

Texas Law Review

PATENTABLE SUBJECT MATTER AND INSTITUTIONAL CHOICE

John M. Golden

SHARE REPURCHASES, EQUITY ISSUANCES, AND THE
OPTIMAL DESIGN OF EXECUTIVE PAY

Jesse M. Fried

BOOK REVIEW—BUT I THOUGHT THE EARTH BELONGED TO THE LIVING

Mark L. Ascher

A PROPOSED WEALTH REDISTRIBUTION SYSTEM BASED ON THE UNDERLYING
PREMISE OF REVENUE SHARING IN AMERICAN PRO SPORTS

THIRD PARTY ASSISTANCE IN DETERMINING OBVIOUSNESS

THE EFFECT OF THE PARLIAMENTARY VOTING SYSTEM ACT
ON THE BRITISH CONSTITUTION

Texas Law Review

A national journal published seven times a year

Recent and Forthcoming Articles of Interest
visit www.texasrev.com for more on recent articles

THE CONSENSUS CONSTITUTION

Justin Driver

March 2011

AMICUS BRIEFS AND THE SHERMAN ACT: WHY ANTITRUST NEEDS A NEW DEAL

Rebecca Haw

April 2011

Individual issue rate: \$15.00 per copy

Subscriptions: \$47.00 (seven issues)

Order from:

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

<http://www.texaslawpublications.com>

Texas Law Review *See Also*

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

“TWO PATHS, ONE RESULT”: A (HEAVILY QUALIFIED) DEFENSE OF CONSENSUS
CONSTITUTIONALISM

Mark Tushnet

TEXAS LAW REVIEW ASSOCIATION

OFFICERS

President-Elect

JAMES A. HEMPHILL
Treasurer

HON. DIANE P. WOOD
President

ANDREA L. FAIR
Executive Director

CARRIN F. PATMAN
Immediate Past President

BOARD OF DIRECTORS

NINA CORTELL
ALISTAIR B. DAWSON
KARL G. DIAL
GARY L. EWELL
STEPHEN FINK
R. JAMES GEORGE, JR.

DIANA M. HUDSON
BRANDON C. JANES
JEFFREY C. KUBIN
D. MCNEEL LANE
LEWIS T. LECLAIR
JOHN B. MCKNIGHT

MICHAEL H. NEWMAN
ERIC J.R. NICHOLS
ELLEN PRYOR
DAVID M. RODI
REAGAN W. SIMPSON
MARK L.D. WAWRO

SCOTT J. ATLAS, *ex officio Director*
OMAR OCHOA, *ex officio Director*

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

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

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

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

The *Texas Law Review* is pleased to consider unsolicited manuscripts for publication but regrets that it cannot return them. Please submit a single-spaced manuscript, printed on one side only, with footnotes rather than endnotes. Citations should conform with the *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* (11th ed. 2008), *The Chicago Manual of Style* (15th ed. 2003), and Bryan A. Garner, *A Dictionary of Modern Legal Usage* (2d ed. 1995).

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

© Copyright 2011, Texas Law Review Association

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

THE UNIVERSITY OF TEXAS SCHOOL OF LAW

ADMINISTRATIVE OFFICERS

LAWRENCE G. SAGER, B.A., LL.B.; *Dean, John Jeffers Research Chair in Law, Alice Jane Drysdale Sheffield Regents Chair.*
MECHELE DICKERSON, B.A., J.D.; *Associate Dean for Academic Affairs, Arthur L. Moller Chair in Bankruptcy Law and Practice.*
ALEXANDRA W. ALBRIGHT, B.A., J.D.; *Associate Dean for Administrative Services, Senior Lecturer.*
EDEN E. HARRINGTON, B.A., J.D.; *Assist. Dean for Clinical Ed. & Public Serv., Dir. of William Wayne Justice Ctr. for Public Interest Law.*
KIMBERLY L. BIAR, B.B.A.; *Assistant Dean for Financial Affairs, Certified Public Accountant.*
CARLA COOPER, B.A., M.A., Ph.D.; *Assistant Dean for Alumni Relations and Development.*
MICHAEL J. ESPOSITO, B.A., J.D., M.B.A.; *Assistant Dean for Continuing Legal Education.*
KIRSTON FORTUNE, B.F.A.; *Assistant Dean for Communications.*
MICHAEL HARVEY; *Assistant Dean for Technology.*
MONICA K. INGRAM, B.A., J.D.; *Assistant Dean for Admissions and Financial Aid.*
DAVID A. MONTROYA, B.A., J.D.; *Assistant Dean for Career Services.*
LESLIE OSTER, A.B., J.D.; *Assistant Dean for Strategic Planning, Senior Lecturer.*
REYMUNDO RAMOS, B.A.; *Assistant Dean for Student Affairs.*

