hughes*[164] to interpret this formulation of "deliberate" discrimination[165]:

> The unlawful administration by state officers of a state statute fair on its face, resulting in its unequal application to those who are entitled to be treated alike, is not a denial of equal protection unless there is shown to be present in it an element of *intentional* or *purposeful* discrimination.[166]

The Court in *Snowden* went on to state that intentional or purposeful discrimination "may appear on the face of the action taken with respect to a particular class or person, or it may only be shown by extrinsic evidence showing a discriminatory design to favor one individual or class over another not to be inferred from the action itself."[167]

Courts differ widely on their interpretations of the test for intentional or purposeful discrimination, resulting in a patchwork of tests that vary from state to state both in terms of their elements and in terms of the difficulty a defendant will have in meeting them. The difficulty of meeting these tests is often tied to whether or not the court believes that the selective enforcement was based on sex.

The court in *State v. McCollum* rejected a test laid out in a previous decision that made it very difficult to prove discriminatory enforcement of a prostitution statute.[168] That test would require a defendant to show that "the failure to prosecute was *selective, persistent, discriminatory and without justifiable prosecutorial discretion*."[169] Instead, the court adopted the standard articulated in *Wayte v. United States*,[170] under which the defendant must show that the prosecution had a discriminatory effect and that it was motivated by a discriminatory purpose.[171] The *McCollum* court interpreted the discriminatory-effect part of the standard as analogous to the first

---

162. 368 U.S. 448 (1962).

163. *Id.* at 456 (emphasis added).

164. 321 U.S. 1 (1944).

165. Weissman, *supra* note 93, at 503.

166. *Snowden*, 321 U.S. at 8 (emphasis added).

167. *Id.* (citations omitted).

168. 464 N.W.2d 44, 48–49 (Wis. Ct. App. 1990).

169. State v. Johnson, 246 N.W.2d 503, 507 (Wis. 1976), *quoted in McCollum*, 464 N.W.2d at 48.

170. 470 U.S. 598 (1985).

171. *McCollum*, 464 N.W.2d at 48 (citing *Wayte*, 470 U.S. at 608).

element a defendant must prove to establish discriminatory enforcement: that similarly situated people are treated differently.[172] The *McCollum* court then determined that johns and prostitutes were similarly situated, that only the women were arrested, and that, therefore, the "state's program of enforcement had a discriminatory effect."[173] Turning next to the discriminatory-purpose part of the standard, the *McCollum* court determined that a report made by a police sergeant stating that the only purpose of his investigation was to arrest women suspected of prostitution was enough to establish discriminatory purpose.[174] This relatively low standard is remarkable, especially given the fact that the same sergeant, through his deposition testimony, attempted to establish that he also had focused on the men.[175] The women charged with prostitution were thus able to meet the burden of proof and the charges against them were properly dropped.[176]

In contrast to the *McCollum* court, the court in *People v. Superior Court of Alameda County*[177] adopted a two-element formulation of the defense of discriminatory enforcement: that the defendant has been "deliberately singled out for prosecution on the basis of some invidious criterion"; and "that the prosecution would not have been pursued except for the discriminatory design of the prosecuting authorities."[178] The *Superior Court* standard collapses all three elements of discriminatory enforcement in the first part of its standard and adds a fourth: that but for the discriminatory purpose of the police officers, the defendant would not have been arrested. This exceptionally high standard was used to deny a finding of discriminatory enforcement even though the defendants presented statistical evidence of the discrepancy in arrest rates between prostitutes and johns, evidence that the Oakland Police Department employed primarily male decoys to arrest women, evidence of six specific cases in which the prostitute was arrested and the john set free, evidence that the customary practice of the Oakland Police Department was to custodially arrest the prostitute while merely citing the john, and evidence that prostitutes, but not johns, were quarantined.[179] The fact that the court refused to find discriminatory enforcement when faced with all of this

172. *Id.* at 49.

173. *Id.* at 50.

174. *Id.* at 50–51.

175. *See id.* at 51.

176. *See id.* at 52 (affirming the lower court's decision to dismiss the charges against the defendants).

177. 562 P.2d 1315 (Cal. 1977).

178. *Id.* at 1320 (internal quotation marks omitted).

179. *See id.* at 1320–23 (discussing this evidence finding no constitutional infirmity in the police department's actions).

evidence to the contrary begs the question: if this evidence is not enough, how much more will it take?

The court in *Commonwealth v. An Unnamed Defendant*[180] followed a third test in evaluating discriminatory-enforcement claims: "that a broader class of persons than those prosecuted has violated the law, that failure to prosecute was either *consistent or deliberate*, and that the decision not to prosecute was based on an impermissible classification such as . . . sex."[181] The defendant in this case submitted evidence of prostitution arrests by the Brockton police, her affidavit stating that the police arrested her and not the john on one occasion[182] and the testimony of a Brockton police detective regarding the department's policy on making prostitution arrests.[183] The court found that the defendant's evidence met the standard for discriminatory enforcement and the charges against her were properly dropped.[184]

In order to make the defense of discriminatory enforcement available to more defendants, courts should reject tests like the one espoused by the court in *Superior Court* that set the standard of proof too high. By doing so, these courts essentially take a defense off the table to which these defendants have a constitutional right. Instead, courts should adopt tests and standards of proof similar to those adopted by the *McCollum* and *Unnamed Defendant* courts. Specifically, courts should allow defendants to prove discriminatory intent with statistics that demonstrate disparities between the arrest and prosecution of prostitutes and johns; testimony from competent representatives of the police department as to discriminatory policies, including the exclusive use of male decoys, that result in higher numbers of arrests of prostitutes than johns, and testimony by the defendant herself that she was arrested on one or multiple occasions while the john was not. Indeed, one judge has gone one step further, refusing to hear cases against prostitutes when the police failed to arrest and prosecute the johns.[185]

---

180. 492 N.E.2d 1184 (Mass. App. Ct. 1986).

181. Commonwealth v. Franklin, 385 N.E.2d 227, 233–34 (Mass. 1978) (emphasis added) (citations omitted), *quoted in Unnamed Defendant*, 492 N.E.2d at 1188.

182. On this occasion, the defendant was performing a sex act with the male in a vehicle when she was arrested by the detective. *Unnamed Defendant*, 492 N.E.2d at 1186. The male was not arrested. *Id.*

183. The detective testified that "'ordinarily' the police do not arrest the driver of the vehicle, or 'radio ahead to another unit' to make an arrest. In cases where the vehicle is followed, and a 'sexual act is [found] going on,' only the female is arrested." *Id.* (alteration in original).

184. *Id.* at 1188.

185. Kandel, *supra* note 71, at 340.

C. *Legitimate State Interests Do Not Sound in Sexism*

The rationales that courts use to support their determinations that prostitutes and johns are not similarly situated are remarkably consistent across state lines and, indeed, are often the same rationales given by courts in other countries.[186] These rationales also double as support for holding that selective enforcement is, in fact, justified. Because these rationales carry such weight with so many courts, it is important that they be grounded in reality and not in sexism. Therefore, courts should scrutinize these justifications closely.

The first rationale has a long history and has been used by many courts across the country: police officers and prosecutors can legitimately focus their efforts on prostitutes because they are the "profiteers" in the transaction.[187] Judges who label all prostitutes "profiteers" make the mistake of attributing the experience of the call girl, who may truly be in control of what she does with whom, to all prostitutes. However, this reasoning completely ignores the prostitutes who do not keep all of the money they earn. Under this rationale, police officers should only focus on prostitutes in order to arrest the real profiteer in the transaction: the pimp.

These courts further mistakenly assume that identifying one party in the transaction as the profiteer renders that party more culpable and, therefore, deserving of punishment.[188] This reasoning ignores the real possibility that the john has an economic advantage over the prostitute. After all, he is the one with the disposable money to spend on commercial sex, while she may have resorted to prostitution out of necessity. In the case of prostitution, each party gives something to the other, and each gets something in return. Their acts are reciprocal, and in a broad sense, they are both profiting.

---

186. *See, e.g.*, Barbara Havelková, *Using Gender Equality Analysis to Improve the Wellbeing of Prostitutes*, 18 CARDOZO J.L. & GENDER 55, 76 (2011) (describing the South African case *State v. Jordan* 2002 (6) SA 642 (CC), in which a majority of the nation's Constitutional Court "found substantive differences between the prostitute and the client, which warranted making the prostitute the primary offender").

187. *See, e.g.*, *Ex parte* Carey, 207 P. 271, 274 (Cal. Dist. Ct. App. 1922) ("The fact that the fallen woman carries on the business of commercialized vice justifies whatever discriminations may be found in the statute.").

188. One court reasoned:

> The customer forms the base of the triangle; the prostitute, male or female, constitutes the largest class of profiteers; and at the apex are the pimp, the panderer, and the bar, restaurant, hotel and motel proprietors who knowingly derive profit from the vicious trade.
>
> In order to most efficiently utilize its limited resources, the vice control unit concentrates on the profit[]eers . . . with special emphasis on those at the apex of the illicit commerce.

People v. Superior Court of Alameda Cnty., 562 P.2d 1315, 1321 (Cal. 1977).