FACULTY EMERITI

HANS W. 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.*
JOHN F. SUTTON, JR., J.D.; *A.W. Walker Centennial Chair Emeritus.*
JAMES M. TREECE, B.A., J.D., M.A.; *Charles I. Francis Professor Emeritus in Law.*
RUSSELL J. WEINTRAUB, B.A., J.D.; *Ben H. & Kitty King Powell Chair Emeritus in Business and Commercial Law.*

PROFESSORS

JEFFREY B. ABRAMSON, B.A., J.D., Ph.D.; *Professor of Law and Government.*
DAVID E. ADELMAN, B.A., Ph.D., J.D.; *Harry Reasoner Regents Chair in Law.*
DAVID A. ANDERSON, A.B., J.D.; *Fred and Emily Marshall Wulff Centennial Chair in Law.*
MARK L. ASCHER, B.A., M.A., J.D., LL.M.; *Joseph D. Jamail Centennial Chair in Law.*
RONEN AVRAHAM, M.B.A., LL.B., LL.M., S.J.D.; *Thomas Shelton Maxey Professor in Law.*
LYNN A. BAKER, B.A., B.A., J.D.; *Frederick M. Baron Chair in Law, Co-Director of Center on Lawyers, Civil Justice, and the Media.*
MITCHELL N. BERMAN, A.B., M.A., J.D.; *Richard Dale Endowed Chair in Law.*
BARBARA A. BINTLIFF, M.A., J.D.; *Joseph C. Hutcheson Professor in Law.*
LYNN E. BLAIS, A.B., J.D.; *Leroy G. Denman, Jr. Regents Professor in Real Property Law.*
ROBERT G. BONE, B.A., J.D.; *G. Rollie White Teaching Excellence Chair in Law.*
OREN BRACHA, LL.B., S.J.D.; *Howrey LLP and Arnold, White, & Durkee Centennial Professor.*
NORMA V. CANTU, B.A., J.D.; *Professor of Law and Education.*
LOFTUS C. CARSON, II, B.S., M. Pub. Aff., M.B.A., J.D.; *Ronald D. Krist Professor.*
ROBERT M. CHESNEY, B.S., J.D.; *Charles I. Francis Professor in Law.*
MICHAEL J. CHURGIN, A.B., J.D.; *Raybourne Thompson Centennial Professor.*
JANE M. COHEN, B.A., J.D.; *Edward Clark Centennial Professor.*
FRANK B. CROSS, B.A., J.D.; *Herbert D. Kelleher Centennial Professor of Business Law.*
WILLIAM H. CUNNINGHAM, B.A., M.B.A., Ph.D.; *James L. Bayless Chair for Free Enterprise.*
JENS C. DAMMANN, J.D., LL.M., Dr. Jur., J.S.D.; *William Stamps Farish Professor in Law.*
JOHN DEIGH, B.A., M.A., Ph.D.; *Professor of Law and Philosophy.*
GEORGE E. DIX, B.A., J.D.; *George R. Killam, Jr. Chair of Criminal Law.*
JOHN S. DZIENKOWSKI, B.B.A., J.D.; *Dean John F. Sutton, Jr. Chair in Lawyering and the Legal Process.*
KAREN L. ENGLE, B.A., J.D.; *Cecil D. Redford Prof. in Law, Director of Bernard & Audre Rapoport Center for Human Rights & Justice.*
KENNETH FLAMM, Ph.D.; *Professor.*
WILLIAM E. FORBATH, A.B., B.A., Ph.D., J.D.; *Lloyd M. Bentsen Chair in Law.*
JULIUS G. GETMAN, B.A., LL.B., LL.M.; *Earl E. Sheffield Regents Chair.*
JOHN M. GOLDEN, A.B., J.D., Ph.D.; *Loomer Family Professor in Law.*
STEVEN GOODE, B.A., J.D.; *W. James Kronzer Chair in Trial and Appellate Advocacy, University Distinguished Teaching Professor.*
LINO A. GRAGLIA, B.A., LL.B.; *A. Dalton Cross Professor.*
CHARLES G. GROAT, B.A., M.S., Ph.D.; *Professor.*
PATRICIA I. HANSEN, A.B., M.P.A., J.D.; *J. Waddy Bullion Professor.*
HENRY T. HU, B.S., M.A., J.D.; *Allan Shivers Chair in the Law of Banking and Finance.*
BOBBY R. INMAN, B.A.; *Lyndon B. Johnson Centennial Chair in National Policy.*
DEREK P. JINKS, B.A., M.A., J.D.; *The Marrs McLean Professor in Law.*
STANLEY M. JOHANSON, B.S., LL.B., LL.M.; *James A. Elkins Centennial Chair in Law, University Distinguished Teaching Professor.*
CALVIN H. JOHNSON, B.A., J.D.; *Andrews & Kurth Centennial Professor.*
EMILY E. KADENS, B.A., M.A., Dipl. M.A., Ph.D., J.D.; *Baker and Botts Professor in Law.*
SUSAN R. KLEIN, B.A., J.D.; *Alice McKean Young Regents Chair in Law.*
SANFORD V. LEVINSON, A.B., Ph.D., J.D.; *W. St. John Garwood & W. St. John Garwood, Jr. Centennial Chair in Law, Professor of Gov't.*
STEFANIE A. LINDQUIST, B.A., J.D., Ph.D.; *A.W. Walker Centennial Chair in Law.*
VIJAY MAHAJAN, M.S.Ch.E., Ph.D.; *John P. Harbin Centennial Chair in Business.*

BASIL S. MARKESINIS, LL.B., LL.D., DCL, Ph.D.; *Jamail Regents Chair*.
 INGA MARKOVITS, LL.M.; *"The Friends of Joe Jamail" Regents Chair*.
 RICHARD S. MARKOVITS, B.A., LL.B., Ph.D.; *John B. Connally Chair*.
 ALOYSIUS P. MARTINICH, B.A., M.A., Ph.D.; *Professor*.
 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.; *Bartlett Cocke Regents Professor in Architecture*.
 LINDA S. MULLENIX, B.A., M. Phil., J.D., Ph.D.; *Morris & Rita Atlas Chair in Advocacy*.
 STEVEN P. NICHOLS, B.S.M.E., M.S.M.E., J.D., Ph.D.; *Clint W. Murchison, Sr. Fellow of Free Enterprise*.
 ROBERT J. PERONI, B.S.C., J.D., LL.M.; *The Fondren Foundation Centennial Chair for Faculty Excellence*.
 H. W. PERRY, JR., B.A., M.A., Ph.D.; *Associate Professor of Law and Government*.
 LUCAS A. POWE, JR., B.A., J.D.; *Anne Green Regents Chair in Law, Professor of Government*.
 WILLIAM C. POWERS, JR., B.A., J.D.; *President, The University of Texas at Austin, Hines H. Baker and Thelma Kelley Baker Chair, University Distinguished Teaching Professor*.
 DAVID M. RABBAN, B.A., J.D.; *Dahr Jamail, Randall Hage Jamail, & Robert Lee Jamail Reg. Chair, Univ. Distinguished Teaching Prof.*
 ALAN S. RAU, B.A., LL.B.; *Mark G. 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*.
 DANIEL B. RODRIGUEZ, B.A., J.D.; *Minerva House Drysdale Regents Chair in Law*.
 WILLIAM M. SAGE, A.B., M.D., J.D.; *Vice Provost for Health Affairs, James R. Dougherty Chair for Faculty Excellence*.
 JOHN J. SAMPSON, B.B.A., LL.B.; *William Benjamin Wynne Professor*.
 CHARLES M. SILVER, B.A., M.A., J.D.; *Roy W. and Eugenia C. MacDonald Endowed Chair in Civil Procedure, Professor in Government, Co-Director of Center on Lawyers, Civil Justice, and the Media*.
 ERNEST E. SMITH, B.A., LL.B.; *Rex G. Baker Centennial Chair in Natural Resources Law*.
 TARA A. SMITH, Ph.D.; *Professor*.
 LEWIS J. SPELLMAN, B.B.A., M.B.A., M.A., Ph.D.; *Professor*.
 JAMES C. SPINDLER, B.A., J.D.; *The Sylvan Lang Professor*.
 MATTHEW L. SPITZER, B.A., Ph.D., J.D.; *Hayden W. Head Regents Chair for Faculty Excellence*.
 JANE STAPLETON, B.S., Ph.D., LL.B., D. Phil.; *Ernest E. Smith Professor*.
 JORDAN M. STEIKER, B.A., J.D.; *Judge Robert M. Parker Endowed Chair in Law*.
 PAUL J. STEKLER, Ph.D.; *Professor*.
 SHIRLEY E. THOMPSON, A.B., A.M., Ph.D.; *Professor*.
 GERALD TORRES, A.B., J.D., LL.M.; *Bryant Smith Chair in Law*.
 GREGORY J. VINCENT, B.A., J.D., Ed.D.; *Professor*.
 WENDY E. WAGNER, B.A., M.E.S., J.D.; *Joe A. Worsham Centennial Professor*.
 LOUISE WEINBERG, A.B., J.D., LL.M.; *William B. Bates Chair for the Administration of Justice*.
 OLIN G. WELLBORN, A.B., J.D.; *William C. Liedtke, Sr. Professor*.
 JAY L. WESTBROOK, B.A., J.D.; *Benno C. Schmidt Chair of Business Law*.
 ABRAHAM L. WICKELGREN, A.B., Ph.D., J.D.; *Bernard J. Ward Professor in Law*.
 ZIPPORAH B. WISEMAN, B.A., M.A., LL.B.; *Thos. H. Law Centennial Professor in Law*.
 PATRICK WOOLLEY, A.B., J.D.; *Beck, Redden & Secrest Professor in Law*.