The second rationale for allowing the state to focus on the prostitute draws a parallel between the enforcement of prostitution laws and the enforcement of narcotics laws and gambling laws, all of which focus on the provider.[189] The supposed similarity between prostitution and the drug trade led one judge to write, "in many areas of unlawful activity, we concentrate upon the professionals engaging in the illegal business, rather than their customers; e.g., drugs and gambling. There is *no reason whatsoever* that compels prostitution to be treated differently."[190]

However, closer examination of prostitution as compared to drug transactions and gambling belies this claim. As Barbara Havelková points out, narcotics are highly addictive and pose significant dangers to the buyer,[191] risks in which the seller does not share. In direct contrast, the risks of prostitution are either distributed equally between seller and buyer (the risk of contracting a sexually transmitted disease), or the risks are mainly incurred by the seller (the risk of physical and sexual violence[192] and the risk that the client will refuse to wear a condom[193]). Some similarities between low-level drug sellers and street prostitutes may be drawn: both may be at the mercy of someone who has the authority to control their behavior, both may be addicted to one substance or many, and both may be poor. However, there are significant differences between those who buy drugs and those who buy sex. Clients in the drug trade are often addicts who may be disempowered themselves. The harms that they inflict are mostly suffered by the addicts themselves and their families. In contrast, johns are not necessarily powerless and have the clear ability to physically, emotionally, and sexually harm the prostitutes they hire. Many prostitutes

189. Blake v. State, 344 A.2d 260, 262 (Del. Super. Ct. 1975) ("The 'seller' is more strictly controlled and more severely punished in any number of criminal statutes."); United States v. Moses, 339 A.2d 46, 55 (D.C. 1975) ("It is not doubted that the major law enforcement efforts in enforcing the statute are directed against the sellers of sex—as is true in the enforcement of the narcotics laws, where sellers are the principal police targets."); State v. Gaither, 224 S.E.2d 378, 379 (Ga. 1976) (Hall, J., concurring specially) ("It does not violate equal protection guarantees for the legislature to criminalize the conduct of the seller but not the buyer of commercialized sex; the enforcement scheme is similar to that typically applied in liquor and drug statutes."); City of Minneapolis v. Buschette, 240 N.W.2d 500, 505 (Minn. 1976) (quoting approvingly the state's contention that "in many instances, the arrest of one seller will prevent more occurrences of [illegal] behavior . . . than the arrest of a number of buyers" and its comparison of narcotics enforcement and prostitution enforcement).

190. City of Columbus v. Leonard, No. 76AP-276, 1976 WL 190152, at *5 (Ohio Ct. App. Aug. 19, 1976) (Whiteside, J., concurring) (emphasis added).

191. Havelková, *supra* note 186, at 78.

192. *See id.* at 67 ("[T]he overwhelming majority of women in prostitution report repeated instances of verbal abuse, physical assault, and rape by both procurers and buyers. Cross-culturally, the prevalence of post-traumatic stress disorder ('PTSD') is at 78–80%, levels which are higher than those of Vietnam veterans . . . ." (footnote omitted)).

193. *See id.* at 78 (arguing that "[t]he only potential physical harm to the client from prostitution is that resulting from sexually transmitted diseases, and here, both sides of the transaction share the risk and the blame at least equally").

are hesitant to report instances of violence to the police because they fear arrest or, worse, abuse at the hands of the police officers themselves.[194] This enables johns to use a discriminatory system that favors the buyers of sex to brutalize prostitutes with near impunity. The result is that, for street prostitutes, violence becomes an expected part of the profession.[195]

The third rationale for limiting enforcement efforts to prostitutes, that prostitutes are the carriers of venereal disease, has a long and troubling pedigree. One case from 1922 exemplifies this logic and is atypical only for the raw sexism of its language:

> In truth, from the standpoint of public health they are sometimes referred to as pestilential and their places of abode as pest houses. . . . The fallen woman occupies a relation to society very analogous to that of the chronic typhoid carrier—a sort of clearing house for the very worst forms of disease. . . . That a woman carrying on the business denounced in the statute is a constant pathological danger no one would question. . . . [This] fact implies the right to seize the offender and detain her, not only for mere purposes of temporary quarantine, but for the laudable purpose of reclaiming her and destroying the probability of a subsequent renewal of the danger.[196]

One could explain this case away as an unfortunate artifact of the Progressive Era and the fervor for reform typical of that time period. However, more recent court decisions demonstrate that judges still deploy the logic, if not the language, of this rationale.[197] In *Reynolds v. McNichols*,[198] the court upheld a Denver "hold and treat" ordinance that authorized the detention in jail, examination, and treatment of one "reasonably suspected of having a venereal disease" due to that person's history of prior arrests for certain offenses, including prostitution.[199] In justifying the ordinance, the court explained, "the ordinance is aimed at the

194. Cooke & Sontag, *supra* note 10, at 471–72, 474.

195. *See id.* at 471–72 (describing several studies indicating that roughly 80% of female sex workers had experienced violence and several other studies showing that 20%–25% of female sex workers had reported violence at the hands of police).

196. *Ex parte* Carey, 207 P. 271, 274 (Cal. Dist. Ct. App. 1922).

197. In fact, fifty-five years after *Ex parte Carey*, the Supreme Court of California upheld the constitutionality of the practice of quarantining only women because "female prostitutes were more likely than their male customers to communicate venereal diseases." People v. Superior Court of Alameda Cnty., 562 P.2d 1315, 1323 (Cal. 1977).

198. 488 F.2d 1378 (10th Cir. 1973).

199. *Id.* at 1380, 1383.

primary source of venereal disease and the plaintiff, being the prostitute, was the potential source, not her would-be customer."[200] This logic ignores the fact that it is impossible for a prostitute, acting alone, to spread venereal disease. Furthermore, it ignores the greater culpability of the john who may go on to have sex with a woman—his wife or girlfriend—who has no notice that her partner may have recently contracted a disease. In contrast, a john who has sex with a prostitute—a stranger—would have notice of the heightened risk of contracting a disease and would be able to take necessary precautions.

VII. Conclusion

The discriminatory enforcement of laws targeting prostitution has a long pedigree that extends back to the period before prostitution was a distinct criminal offense.[201] Equally ancient are the justifications for bringing the force of the law down on prostitutes while letting the johns go free: prostitutes are fallen women, irredeemably lost, perpetrators of nocturnal vice, pest-house dwellers, profiteers, and outcasts.[202] Armed with these justifications, police officers get to work—conducting street sweeps; posing as johns in cars, clubs, and massage parlors; and offering actual johns, caught with their pants down, a deal: "testify against the girl and we'll forget all about you."[203]

The defense of discriminatory enforcement offers women charged with prostitution the important opportunity to force police departments to abandon these practices.[204] However, these women need the cooperation of the courts. A woman seeking dismissal of a prostitution charge on the ground of discriminatory enforcement already faces an uphill battle.[205] She must gather evidence of intentional and invidious selective enforcement, often through statistical evidence or police testimony.[206] This requires not only courage but also access to resources that women in this position, already disempowered and now facing a criminal prostitution charge, may not possess.[207]

Thus it is important that courts, given a defendant claiming discriminatory enforcement, seriously consider the evidence offered by the defendant and refuse to dismiss it or explain it away. By dropping prostitution charges against women who demonstrate discriminatory

200. *Id.* at 1383.
201. *See supra* subparts III(A)–(B).
202. *See supra* subpart VI(C).
203. *See supra* Part IV.
204. *See supra* subpart V(B).
205. *See supra* Part VI.
206. *See supra* subpart VI(B).
207. *See supra* note 98 and accompanying text.

enforcement, the courts have the opportunity to send a powerful message: that the laws of the United States are applied with an even hand, that prostitution is not just a woman's crime, and that johns, so long the silent and anonymous actor, are to be held equally accountable.

—*Elizabeth M. Johnson*

Mind the GAAP: Moving Beyond the Accountant–Attorney Treaty*

I. Introduction

Imagine an investor considering a company's stock. Before buying, the investor diligently examines the company's financial statements and Securities and Exchange Commission (SEC) filings. These documents mention a handful of pending lawsuits and lay out the bare facts of the suit. They simultaneously assure the investor that the company has a number of good defenses and is diligently defending the suits. The financial statements similarly mention the litigation in a footnote but go on to say any potential loss cannot be reasonably estimated. Though these disclosures have perhaps given the investor some pause, a lack of legal expertise and knowledge of the company's operations prevents the investor from gleaning much from the disclosures. With these concerns in mind, the investor acquires a stake in the company, trusting that any reliable disclosure system would require more substantive discussion if the lawsuits were truly presented major problems for the company. Soon, however, the company is slammed with a judgment that will bankrupt it (or settles for a startling sum), and much, if not all, of the investment is lost.

Given the current reporting regime and the generally accepted accounting principles (GAAP) for loss contingencies, this story is hardly a fantasy. Such an event should not be acceptable in a system that focuses on transparency and disclosure. If regulators are to fulfill their mission of providing ample, high-quality information so that investors can make informed decisions,[1] the disclosure of contingent liabilities in the litigation context must be strengthened. Indeed, the Financial Accounting Standards Board (FASB)—the body tasked with articulating GAAP[2]—unsuccessfully attempted to reform this area beginning in 2008.[3] Among the reforms most

---

* I am grateful to the editors of the *Texas Law Review*—especially Elizabeth Johnson, Dina McKenney, and Spencer Patton—for their hard work editing this Note, to Professor Henry Hu for helpful comments during the writing process, and above all to my family and friends for their constant love and support.

1. Henry T.C. Hu, *Too Complex to Depict? Innovation, "Pure Information," and the SEC Disclosure Paradigm*, 90 TEXAS L. REV. 1601, 1614–15 (2012).

2. *See Facts about FASB*, FIN. ACCT. STANDARDS BOARD, http://www.fasb.org/jsp/FASB/Page/SectionPage&cid=1176154526495 ("Since 1973, the Financial Accounting Standards Board (FASB) has been the designated organization in the private sector for establishing standards of financial accounting . . . .").

3. *See* FIN. ACCOUNTING STANDARDS BD., PROPOSED STATEMENT OF FINANCIAL ACCOUNTING STANDARDS: DISCLOSURE OF CERTAIN LOSS CONTINGENCIES (2008) [hereinafter FASB EXPOSURE DRAFT 1], *available at* http://www.fasb.org/cs/BlobServer?blobkey=id&

quickly rejected was a provision requiring companies to report settlement valuations of certain pending litigation.[4] Due to the overwhelmingly negative response to *any* reform in this area, the reporting of settlement valuations received inadequate attention as a viable approach. This Note recognizes and develops the viability of a system of disclosure based on settlement value and its potential to satisfy warring factions of attorneys and accountants. The proposed disclosure system would require dollar amounts offered by disclosing companies in settlement negotiations to form the baseline for quantifying losses that may not otherwise trigger current disclosure requirements because the potential losses cannot be reasonably estimated.