ASSISTANT PROFESSORS

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

SENIOR LECTURERS, WRITING LECTURERS, AND CLINICAL PROFESSORS

WILLIAM P. ALLISON, B.A., J.D.; <i>Clinical Prof., Dir. of Criminal Defense Clinic, Co-Dir. of Actual Innocence Clinic</i> .	JEANA A. LUNGWITZ, B.A., J.D.; <i>Clinical Professor, Director of Domestic Violence Clinic</i> .
MARJORIE I. BACHMAN, B.S., J.D.; <i>Clinical Instruct., Staff Attorney</i> .	TRACY W. MCCORMACK, B.A., J.D.; <i>Lect.; Dir. Trial Advoc. Prg.</i>
PHILIP C. BOBBITT, A.B., J.D., Ph.D.; <i>Distinguished Sr. Lecturer</i> .	ROBIN B. MEYER, B.A., M.A., J.D.; <i>Lecturer</i> .
KAMELA S. BRIDGES, B.A., B.J., J.D.; <i>Lecturer</i> .	JANE A. O'CONNELL, B.A., M.S., J.D.; <i>Lecturer, Associate Director for Patron Services, Instruction, and Research</i> .
CYNTHIA L. BRYANT, B.A., J.D.; <i>Clinical Professor, Director of Mediation Clinic</i> .	RANJANA NATARAJAN, B.A., J.D.; <i>Clinical Professor, Director of National Security Clinic</i> .
PAUL J. BURKA, B.A., LL.B.; <i>Senior Lecturer</i> .	ROBERT C. OWEN, A.B., M.A., J.D.; <i>Clinical Professor</i> .
MARY R. CROUTER, A.B., J.D.; <i>Lecturer, Assistant Director of William Wayne Justice Center for Public Interest Law</i> .	SEAN J. PETRIE, B.A., J.D.; <i>Lecturer</i> .
TIFFANY J. DOWLING, B.A., J.D.; <i>Clinical Instructor, Staff Attorney</i> .	WAYNE SCHIESS, B.A., J.D.; <i>Sr. Lecturer, Dir. of Legal Writing</i> .
LORI K. DUKE, B.A., J.D.; <i>Clinical Professor</i> .	STACY ROGERS SHARP, B.S., J.D.; <i>Lecturer</i> .
ARIEL E. DULITZKY, J.D., LL.M.; <i>Clinical Professor</i> .	PAMELA J. SIGMAN, B.A., J.D.; <i>Lect., Dir. Juvenile Justice Clinic</i> .
ELANA S. EINHORN, B.A., J.D.; <i>Lecturer</i> .	DAVID S. SOKOLOV, B.A., M.A., J.D., M.B.A.; <i>Distinguished Senior Lecturer, Director of Student Life</i> .
TINA V. FERNANDEZ, A.B., J.D.; <i>Lecturer</i> .	LESLIE L. STRAUCH, B.A., J.D.; <i>Clinical Prof. Child. Rights Clinic</i> .
ELIZABETH FRUMKIN, B.A., M.A., J.D.; <i>Lecturer</i> .	GRETCHEN S. SWEEN, B.A., M.A., Ph.D., J.D.; <i>Lecturer</i> .
DENISE L. GILMAN, B.A., J.D.; <i>Clinical Prof., Dir. Immigr. Clinic</i> .	MELINDA E. TAYLOR, B.A., J.D.; <i>Sr. Lect., Dir. Envtl. Law Clinic</i> .
BARBARA HINES, B.A., J.D.; <i>Clinical Prof., Dir. Immigr. Clinic</i> .	HEATHER K. WAY, B.A., B.J., J.D.; <i>Lecturer, Director of Community Development Clinic</i> .
KRISTINE A. HUSKEY, B.A., J.D.; <i>Dir. Nil. Sec. & Hum. Rts. Clinic</i> .	ELIZABETH M. YOUNGDALE, B.A., M.L.I.S., J.D.; <i>Lecturer</i> .
HARRISON KELLER, B.A., M.A., Ph.D.; <i>Vice Provost for Higher Education Policy, Senior Lecturer</i> .	