Part II of this Note will examine the existing reporting system for litigation contingencies and its current inadequacy. Part III will outline a proposed reform to meet this issue that focuses on disclosure of the value of settlements offered by the reporting company. Part IV will examine the practical obstacles and objections the proposed reform would have to overcome. Part V will briefly conclude.

II. Examining the Problem

A. Existing Reporting Requirements

The existing regime for reporting litigation contingencies originates in both SEC rules and the accounting standards that establish the rules for preparing financial statements. This information is disclosed in mandated periodic filings with the SEC and made available to the investing public via forms 10-K (annually) and 10-Q (quarterly).[5] Regulation S-K is the most basic toolkit for SEC disclosure filings and explains to reporting companies what must be included in disclosure documents.[6] The primary provision

---

blobwhere=1175818814390&blobheader=application%2Fpdf&blobcol=urldata&blobtable=MungoBlobs (proposing standards to enhance the disclosure requirements for loss contingencies); *see also Project Updates*, FIN. ACCT. STANDARDS BOARD, http://www.fasb.org/jsp/FASB/FASBContent_C/ProjectUpdatePage&cid=900000011071 (last updated July 12, 2012) (noting that the FASB voted to remove the disclosure project from its agenda in 2012).

4. *See* FASB EXPOSURE DRAFT 1, *supra* note 3, at iii; *see also, e.g.*, Letter from Michael H. Gibbs, Vice President & Gen. Counsel, Whataburger, to Dir., Fin. Accounting Standards Bd. (July 21, 2008), *available at* http://www.fasb.org/cs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1175818420130&blobheader=application%2Fpdf&blobheadername2=Content-Length&blobheadername1=ContentDisposition&blobheadervalue2=42138&blobheadervalue1=filename%3D52225.pdf&blobcol=urldata&blobtable=MungoBlobs ("We are greatly concerned about the proposed expanded disclosures in financial statements of pending or anticipated litigation.").

5. Securities Exchange Act of 1934 § 13(a), 15 U.S.C. § 78m(a) (2012); 17 C.F.R. § 240.13a-1 (2013) (requiring annual reports); 17 C.F.R. § 240.13a-13 (2013) (requiring quarterly reports).

6. *See* Edmund W. Kitch, *The Theory and Practice of Securities Disclosure*, 61 BROOK. L. REV. 763, 800 (1995) (describing Regulation S-K as "one common, master form" for disclosures under both the '33 Act and the '34 Act).

relating specifically to potential litigation losses in Regulation S-K is Item 103, which requires disclosure of various factual information regarding litigation falling outside of "ordinary routine litigation."[7] The instructions to Item 103 clarify that suits claiming damages of less than 10% of the company's assets need not be disclosed.[8] These limitations on disclosure are presumably to keep the reporting burdens on companies manageable.[9] The disclosures required under Item 103 are predominantly factual and perusing the litigation section of a filing is often little more than a lesson in civil procedure as various steps of the litigation are described.[10] As such, disclosures under Item 103 are of little help to investors looking for substantive information.

In addition to these factual disclosures, Item 303 is often implicated and requires large litigation contingencies to be discussed under the management's discussion and analysis (MD&A) section of the filing—a key piece of securities disclosure.[11] Item 303 requires litigation contingencies to be discussed in the MD&A to the extent that the potential losses represent a "known . . . uncertaint[y] . . . that the registrant reasonably expects will have a material . . . impact on . . . revenues."[12] Despite these requirements, though, management might merely exclaim that they have excellent defenses to completely frivolous claims. Furthermore, this section may merely incorporate the appropriate footnote from the financial statements by reference and provide little actual discussion or analysis.[13] This likely stems from the reluctance of reporting companies to reveal any information that could prejudice their claims in pending litigation.[14]

---

7. SEC Regulation S-K, Item 103, 17 C.F.R. § 229.103 (2013).

8. *Id.*

9. *Cf.* Michael A. Harrison, Note, *Regulation FD's Effect on Fixed-Income Investors: Is the Public Protected or Harmed?*, 77 IND. L.J. 189, 218 (2002) (examining the interplay of compliance costs and the volume of required disclosure).

10. *See* Karen M. Hennes, *Disclosure of Contingent Legal Liabilities*, 33 J. ACCT. & PUB. POL'Y (forthcoming 2014) (manuscript at 5), *available at* http://ac.els-cdn.com/S0278425413000926-main.pdf?_tid=8ca9d548-8127-11e3-80ca-00000aab0f01&acdnat=1390149441_597c9cb42ffcb296f3995092ef31cac0 ("Overall, the consistent conclusion from prior studies on litigation contingency disclosures is that firms provide fairly limited information on the estimated loss amount.").

11. *See* Troy A. Paredes, *Blinded by the Light: Information Overload and its Consequences for Securities Regulation*, 81 WASH. U. L.Q. 417, 425 (2003) (calling the MD&A "[o]ne of the most important textual disclosures").

12. SEC Regulation S-K, Item 303, 17 C.F.R. § 229.303 (2013).

13. *See, e.g.*, Anadarko Petroleum Corp., Quarterly Report (Form 10-Q), at 38, 44 (July 27, 2011) [hereinafter Anadarko 10-Q] (referencing appropriate footnotes from financial statements concerning litigation contingencies and failing to further elaborate).

14. *See infra* notes 28–48 and accompanying text.

Audited financial statements are the other major prong of corporate disclosure and must be included in periodic filings.[15] The current GAAP for contingent liabilities of all types is found in the Statement of Financial Accounting Standards No. 5 (SFAS 5).[16] As a threshold matter, SFAS 5—like Regulation S-K Item 103—includes a materiality requirement before any discussion of a contingent liability is required.[17] The standard defines a loss contingency as an "existing condition, situation, or set of circumstances involving uncertainty . . . that will ultimately be resolved when one or more future events occur or fail to occur."[18] Under SFAS 5, losses must be accrued on financial statements when two factors are met. It must be "probable that . . . events will occur confirming the fact of the loss" and "[t]he amount of the loss [must] be [capable of being] reasonably estimated."[19] Disclosure (though not accrual) is still required if there is a "reasonable possibility" a loss will occur.[20] Such a disclosure should try to "estimate . . . the possible loss or range of loss or state that such an estimate cannot be made."[21] These disclosures are placed in a footnote to the financial statements.[22] Rather than provide concrete percentages to define terms like "probable" and "reasonably possible," SFAS 5 instead allows "significant leeway for professional judgment in the disclosure decision."[23]

---

15. *See* SEC Regulation S-X, 17 C.F.R. §§ 210.3-01–.3-02 (2013) (requiring companies filing periodic reports to include audited financials in those reports).

16. ACCOUNTING FOR CONTINGENCIES, Statement of Fin. Accounting Standards No. 5 (Fin. Accounting Standards Bd. 1975) [hereinafter SFAS 5], *available at* http://www.fasb.org/cs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1175820910926&blobheader=application%2Fpdf&blobcol=urldata&blobtable=MungoBlobs. The FASB codified all accounting standards as of 2009 in their Accounting Standards Codification (ASC). FIN. ACCOUNTING STANDARDS BD., NOTICE TO CONSTITUENTS (V 4.8) ABOUT THE CODIFICATION 4 (2012), *available at* https://asc.fasb.org/imageRoot/80/34350180.pdf. SFAS 5 is primarily codified at ASC 450. KPMG, THE FASB ACCOUNTING STANDARDS CODIFICATION CROSS-REFERENCE QUICK GUIDE 1 (2010), *available at* http://us.kpmg.com/microsite/attachments/us_accounting_bulletin_2010/fasb-asc-quick-guide.pdf. To avoid confusion and because the codification took place concurrently with the proposed changes in SFAS 5 discussed in this Note, I will use the precodification terminology.

17. *See* SFAS 5, *supra* note 16, at 8 (qualifying the first twenty paragraphs with the statement that "[t]he[se] provisions . . . need not be applied to immaterial items"). A determination of what information is material in a given circumstance is fraught with uncertainty and outside the scope of this Note. Instead, this Note will assume and use as examples only cases that are clearly material and may even threaten the continued viability of the company as a going concern.

18. *Id.* para. 1.

19. *Id.* para. 8.

20. *Id.* para. 10.

21. *Id.*

22. Robert D. Fesler & J. Larry Hagler, *Litigation Disclosures Under SFAS No. 5: A Study of Actual Cases*, ACCT. HORIZONS, Mar. 1989, at 10, 11.

23. *Id.* at 12.

This has led at least one author to suggest that "accounting for contingencies more closely resembles an art than a science."[24]

These provisions, working together, are meant to provide investors with enough information regarding potential litigation losses of a company to make an informed investment decision.[25] It is far from clear, however, that this is actually the case.

B. Inadequacies of the Current System

A close inspection of the current disclosure system for litigation contingencies and how the system works in practice reveals that the disclosure system is not accomplishing the enunciated goal of transparency. One of the primary goals of the entire SEC disclosure system is to provide investors of all types enough factual information about an investment to make what they believe to be wise investment decisions.[26] The system of reporting for litigation contingencies falls short of this ideal in several respects.

First, the process by which financial reports are prepared is not conducive to transparency. Evaluation of litigation contingencies is outside the realm of auditor expertise as it requires at least some degree of legal judgment.[27] Auditors must therefore rely on the responses of corporate attorneys to audit-inquiry letters.[28] The problem arises from attorney hesitance to reveal any useful information in this process. This hesitance stems from a desire to avoid disclosing information that might prejudice the company's stance in the pending litigation.[29] The quality and usefulness of attorney responses to audit inquiries is suspect at best.[30] One commentator has suggested that "[i]n most instances, the auditor could . . . obtain more information just by reviewing the pleadings filed at the local courthouse."[31] The inadequacy of attorney responses to audit inquiries was presumably recognized when the ABA and the American Institute of Certified Public

---

24. Matthew J. Barrett, *Opportunities for Obtaining and Using Litigation Reserves and Disclosures*, 63 OHIO ST. L.J. 1017, 1032 (2002).

25. *See id.* at 1030–31.

26. Hu, *supra* note 1.

27. *See* W.R. Koprowski et al., Essay, *Financial Statement Reporting of Pending Litigation: Attorneys, Auditors, and Differences of Opinions*, 15 FORDHAM J. CORP. & FIN. L. 439, 439 (2010) (discussing auditor reliance on attorney responses to audit inquiries).

28. *Id.* at 446–47.

29. *See id.* at 453 ("The uncertainty about the likely application and potential waiver of the attorney-client privilege and . . . information covered by the work product doctrine[] creates an incentive for this limited disclosure[] to the auditor.").

30. *See* M. Eric Anderson, *Talkin' 'Bout My Litigation—How the Attorney Response to an Audit Inquiry Letter Discloses as Little as Possible*, 7 TRANSACTIONS: TENN. J. BUS. L. 143, 144 (2005) (suggesting attorney responses provide no more information than could be gleaned from a trip to the county courthouse).

31. *Id.*

Accountants formed a joint task force "to address concerns regarding language used by auditors in audit-inquiry letters . . . and responses by attorneys."[32] The task force was dissolved before any reform was accomplished.[33]

The hostile attitude of lawyers toward substantive disclosure in response to audit inquiries is perhaps best demonstrated by the ABA's 1976 guidance on responses to auditors' inquiries: "[i]t is not . . . believed necessary, or sound public policy, to intrude upon the confidentiality of the lawyer-client relationship in order to command [confidence in the auditing process]."[34] Despite its bellicose rhetoric, this document is known as part of the "treaty" between accountants and lawyers on this subject because it at least set up some workable framework for attorney responses to audit inquiries.[35] It is telling that the document opens by stating "[t]he public interest in protecting the confidentiality of lawyer-client communications is fundamental."[36] It goes on to warn lawyers that "an adverse party may assert that any evaluation of potential liability is an admission."[37]

The ABA's position reflects the view that the "most basic fear is that any evaluation or assessment of a client's liability . . . will be disclosed to an adversarial party and used against the client in a subsequent court proceeding."[38] Though the pronouncements of the ABA are not binding on attorneys,[39] here they are indicative of a general wariness toward audit letters by the legal profession. Thus, the battle lines between accountants and attorneys are drawn with a seemingly intractable divide in what has been described as a "clash of cultures."[40] Under this guidance from the ABA, attorney responses to audit inquiries were so arcane that accountants needed guidance to essentially "translate" audit letter responses.[41] There

32. *Highlights of Technical Activities*, IN OUR OPINION (Am. Inst. of Certified Pub. Accountants Audit & Attest Standards Grp., New York, N.Y.), Spring 2004, at 4, 7.

33. Anderson, *supra* note 30.

34. Am. Bar Ass'n, *Statement of Policy Regarding Lawyers' Responses to Auditors' Requests for Information*, 31 BUS. LAW. 1709, 1710 (1976) [hereinafter ABA Policy Statement].

35. *See* Interview with Michael Young, Att'y, Willkie Farr & Gallagher LLP (Apr. 2011) [hereinafter Young Interview], *in* Kim Nilsen, *Renewed Focus on Loss Contingency Disclosures*, J. ACCT., Apr. 2011, at 37, 38 ("About 35 years ago, the accounting profession and the legal profession got together . . . and put in place a structure for auditor-lawyer dialogue. The document capturing that structure has informally become known as 'the treaty.'").

36. ABA Policy Statement, *supra* note 34, at 1709.

37. *Id.* at 1711.

38. Anderson, *supra* note 30.

39. *Cf.* Ted Schneyer, *An Interpretation of Recent Developments in the Regulation of Law Practice*, 30 OKLA. CITY U. L. REV. 559, 594–95 (2005) (noting that the ABA ethics rules and standards for criminal defense are nonbinding).

40. Young Interview, *supra* note 35.

41. Anderson, *supra* note 30, at 166.

have been calls to soften the ABA's stance on these responses,[42] but in light of the very real possibilities that disclosure could prejudice companies in ongoing litigation, the ABA's position seems justified to protect the sanctity of the attorney–client relationship. Such relationship has long been seen as fundamental to the legal profession.[43]

Concerns over the possibility of audit responses prejudicing clients are not figments of overactive legal imaginations. Courts are split on whether disclosure to auditors acts as a waiver of the attorney–client privilege.[44] Though some jurisdictions protect these disclosures using the work product doctrine (which protects work prepared in anticipation of litigation), this is not uniformly true.[45] Such information could subsequently be discoverable and used against a party in either open court or settlement negotiations.[46] Indeed, one scholar has written on litigation advantages that can be derived from the *current* reporting system for litigation contingencies.[47] His basic premise involves using disclosure documents and the accruals appearing therein (both explicit and implicit) to reverse engineer a company's valuation of the litigation and urging litigators to "recognize the discovery possibilities of obtaining accounting information and supporting data regarding contingencies."[48] In so doing, adverse parties could gain a leg up both in settlement negotiations and in the courtroom.[49] Thus, any attempt at broadening disclosure must take concerns about prejudicing ongoing litigation seriously.

An additional problem with the current regime is the latitude companies are given to estimate liabilities in their responses to audit inquiries. If such liabilities cannot be reasonably estimated, no accrual is taken no matter how likely the contingency is to occur.[50] Empirical work in this area indicates that applications of these standards vary widely, and the end result is often that no loss at all is reflected on financial statements.[51]

---

42. *See, e.g.*, *id.* at 167 ("A change in the ABA Statement making it less hostile to the process would be productive.").

43. *See, e.g.*, United States v. Zolin, 491 U.S. 554, 562 (1989) (discussing the significance and tradition of the attorney–client relationship).

44. For a survey of cases on both sides of this issue, see generally Anderson, *supra* note 30, at 157–63.

45. *See id.*

46. *See* Barrett, *supra* note 24, at 1081–82 (noting the potential for sophisticated litigators to effectively utilize the discovery process by examining accounting information).

47. *See id.*

48. *Id.* at 1081.

49. *See id.* at 1103 (highlighting the advantages of using discovery requests and subpoenas to gain access to litigation-contingency information).

50. *See* SFAS 5, *supra* note 16, para. 8. Of course, if the probability reached 100%, it would no longer be a contingent liability. *See id.* para. 1.

51. OFFICE OF THE CHIEF ACCOUNTANT ET AL., U.S. SEC. & EXCH. COMM'N, REPORT AND RECOMMENDATIONS PURSUANT TO SECTION 401(C) OF THE SARBANES-OXLEY ACT OF 2002 ON ARRANGEMENTS WITH OFF-BALANCE SHEET IMPLICATIONS, SPECIAL PURPOSE ENTITIES, AND

The decision to omit a potential litigation loss from financial statements might often come about due to the inherent difficulty of estimating the amount of loss. In a hypothetical lawsuit in which the plaintiff claims $1 billion in damages, the final recovery could range anywhere from zero to $1 billion. The range might broaden even further if a plaintiff seeks unspecified punitive damages. The ABA policy statement may put it best in saying "the amount or range of potential loss will normally be as inherently impossible to ascertain, with any degree of certainty, as the outcome of the litigation."[52] The problem has not gone unnoticed by the SEC. The SEC's Deputy Chief Accountant Scott Taub noted in 2004 that it is "somewhat surprising the number of instances where zero is considered the low end of a range with no number more likely than any other right up until a large settlement is announced."[53] Despite Mr. Taub's statements, however, given the ABA's hostility to meaningful disclosure, such numbers should come as no surprise.

C. *Examples of the Problem*

Unfortunately, the concerns on the accountants' side are just as convincing. "Surprise" losses are far from unheard of.[54] The need for enhanced disclosure in this area can perhaps best be seen on an anecdotal level. Though an empirical study of disclosure failings is well beyond the scope of this Note, a couple of examples will help to illustrate the current problem in disclosure of litigation contingencies.

A dated—though no less damning—example of the problem can be seen in the litigation between Pennzoil and Texaco during the late 1980s. Pennzoil filed suit, alleging Texaco had tortiously interfered with Pennzoil's merger agreement with Getty Oil.[55] The case proceeded to trial where a jury ultimately awarded Pennzoil $10.53 billion in damages.[56] The

---

TRANSPARENCY OF FILINGS BY ISSUERS 69–70 (2005) [hereinafter SEC, REPORT AND RECOMMENDATIONS] (presenting data from a study of filings by issuers related to contingencies and noting that not only did "disclosures about contingent obligations vary widely in terms of format and location" but less than 10% of issuers in the sample "report that they have recognized any amount of liability on their balance sheets for any legal contingent obligation" (emphasis omitted)).

52. ABA Policy Statement, *supra* note 34, at 1714.

53. Scott A. Taub, Deputy Chief Accountant, U.S. Sec. & Exch. Comm'n, Remarks at the University of Southern California Leventhal School of Accounting SEC and Financial Reporting Conference (May 27, 2004), *available at* http://www.sec.gov/news/speech/spch052704sat.htm.

54. For a further discussion of such "surprise" events, see James S. Johnson, Comment, *The Accountable Attorney: A Proposal to Revamp the ABA's 1976 Statement of Policy Regarding Lawyers' Responses to Auditors' Requests for Information*, 14 TEX. WESLEYAN L. REV. 27, 30–32 (2007).

55. Edward B. Deakin, *Accounting for Contingencies: The Pennzoil-Texaco Case*, ACCT. HORIZONS, Mar. 1989, at 21, 21–22.

56. *Id.* at 24.

judgment eventually forced Texaco into bankruptcy protection.[57] Throughout the protracted legal battle, Texaco did not accrue any losses in its disclosures to investors.[58] Indeed, "entry of a liability did not occur until actual payment was agreed upon."[59] Such an example serves as a harrowing worst-case scenario for failure of disclosure in this realm.