ADJUNCT PROFESSORS AND OTHER LECTURERS

ROBERT J. ADAMS, JR., B.S., M.B.A., Ph.D.
 WILLIAM R. ALLENSWORTH, B.A., J.D.
 WILLIAM G. BARBER, B.S.Ch.E., J.D.
 SHARON C. BAXTER, B.S., J.D.
 WILLIAM H. BEARDALL, JR., B.A., J.D.
 JERRY A. BELL, B.A., J.D.
 ALLISON H. BENESCH, B.A., M.S.W., J.D.
 CRAIG R. BENNETT, B.S., J.D.
 JAMES B. BENNETT, B.B.A., J.D.
 MELISSA J. BERNSTEIN, B.A., M.L.S., J.D.
 ROBERT S. BICKERSTAFF, B.A., J.D.
 MURFF F. BLEDSOE, B.A., J.D.
 WILLIAM P. BOWERS, B.B.A., J.D., LL.M.
 STACY L. BRAININ, B.A., J.D.
 ANTHONY W. BROWN, B.A., J.D.
 JAMES E. BROWN, LL.B.
 TOMMY L. BROYLES, B.A., J.D.
 W. A. BURTON, JR., B.A., M.A., LL.B.
 JOHN C. BUTLER, B.B.A., Ph.D.
 BENEDICTE CALLAN, B.A., Ph.D.
 AGNES E. CASAS, B.A., J.D.
 RUBEN V. CASTANEDA, B.A., J.D.
 EDWARD A. CAVAZOS, B.A., J.D.
 DAN J. CHRISTENSEN, B.B.A., J.D.
 JEFF CIVINS, A.B., M.S., J.D.
 JUDGE LEIF M. CLARK, B.A., M.Div., J.D.
 GARY COBB, B.A., J.D.
 ELIZABETH COHEN, B.A., M.S.W., J.D.
 LYNN R. COLEMAN, B.A., LL.B.
 JAMES W. COLLINS, B.S., J.D.
 PATRICIA J. CUMMINGS, B.A., J.D.
 DICK DEGUERIN, B.A., LL.B.
 MICHELE Y. DEITCH, B.A., M.S., J.D.
 ADAM R. DELL, B.A., J.D.
 RICHARD D. DEUTSCH, B.A., J.D.
 CASEY D. DUNCAN, B.A., M.L.I.S., J.D.
 PHILIP DURST, B.A., M.A., J.D., Ph.D.
 JAY D. ELLWANGER, B.A., J.D.
 KAITLIN T. FARRELL, B.A., J.D.
 ARLEY D. FINLEY, B.A., M.B.A., J.D.
 JOHN C. FLEMING, B.A., J.D.
 MARIA FRANKLIN, B.A., M.A., Ph.D.
 DAVID C. FREDERICK, B.A., Ph.D., J.D.
 GREGORY D. FREED, B.A., J.D.
 FRED J. FUCHS, B.A., J.D.
 MICHAEL GAGARIN, Ph.D.
 GERALD J. GALOW, B.A., J.D.
 JAMES B. GAMBRELL, B.S.M.E., M.A., LL.B.
 FRANCIS J. GAVIN, B.A., M.S.T., M.A., Ph.D.
 CHARLES E. GHOLZ, B.S., Ph.D.
 MICHAEL J. GOLDEN, A.B., J.D.
 JULIE E. GRANTHAM, B.A., J.D.
 SHERRI R. GREENBERG, B.A., M.Sc.
 ROBERT L. GROVE, B.S., J.D.
 DAVID HALPERN, B.A., J.D.
 ELIZABETH HALUSKA-RAUSCH, B.A., M.A., M.S., Ph.D.
 JETT L. HANNA, B.B.A., J.D.
 KELLY L. HARAGAN, B.A., J.D.
 CLINT A. HARBOUR, B.A., J.D., LL.M.
 ROBERT L. HARGETT, B.B.A., J.D.