Of course, much has changed in the world of securities disclosure since the late 1980s. Of particular importance is the more prominent role of the MD&A, with stricter rules and enforcement of those rules.[60] Thus, a more recent example is needed to adequately illustrate the continuing inadequacy of litigation-contingency disclosures. Such an example, like the Texaco case before it, can be drawn from the energy industry.

Anadarko is one of the largest independent oil and gas exploration and production companies in the world.[61] It was ranked 207th on the most recent Fortune 500 list, nestled between the likes of Monsanto and Starbucks.[62] On April 20, 2010, the Deepwater Horizon oil platform exploded in the Gulf of Mexico, killing eleven people and leaving crude oil gushing into the ocean.[63] The well, in which Anadarko owned a 25% share, was a joint venture between BP, Anadarko, and MOEX Offshore.[64] Under the joint operating agreement in effect between those parties, each company was responsible for its proportional share of the costs from the well, including those resulting from the negligence (though not gross negligence) of the operator (BP).[65] Given that damages resulting from the explosion and subsequent oil spill were then estimated to be at least $22 billion,[66] the amount of money riding on a determination of whether BP's actions on the platform were ordinary negligence or gross negligence was considerable.

---

57. *Id.* at 21, 24.

58. *See id.* at 22, 27.

59. *Id.* at 27.

60. For an examination of the development of the MD&A requirements, see Paredes, *supra* note 11, at 425–26.

61. *About Anadarko*, ANADARKO, http://www.anadarko.com/About/Pages/Overview.aspx.

62. *Fortune 500*, FORTUNE, CNN MONEY (2014), http://money.cnn.com/magazines/fortune/fortune500/2013/full_list/index.html?iid=F500_sp_full.

63. Julia Werdigier, *Ending Dispute, Well Partner Settles with BP for $4 Billion*, N.Y. TIMES, Oct. 17, 2011, http://www.nytimes.com/2011/10/18/business/bp-recovers-4-billion-from-anadarko-for-gulf-spill.html?_r=1&.

64. *Id.*

65. *See* Ben Casselman, *Anadarko Blames BP for Rig Disaster*, WALL ST. J., June 19, 2010, http://online.wsj.com/article/SB10001424052748704122904575315244230579982.html; James Herron, *Anadarko Would Take Huge Hit if Forced to Pay into BP's $20 Billion Oil Spill Fund*, THE SOURCE, WALL ST. J. (June 21, 2010, 11:49 AM), http://blogs.wsj.com/source/2010/06/21/anadarko-would-take-huge-hit-if-forced-to-pay-into-bps-20-billion-oil-spill-fund/.

66. *See* Herron, *supra* note 65 (stating that BP had already spent $2 billion and had agreed to a compensation fund of $20 billion).

Anadarko's potential liability under the joint operating agreement threatened their continued viability as a company.[67]

Anadarko settled with BP by agreeing to pay $4 billion to satisfy any claim BP would have against it under the joint operating agreement.[68] An examination of the company's SEC filings during this time illustrates the need for reform in the disclosure of litigation contingencies. The 10-Q filed immediately prior to the announcement of the settlement provides no helpful indication as to what the size of the liability would be. The MD&A mentions the event and refers readers to the financial statement notes for analysis.[69] Though the relevant note is replete with factual information spanning eleven pages,[70] the bottom line regarding the incident is captured by a single paragraph:

> After applying the relevant accounting guidance to the Company's Deepwater Horizon event-related contingent liabilities, the Company's aggregate liability accrual for these amounts is zero as of June 30, 2011. The zero liability accrual is not intended to represent an opinion of the Company that it will not incur any future liability . . . . Rather, [it] is based on currently available facts and the application of accounting rules . . . where the relevant accounting rules do not allow for loss recognition where a potential loss is not considered "probable" or cannot be reasonably estimated.[71]

Thus, from June 30 to October 17 Anadarko's liability for the events surrounding the Deepwater Horizon spill went from zero to $4 billion. These examples amply depict exactly the problem Mr. Taub was referencing in his speech.[72]

Such examples "demonstrate that financial statement disclosure is not moving in a direction consistent with the kind of transparency that . . . investors desire."[73] Empirical work has also shown that disclosures of litigation contingencies have often been too little and too late. For example, one study found that 45% of sampled cases involving material losses lacked any disclosure that such results would have a material economic impact on the companies involved until the quarter in which the case was resolved.[74] Additionally, the results of the study suggested that investors did not fully

---

67. *Id.*

68. Werdigier, *supra* note 63.

69. Anadarko 10-Q, *supra* note 13, at 35, 38.

70. *Id.* at 7–17.

71. *Id.* at 8.

72. *See* Taub, *supra* note 53 (criticizing situations where companies do not accrue because they find zero as likely a number as any other).

73. Koprowski et al., *supra* note 27, at 444.

74. Feng Chen et al., *Auditor Quality and Litigation Loss Contingency Disclosures* 18, 52 fig.1 (Rotman Sch. of Mgmt., Working Paper No. 2179148, 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2179148.

anticipate these losses.[75] Similarly, an empirical study by the SEC found that a very small percentage of eventual liability was ever accrued prior to the time it was actually paid.[76] The Investors Technical Advisory Committee (ITAC)—an arm of the FASB, comprised of investors with accounting expertise, intended to give the FASB the perspective of investors[77]—has similarly expressed concern that large settlements often take investors by surprise.[78]

III. Towards a Solution

A. Laying the Foundation

Any proposed solution must carefully balance the interests of companies in not revealing information that might damage their position in ongoing litigation. I propose a type of middle ground between accountants and lawyers whereby investors are given useful, nonpublic information regarding large litigation losses and companies avoid prejudicing pending litigation with such disclosures. This proposal focuses on the process of litigants reaching a settlement agreement.

In modern commercial litigation, most cases settle.[79] Indeed, settlement is so preferred as a means of dispute resolution that the Federal Rules of Civil Procedure encourage judges to promote settlement or other alternative resolutions to adjudication.[80] It is relatively rare for a company to hurtle toward trial in a case without attempting to settle it.[81] The benefits of settlement are myriad and include avoiding costly attorney's fees, escaping negative publicity, and allowing parties to bargain for a mutually

---

75. *Id.* at 30–31.

76. SEC, REPORT AND RECOMMENDATIONS, *supra* note 51.

77. *Investor Advisory Committee (IAC)*, FIN. ACCT. STANDARDS BOARD, http://www.fasb.org/investors_technical_advisory_committee/.

78. Letter from Lynn Turner, Member, Investors Technical Advisory Comm., to Technical Dir., Fin. Accounting Standards Bd. (Aug. 15, 2008) [hereinafter ITAC Comment Letter], *available at* http://www.fasb.org/cs/BlobServer?blobkey=id&blobwhere=1175819498531&blobheader=application%2Fpdf&blobcol=urldata&blobtable=MungoBlobs.

79. *See* Marc Galanter & Mia Cahill, *"Most Cases Settle": Judicial Promotion and Regulation of Settlements*, 46 STAN. L. REV. 1339, 1339–40 (1994) (disputing estimates that 85%–95% of all cases settle because such estimates include cases settled after commencement of trial and claiming the actual number is closer to two-thirds). Regardless of the true percentage of cases that actually settle, it is apparent that settlement is significantly more prevalent than taking cases all the way to final judgment. *See id.* at 1339 ("[S]ettlement is the most frequent disposition of civil cases in the United States . . . .").

80. FED. R. CIV. P. 16 (providing that the court may order a pretrial conference for the purpose of facilitating settlement).

81. *Cf.* Samuel R. Gross & Kent D. Syverud, *Getting to No: A Study of Settlement Negotiations and the Selection of Cases for Trial*, 90 MICH. L. REV. 319, 320 (1991) ("[T]he nature of our civil process drives parties to settle so as to avoid the costs, delays, and uncertainties of trial, and, in many cases, to agree upon terms that are beyond the power or competence of courts to dictate.").

acceptable resolution.[82] Indeed, one scholar has indicated that from a purely economic perspective (assuming nonzero attorney's fees) *every* case should settle.[83] Because of the importance of the settlement process, providing investors a window into the progress of negotiations would be a valuable informational tool that would serve the aim of the disclosure system. Given the prevalence of settlements in modern litigation, it is odd that companies need not disclose anything about settlement negotiations. Indeed, surprise settlements rather than trial losses seemed of more concern to ITAC in a recent reform proposal.[84]

Given the prevalence of settlement in modern commercial litigation, I propose the use of settlement offers made by the defendants themselves to set a nonzero floor for accrual and disclosure of contingent-litigation liabilities. In situations where the liability has been deemed to be "probable," the minimum accrual under a revised SFAS 5 would be that of the maximum settlement value offered by the reporting company rather than allowing companies to get by with a boilerplate statement that a potential loss cannot be reasonably estimated. Similarly, the same value would be disclosed in the financial statement footnotes (though not accrued) if a loss is deemed to be reasonably possible.

By using the highest settlement number offered as a baseline, the consequences feared most by corporate counsels can be avoided while strengthening the disclosure regime. The Federal Rules of Evidence contain a rule expressly excluding statements made in settlement negotiation from admissibility in court.[85] State rules of evidence largely mirror the Federal Rules.[86] Thus, corporate concern that these disclosures would prejudice them in court would be largely unfounded. The proposed system would also allay corporate fears that more disclosure could damage their position in settlement negotiations. Unlike adversaries getting their hands on a company's subjective evaluation of the true value of a case, disclosure of settlement offers would provide no new information to opponents. Because, by definition, a settlement offer has already been communicated to the other side, this information provides no benefit to plaintiffs.

---

82. *See* Galanter & Cahill, *supra* note 79, at 1350 tbl.1.

83. Scott A. Moss, *Illuminating Secrecy: A New Economic Analysis of Confidential Settlements*, 105 MICH. L. REV. 867, 873 (2007).

84. *See* ITAC Comment Letter, *supra* note 78 (discussing the issue that arises when investors are surprised by previously undisclosed settlements but making no explicit mention of trial losses).

85. FED. R. EVID. 408(a).

86. Stephen D. Easton, *Can We Talk?: Removing Counterproductive Ethical Restraints upon Ex Parte Communication Between Attorneys and Adverse Expert Witnesses*, 76 IND. L.J. 647, 658 n.44 (2001).

B. Necessary Limitations

Several limitations on the idea of using settlement offers made by defendants as a proxy for subjective evaluations of cases must be imposed in order to make the system practicable. First, the disclosure system would still only be invoked in material litigation. This limitation is primarily due to ease of compliance. The existing disclosure rules in this area already have materiality requirements,[87] and it would not be feasible to do away with such requirements for settlement-offer disclosure. It would be extremely cumbersome if a company had to disclose settlement information about every slip and fall case pending against it. Indeed, a workable solution might even sidestep the materiality question entirely and focus only on litigation that could possibly threaten the company as a going concern. Such a solution would not compromise disclosure goals, though, as it is unlikely that investors are concerned with a company's death by a thousand cuts but rather the single "surprise" events mentioned previously.[88]

By keeping compliance costs in mind, this system avoids the primary pitfall of another seemingly attractive solution: that of independent third-party legal evaluation. Such a system would require as part of the audit that an independent legal review take place and value the litigation.[89] Requiring such third-party oversight, however, would be very expensive and would introduce another layer of work to an auditing process that is already time sensitive.[90]

An additional important limitation on the scope of the proposed disclosure would be to allow the companies to aggregate the values of multiple cases into a single set of disclosure numbers. Merely requiring disclosure of settlement offers in individual cases might still prejudice ongoing and future litigation as it could provide adverse parties with a window into a particular company's negotiating strategy. This aggregation approach was embraced in a failed reform undertaken by the FASB.[91] Using the concept of aggregation of individual settlement offers prior to disclosure as proposed by the FASB would allow companies to avoid wholly revealing their settlement negotiation strategies and would ameliorate a set of potential objections from issuers. Of course this

87. Kitch, *supra* note 6, at 819–20.

88. *See supra* notes 75–78 and accompanying text.

89. Johnson, *supra* note 54, at 48–49.

90. Koprowski et al., *supra* note 27, at 456–57.

91. FIN. ACCOUNTING STANDARDS BD., PROPOSED ACCOUNTING STANDARDS UPDATE: CONTINGENCIES (TOPIC 450): DISCLOSURE OF CERTAIN LOSS CONTINGENCIES 41–42 (2010) [hereinafter FASB EXPOSURE DRAFT 2], *available at* http://www.fasb.org/cs/BlobServer?blobkey=id&blobwhere=1175823559187&blobheader=application%2Fpdf&blobcol=urldata&blobtable=MungoBlobs.

problem would still remain for smaller companies that might only have one material litigation taking place at any one time.

C. *The Proposal in Practice*

To illustrate how the proposed system would work in practice, take the example of Anadarko's experience with the Deepwater Horizon disaster. Given that BP had settled its claims against its other joint venture partner, MOEX, well before the Anadarko settlement,[92] it is prudent to assume that settlement negotiations between Anadarko and BP had commenced by the time the June 30th 10-Q was released. Thus, under the proposed disclosure rules the minimum value of losses Anadarko would have had to report would have been the highest value of settlement they had offered to BP in the course of those negotiations. As settlement negotiations had likely already begun, this number would not be the zero Anadarko reported in reality[93] but instead a number that more closely reflected the value of the settlement actually reached. Of course the value of the final settlement would not map precisely to the settlement values disclosed in periodic filings; the proposed system, however, would get closer to the actual number than allowing companies the easy out of declining to recognize *any* amount as more likely than zero. Thus, the application of the proposal in a hypothetical situation like Anadarko's would have provided investors a more accurate advance warning of the large settlement that was to come.

IV. Practical Concerns

A. *Obstacles*

Prior attempts at reforming the reporting of contingent liabilities have failed, with accountants and lawyers proving unable to reach a meaningful consensus on the issue.[94] Dissatisfaction with the accounting rules for contingent-litigation liabilities is nothing new. Indeed, as early as 1978 there were rumblings of discontent and pushes for reform from some in the accounting profession.[95] The most recent attempt at reform came from the FASB beginning in 2008. The reforms were proposed for comment in two different "exposure drafts."[96] The fate of the FASB's attempted changes is instructive because it illustrates the main concerns of both sides of the debate and just how intractable a solution might be. Under the first exposure draft, the concerns previously discussed regarding SFAS 5 were largely addressed. For example, a major change in the exposure draft

92. Werdigier, *supra* note 63.
93. *See supra* notes 69–71 and accompanying text.
94. *See supra* notes 31–33 and accompanying text.
95. Fesler & Hagler, *supra* note 22, at 10.
96. FASB EXPOSURE DRAFT 1, *supra* note 3; FASB EXPOSURE DRAFT 2, *supra* note 91.

involved advancing the position that financial-statement users "prefer[red] to have a highly uncertain estimate supplemented with a qualitative description [rather] than no quantification of a potential loss as commonly occurs in existing practice."[97] Additionally, the draft required tabular disclosure of the amount a company accrued internally in order to allow users to better see the changes in this information from period to period.[98] Further, anticipating corporate criticism on the matter, the draft permitted companies to aggregate their litigation contingencies in order to minimize the risk that those contingencies could be used against them.[99] This was done to avoid exactly the type of reverse engineering discussed previously.[100] Corporate counsel still vehemently opposed the reforms.[101]

The aim of the FASB amendments was clear: to provide a window into management's valuation of the case. In the face of strong opposition from issuers of financial statements and their counsel, however, the FASB voted on July 9, 2012, to remove the issue from its agenda, having made no changes to the accounting standard.[102] Eventually, the various stakeholders met and a consensus emerged that "[d]isclosures about litigation contingencies should focus on the contentions of the parties, rather than predictions about the future outcome."[103] The problem with this consensus should be apparent: the contentions of the parties are already public record; they are called pleadings.[104] Meaningful improvement in disclosure must make nonpublic information available to investors. My proposed solution provides such nonpublic information by offering a window into the settlement process. The proposed FASB changes were abandoned because they did not strike the proper balance.[105] Given the backlash from corporate counsel against the FASB exposure drafts, it is unlikely that companies will quietly allow their true subjective evaluations to be disclosed to the public. This is where disclosure of settlement offers can strike a middle ground between the two camps.

---

97. FASB EXPOSURE DRAFT 1, *supra* note 3, at 11.

98. *Id.* at 12.

99. *Id.*

100. *See supra* notes 44–49 and accompanying text.

101. FIN. ACCOUNTING STANDARDS BD., DISCLOSURE OF CERTAIN LOSS CONTINGENCIES COMMENT LETTER SUMMARY 2 (2010), *available at* http://www.fasb.org/cs/BlobServer?blobkey=id&blobwhere=1175821780488&blobheader=application%2Fpdf&blobcol=urldata&blobtable=MungoBlobs (noting that 289 out of 339 commenters did not support the changes proposed).

102. *Project Updates*, *supra* note 3.

103. *Summary of Board Decisions*, FIN. ACCT. STANDARDS BOARD, http://www.fasb.org/cs/ContentServer?c=FASBContent_C&pagename=FASB%2FFASBContent_C%2FActionAlertPage&cid=1176156416898.

104. *See supra* note 31 and accompanying text.

105. *See supra* text accompanying notes 94–101.

The FASB's initial exposure draft for changes to SFAS 5 did (at the request of financial-statement users) at least seek comments on reflecting proposed settlement values on financial statements, but the FASB seemed to have already decided against the issue because "often [settlement] offers expire quickly and may not reflect the status of negotiations only a short time later."[106] The decision not to proceed along these lines was largely supported by commenters,[107] but the comments on this topic may well have been the result of a generalized knee-jerk reaction to the exposure draft as a whole. As has been discussed, a vast majority of respondents reacted negatively to the initial exposure draft.[108] I suggest that the FASB threw the baby out with the bathwater in this respect. This is especially likely given that just over half the commenters were preparing companies (and their lawyers) that, as has been discussed, desire to disclose as little information as possible. Rather than attempt a sea change as the FASB did with its most recent foray into this area, a more prudent approach (and one I advocate in this Note) is incremental change.

A small handful of commenters supported the idea of using settlement values to drive disclosure of uncertain litigation contingencies. These commenters recognized the ability of settlement numbers to effectively establish upper and lower bounds of a potential loss even when such losses would be extremely difficult to quantify otherwise.[109] Unlike these commenters, I do not propose allowing settlement offers from plaintiffs to set the high end of the potential loss spectrum. Such a system skews the bargaining dynamics too far in favor of the plaintiff and could possibly push parties away from settlement. Putting plaintiffs in a position of being able to dictate the high end of potential losses disclosed (or in some cases accrued) by defendants would be a powerful piece of leverage in the hands of negotiators.

In response to the failure to tighten auditing requirements, the SEC has recently ramped up the rhetoric over more vigorous enforcement of existing rules. In a 2011 speech, the SEC's Corporation Finance Division's Chief Accountant Wayne Carnall warned of increased SEC scrutiny in this area, with particular emphasis on the MD&A.[110] Similarly, SEC Chairman Mary Schapiro noted in 2011:

---

106. FASB EXPOSURE DRAFT 1, *supra* note 3, at iii.

107. FIN. ACCOUNTING STANDARDS BD., DISCLOSURE OF CERTAIN LOSS CONTINGENCIES FINAL COMMENT LETTER SUMMARY 18 (2009), *available at* http://www.fasb.org/cs/BlobServer?blobkey=id&blobwhere=1175819438449&blobheader=application%2Fpdf&blobcol=urldata&blobtable=MungoBlobs.

108. *See supra* note 101 and accompanying text.

109. *See, e.g.*, ITAC Comment Letter, *supra* note 78 (stating that settlement offers should be disclosed and noting that they provide boundaries for the amount of the potential loss).

110. *See NYC Bar Program, "Litigation Contingency Reporting: Where Are We with FASB and the SEC?" New FASB Exposure Draft and SEC Scrutiny*, METROPOLITAN CORP. COUNS., Apr. 2011, at 35.

> [O]ur disclosure teams are asking institutions to clarify their exposure to potential losses due to litigation .... In the past, companies have often claimed that they were unable to accurately calculate their exposure, or they failed altogether to provide this information-arguing that doing so would prejudice their positions. We are asking that they begin providing this information if they have not been already, and that they ensure they refine their calculations over time as events and circumstances change and new information is obtained. These are not new requirements-they are currently what the accounting standard requires.[111]

It remains to be seen whether this heightened rhetoric is anything more than empty bluster. There is a big difference between SEC staff making statements on an issue and actually dedicating a portion of their extremely limited enforcement resources to these matters.[112]

The problem with this increased scrutiny (even if it comes to fruition) is that it will likely not solve the problem. The losses involved really cannot be reasonably estimated. Anadarko truly did not know how much it would have to pay as a result of the Deepwater Horizon disaster.[113] The challenge, instead, is developing some way of cutting through the veil of secrecy necessary to an adversarial justice system to discover the company's true assessment of its liability while avoiding a prejudicial effect on the company. The disclosure of settlement-offer values offers just such an opportunity.

In addition to overcoming the difficult history of reform in this area, another stumbling block in any reform attempt, or even enhanced enforcement of the disclosure rules related to litigation contingencies, is the tangled regulatory situation. The groups with a part to play in disclosure in this area include an alphabet soup of the SEC, the FASB, the ABA, and the Public Company Accounting Oversight Board (PCAOB). As one commentator from the accounting world put it: "Litigation contingencies are somewhat unique in that their disclosure is affected by three groups: accountants, attorneys, and federal regulators."[114] The SEC is given

---

111. Mary L. Schapiro, Chairman, U.S. Sec. & Exch. Comm'n, Remarks Before the Women in Finance Symposium (July 12, 2011), *available at* http://www.sec.gov/news/speech/2011/spch071211mls.htm.

112. *See* Ross MacDonald, Note, *Setting Examples, Not Settling: Toward a New SEC Enforcement Paradigm*, 91 TEXAS L. REV. 419, 443 & n.170 (2012) (discussing the SEC's limited resources).

113. *See* John Schwartz, *Liability Questions Loom for BP and Ex-Partners*, N.Y. TIMES, June 24, 2010, http://www.nytimes.com/2010/06/25/us/25liability.html?_r=0 (discussing "the legal endgame of sorting out the final tab among the companies that owned the well and worked on the Deepwater Horizon rig"); *see also* Ronen Perry, *The* Deepwater Horizon *Oil Spill and the Limits of Civil Liability*, 86 WASH. L. REV. 1, 50 (2011) (declaring that Anadarko may "[t]echnically" be held liable for the spill).

114. Fesler & Hagler, *supra* note 22.

statutory authority to dictate what accounting standards govern these financial statements.[115] Although the SEC has promulgated Regulation S-X to address accounting issues, these rules generally prescribe only the form of disclosure rather than the substance and standards of the financial statements.[116] Instead of directly articulating the standards for these financial statements, the SEC has long deferred to the FASB to promulgate GAAP.[117] As the FASB's website puts it, "the [SEC]'s policy has been to rely on the private sector for this function to the extent that the private sector demonstrates ability to fulfill the responsibility in the public interest."[118] The FASB is a private entity formed in 1973 to "establish[] standards of financial accounting that govern the preparation of financial reports by nongovernmental entities."[119] Though the SEC does not directly oversee the FASB or appoint its members, the fact that the SEC has the ultimate authority to set out GAAP undoubtedly exerts considerable influence over the FASB.[120]

The PCAOB came into existence with the Sarbanes-Oxley Act due to "deep failings in the U.S. accounting profession's ability to regulate itself"[121] and is tasked primarily with overseeing the audit firms that review the financial statements of reporting companies.[122] The creation of the PCAOB took the luxury of self-regulation away from the accounting profession.[123] Instead, the members of the PCAOB are appointed by the SEC in consultation with the Federal Reserve Board of Governors and the Secretary of the Treasury.[124] As an additional measure of control, the SEC oversees the PCAOB's budget.[125]

---

115. Securities Exchange Act of 1934 § 13(b), 15 U.S.C. § 78m(b)(1) (2012); *Facts About FASB*, *supra* note 2.

116. Arthur Acevedo, *The Fox and the Ostrich: Is GAAP a Game of Winks and Nods?* 12 TRANSACTIONS: TENN. J. BUS. L. 63, 91 (2010).

117. *See* Commission Statement of Policy Reaffirming the Status of the FASB as a Designated Private-Sector Standard Setter, Securities Act Release No. 8221, Exchange Act Release No. 47,743, Investment Company Act Release No. 26,028, 68 Fed. Reg. 23,333, 23,333 (May 1, 2003) (continuing to acknowledge that FASB standards reflect "generally accepted" accounting standards); *see also* Koprowski et al., *supra* note 27, at 441–42 (discussing the history of the FASB as the accounting-standard setter).

118. *Facts about FASB*, *supra* note 2.

119. *Id.*

120. *Cf.* Acevedo, *supra* note 116, at 126–27 (calling for more government control over accounting standards).

121. Paul S. Atkins, Comm'r, U.S. Sec. & Exch. Comm'n, Statement Before the Open Meeting Regarding PCAOB and FASB Budget Review (Mar. 3, 2005), *available at* http://www.sec.gov/news/speech/spch030305psa3.htm.

122. *See About the PCAOB*, PUB. COMPANY ACCT. OVERSIGHT BOARD (2014), http://pcaobus.org/About/Pages/default.aspx ("The PCAOB is a nonprofit corporation established by Congress to oversee the audits of public companies . . . .").

123. *See id.* (explaining that public-company auditors were previously self-regulated).

124. *Id.*

125. Atkins, *supra* note 121.

With so many different organizations and so many different constituencies, any reform must be able to appease all groups. By seeking a compromise position between accountants and lawyers, the proposed disclosure of settlement offer values might be able to generate sufficient buy-in from these diverse groups to gain acceptance. No proposal to solve the current disclosure problem can fully appease all parties. Corporations, as evidenced by their harsh reaction to the FASB exposure drafts, would undoubtedly prefer to maintain the status quo. Using settlement offers instead of subjective valuations of cases will allay many of the concerns that corporate counsel have expressed about the disclosures prejudicing ongoing cases. Such information is useless to the other side in active litigation, as it has necessarily been communicated to them. The disclosed number may establish a floor in negotiations, but that floor has been chosen by the disclosing company. Rules of evidence would similarly prevent disclosed information from prejudicing a company in court.[126] On the other hand, these numbers are of considerable value to investors as they are not available in the public realm. Unlike the consensus reached that disclosures should focus on the contentions of the parties, disclosure of settlement offers gives investors a new window into the settlement process that they previously lacked.

B. Objections

Perhaps the clearest objection to the use of settlement numbers in a disclosure setting is that the proposal does not appreciate the interests of litigants in keeping the settlement process confidential. Defendants have very real economic incentives to keep settlements confidential.[127] These incentives include reputational concerns and the likelihood of costly "copycat" lawsuits.[128] A related concern is the ability to reverse engineer the disclosure to determine how much money a particular case was settled for. Such knowledge could be useful for subsequent plaintiffs looking for insight into a company's negotiating strategy.

Several factors, however, push against concerns that settlement confidentiality will be lost. First is the simple fact that once the suits are filed the allegations are public record and have been disclosed in pleadings. Furthermore, other reporting requirements already call for disclosure of factual information about the case.[129] Once this information is in the public realm, a proposed settlement value alone can likely do little harm.

---

126. *See* FED. R. EVID. 408(a) (excluding any statement made as part of compromise negotiations).

127. *See* Moss, *supra* note 83, at 878–80 (demonstrating that confidential settlements increase settlement range because of the value defendants place on confidentiality).

128. *Id.* at 878, 879 n.58.

129. *See* SEC Regulation S-K, Item 103, 17 C.F.R. § 229.103 (2013) (requiring factual disclosure about material litigation).

Additionally, the aggregation of suits on the balance sheet will help to preserve the desired confidentiality to a point. Though the comment letters against the FASB draft demonstrate the shortcomings of the aggregation principle in this respect,[130] reverse engineering reported settlement figures would be considerably less dangerous than reverse engineering subjective case valuations. On balance, the interests in confidentiality of settlement negotiations and agreements are outweighed by the importance of improving the disclosure system for investors.

A related concern is that many settlement offers include nondisclosure provisions. These concerns were raised by some commenters to the initial FASB exposure draft.[131] This fact, however, need not kill the proposal before it gets off the ground, as some commentators suggest. Instead, the nondisclosure provisions could simply be redrafted to take into account the fact that a party may need to disclose the offered value as part of its securities filings. As discussed, the aggregation of these values should obviate most concerns about reverse engineering the disclosures to determine information about a particular settlement negotiation.

Another concern is whether Federal Rule of Evidence 408 would apply broadly enough to keep proposed settlement value disclosure out of trial proceedings. As has been discussed, if such numbers were admissible in court it would seriously prejudice the cases of reporting companies.[132] Though the contents of the negotiations themselves clearly fall within Rule 408, disclosure to outside auditors might fall outside of this rule. There are two key prerequisites to invoking Rule 408: (1) there must be a disputed claim and (2) the Rule only applies to "communications made or conduct that occurs 'in compromise negotiations.'"[133] While the first requirement easily applies to negotiations of pending-litigation settlement, it is possible that audit-inquiry responses from attorneys could be found not to be made "in compromise negotiations." A possible alternative basis for excluding disclosures of settlement offers in the audit process would be Rule 403, which allows exclusion of anything that, although "relevant[,] . . . [has] probative value [that] is substantially outweighed by a danger of . . . unfair prejudice, confusing the issues [or] misleading the jury."[134] Though an in-depth analysis of the Federal Rules of Evidence is outside the scope of this Note, consideration of the evidentiary questions raised by the proposed disclosure system would be necessary in order for the desired benefits to be realized.

---

130. *See supra* note 101 and accompanying text.

131. FIN. ACCOUNTING STANDARDS BD., *supra* note 107, at 19.

132. *See supra* notes 29–49 and accompanying text.

133. Wayne D. Brazil, *Protecting the Confidentiality of Settlement Negotiations*, 39 HASTINGS L.J. 955, 960 (1988).

134. FED. R. EVID. 403.

One concern with causing companies to take a hit on their balance sheets when making offers to settle pending litigation is that it may push management away from settling cases. Such concerns, however, overlook the very real benefits of settling a case (particularly one large enough to get over the materiality hurdle). Of more concern, however, is the possibility that requiring accrual of offers made in the course of settlement negotiations will distort the negotiation process. There is no doubt that plaintiffs will enter negotiations armed with the knowledge that any offer made by the defendants will have consequences that resound far beyond the negotiating table. Defendants will know this too. This might incentivize defendants to be more hesitant in negotiations in order to protect the company's balance sheet in the short term, perhaps at the expense of longer term interests. It is conceivable, given the incentives for management to keep earnings per share in line with expectations, that settlement negotiations could be seriously and negatively affected.[135]

A key concern of any new rule is the ease with which such a rule can be circumvented or "gamed." Disclosure of proposed settlement values is not without concerns of this type. Though such actions would likely run afoul of antifraud provisions of the securities laws,[136] enforcement would be difficult because beginning negotiations with low numbers is a legitimate negotiating tactic that might be indistinguishable from an attempt to game the disclosure system.[137] Companies might invariably make only low settlement offers in order to avoid large disclosures on their balance sheets. If companies refuse to allow their settlement offers to rise above an inexplicably low level, however, they will be unable to settle their cases and take advantage of the numerous benefits of settlement discussed above.[138] Particularly in cases large enough to threaten the company as a going concern, the dangers of not settling likely outweigh any incentive management would have to game the disclosure system by offering only extremely low settlement values.

Another potential source for gaming a rule requiring disclosure of settlement offers is the commonplace confidentiality norms in the context of mediation.[139] Though an in-depth analysis of the diverse laws and court

135. For a discussion of short-termism from different market participants, see generally Lynne L. Dallas, *Short-Termism, the Financial Crisis, and Corporate Governance*, 37 J. CORP. L. 265 (2012).

136. *See* SEC Rule 10b-5, 17 C.F.R. § 240.10b-5 (2013) (prohibiting untrue statements or omissions of material fact).

137. *See* Robert S. Adler & Elliot M. Silverstein, *When David Meets Goliath: Dealing with Power Differentials in Negotiations*, 5 HARV. NEGOT. L. REV. 1, 75–76 (2000) (noting that there is "virtual unanimity among the experts" that an initial offer should be as high or low as reasonably possible).

138. *See supra* note 82 and accompanying text.

139. For a discussion of mediation confidentiality, see Sarah Rudolph Cole, *Protecting Confidentiality in Mediation: A Promise Unfulfilled?*, 54 U. KAN. L. REV. 1419 (2006).

rules regarding mediation confidentiality is outside the scope of this Note, a workable disclosure rule cannot allow itself to be avoided simply by taking a case to mediation and thereby cloaking the entire process in a shroud of confidentiality. Such a rule would do little to provide investors meaningful information, though it would be a boon to mediators. The Uniform Mediation Act creates a privilege that keeps information revealed in mediation from being admitted as evidence or discovered in a court proceeding[140] but has an exception if such information is required to be disclosed by law.[141] A requirement of disclosure of settlement offers by the securities laws would presumably bring such an exception into play. Requiring disclosure of these values would not harm the mediation process because the rationale for the normal rules of strict confidentiality is not implicated by a company disclosing its own willingness to settle a case.[142] Because the proposal does not require any disclosure of offers made by the other side, the concerns implicated by mediation confidentiality fail to surface.

Another practical concern involves the potential unreliability of offered settlement values as an indicator of potential outcomes. It could be that using the numbers thrown about in settlement negotiations is not a better system for valuing litigation than allowing companies to consider zero just as likely an outcome as any other. A study conducted by Professor Gerald Williams provides a telling illustration in this regard.[143] Williams asked fourteen pairs of practicing attorneys to negotiate a settlement of the same personal injury case.[144] The resulting settlements ranged from $15,000 to $95,000.[145] Additionally, the defendant's opening demands (a number that would be particularly relevant under the proposed disclosure system) ranged from $3,000 to $50,000.[146] It is difficult to extrapolate these results into the context of the complex commercial litigation that would trigger disclosure, but the variance may well be larger in those more complex and difficult-to-value situations. Of course the disparity between opening demands may merely reflect differing negotiation strategies on the parts of various lawyers.[147] Despite the difficulty and large probable variances in valuation from firm to firm, however, it must still be borne in mind that the current system essentially allows companies to punt and give

140. UNIFORM MEDIATION ACT § 4 (2001).

141. *Id.* § 6(a)(2).

142. *See* Cole, *supra* note 139, at 1419 (citing fear that things said in mediation would later be used against a party as the rationale for mediation confidentiality).

143. *See* GERALD R. WILLIAMS, LEGAL NEGOTIATION AND SETTLEMENT 6–7 (1983) (providing the results of an experiment regarding lawyer negotiating outcomes).

144. *Id.* at 6.

145. *Id.*

146. *Id.* at 7 tbl.1–1.

147. *See id.* at 73–77 (discussing various positioning strategies in settlement negotiations).

investors no hint of what a likely settlement value might be. In the disclosure context, surely a company's best guess at an opening settlement value is of more use to investors than no information at all.

The FASB discounted the idea of settlement offer disclosure before even releasing its first exposure draft because offers in settlement negotiation might change rapidly and might not reflect the reality on the ground.[148] These concerns, however, are likely overblown. The period between when initial negotiating positions are staked out and when the parties start to close in on an agreement is often measured in months if not years.[149] Indeed, it often takes the looming threat of trial to push parties to proceed beyond their initial negotiation postures.[150] The FASB's concerns, though perhaps valid on their face, do not appear to take into account the dynamics of bargaining in the shadow of the adjudicatory system.

This is not to suggest, however, that the concerns expressed in the exposure draft over the short-term significance of settlement numbers are wholly without merit. Indeed, disclosure of settlement offers implicates the potential ongoing duty of issuers to update their disclosures with material information. Though the contours of this duty vary from jurisdiction to jurisdiction, most circuits have adopted some duty to update in certain circumstances.[151] This duty can be largely avoided, particularly in the contingent-liability area, if "the initial statements are accompanied by sufficient cautionary language warning investors that the statements should not be interpreted as having any implicit representation regarding subsequent events or future disclosures by the company."[152] Thus, capable drafting of settlement value disclosures should be able to avoid a duty to update under the securities laws. Because periodic filings are made every quarter, significant changes in settlement negotiations could occur from one filing to the next. The Anadarko example is again illustrative. Though this Note has largely assumed that settlement negotiations had already begun by the time the 10-Q in question had been filed, it is conceivable that such

148. FASB EXPOSURE DRAFT 1, *supra* note 3, at iii.

149. *See* WILLIAMS, *supra* note 143, at 78.

150. *Id.*

151. *See* ERIC R. SMITH ET AL., DUTY TO UPDATE PREVIOUSLY DISCLOSED INFORMATION 2–4 (2011), *available at* http://www.venable.com/files/Publication/d90ad0bd-0947-4956-aa70-1026f1ac03be/Presentation/PublicationAttachment/cf3d0d7f-ca04-4b19-96e1-1510529d9821/Duty_to_Update_Previously_Disclosed_Information.pdf (examining the approaches various circuits have taken with the duty to update). Indeed, only the Seventh Circuit has consistently resisted imposing any duty to update on issuers. *Id.* at 4.

152. *See* JONI S. JACOBSEN ET AL., DISCLOSURE DUTIES ARISING UNDER SECTION 10(B): WHEN TO CORRECT OR UPDATE 11 (2011), *available at* http://www.kattenlaw.com/files/upload/Disclosure_Duties_Arising_Under_Section_10B_When_To_Correct_Or_Update.pdf (discussing ways to negate the duty to update); SMITH ET AL., *supra* note 151, at 4 (recommending nonboilerplate cautionary language be included in disclosure documents to avoid a duty to update).

negotiations had actually not yet begun. If this were the case, the proposed solution would have generated the same outcome as reality. Nevertheless, it is important to avoid letting the perfect be the enemy of the good. Continuous disclosure on all fronts would be a far too cumbersome system to be workable, and well-drafted disclosures should inform investors that volatile information like settlement offers is subject to change.

Though there are a number of objections that can be raised to the disclosure of offered settlement values, on balance the proposed system can accomplish the goal of bridging the current gap between accountants and lawyers.

V. Conclusion

In sum, the current system of litigation contingency reporting does not adequately protect investors from large, unexpected settlements and judgments. By turning the focus on a company's attempt to settle the large claims against it, the information gap can be partially filled. Though the proposal outlined in this Note does not alleviate the problem completely, it walks the fine line between greater protection of investors and protecting the reporting companies' interests in ongoing and future litigation.

Any proposed solution to the problem of disclosure in this area must successfully address the concerns of corporations discussed above. Additionally, though, any proposal must also require companies to reveal some degree of nonpublic information. The numbers introduced in settlement negotiations meet the requirements of both sides of this cultural divide. Though a number of practical obstacles and concerns must be overcome in order for disclosure of settlement offers to function as a viable proxy for subjective valuations by a reporting company, the interests of investors and the need for transparency in the financial system call for a workable compromise on the issue. Disclosure of settlement offer values can be that compromise.

—Jamie L. Yarbrough

* * *

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