 JAMES C. HARRINGTON, B.A., M.A., J.D.
 CHRISTOPHER S. HARRISON, Ph.D., J.D.
 JOHN R. HAYS, JR., B.A., J.D.
 P. MICHAEL HEBERT, A.B., J.D.
 STEVEN L. HIGHLANDER, B.A., Ph.D., J.D.
 SUSAN J. HIGHTOWER, B.A., M.A., J.D.
 WILLIAM M. HINES, III, B.B.A., J.D.
 JAMES C. HO, B.A., J.D.
 KENNETH E. HOUP, JR., J.D.
 RANDY R. HOWRY, B.J., J.D.
 MONTY G. HUMBLE, B.A., J.D.
 DIRK M. JORDAN, B.A., J.D.
 SEAN D. JORDAN, B.A., J.D.
 PATRICK O. KEEL, B.A., J.D.
 DOUGLAS L. KEENE, Ph.D.
 CHARI L. KELLY, B.A., J.D.
 ROBERT N. KEPPLER, B.A., J.D.
 MARK L. KINCAID, B.B.A., J.D.
 AMI L. LARSON, B.A., J.D.
 KEVIN R. LASHUS, B.A., J.D.
 JODI R. LAZAR, B.A., J.D.
 KEVIN L. LEAHY, B.A., J.D.
 MAURIE A. LEVIN, B.A., J.D.
 JAMES LLOYD-LOFTIS, B.B.A., J.D.
 ANDREW F. MACRAE, B.J., J.D.
 JIM MARCUS, B.A., J.D.
 PETER D. MARKETOS, B.A., J.D.
 HARRY S. MARTIN, A.B., M.L.S., J.D.
 FRANCES L. MARTINEZ, B.A., J.D.
 LAURA A. MARTINEZ, B.A., J.D.
 RAY MARTINEZ, III, B.A., J.D.
 LISA M. MCCLAINE, B.A., J.D., LL.M.
 STEPHANIE G. MCFARLAND, B.A., J.D.
 BARRY F. MCNEIL, B.A., J.D.
 ANGELA T. MELINARAAB, B.F.A., J.D.
 JO A. MERICA, B.A., J.D.
 RANELLE M. MERONEY, B.A., J.D.
 ELIZABETH N. MILLER, B.A., J.D.
 DARYL L. MOORE, B.A., M.L.A., J.D.
 EDWIN G. MORRIS, B.S., J.D.
 MANUEL H. NEWBURGER, B.A., J.D.
 DAVID G. NIX, B.S.E., LL.M., J.D.
 PATRICK L. O'DANIEL, B.B.A., J.D.
 M. A. PAYAN, B.A., J.D.
 MARK L. PERLMUTTER, B.S., J.D.
 LOUIS T. PIRKEY, B.S.Ch.E., J.D.
 JUDGE ROBERT L. PITMAN, B.S., J.D.
 ELIZA T. PLATTS-MILLS, B.A., J.D.
 LAURA L. PRATHER, B.B.A., J.D.
 JONATHAN PRATTER, B.A., M.L.S., J.D.
 VELVA L. PRICE, B.A., J.D.
 BRIAN C. RIDER, B.A., J.D.
 GRETCHEN RITTER, B.S., Ph.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.
 AMANDA M. RUNYON, B.B.A., J.D.
 SARA J. RUSHING, B.A., LL.M., J.D.
 MATTHEW C. RYAN, B.A., J.D.
 KAREN R. SAGE, B.A., J.D.
 MARK A. SANTOS, B.A., J.D.
 CHRISTOPHE H. SAPSTEAD, B.A., J.D.
 SUSAN SCHULTZ, B.S., J.D.
 AMY J. SCHUMACHER, B.A., J.D.
 MARCUS F. SCHWARTZ, B.B.A., J.D.
 SUZANNE SCHWARTZ, B.J., J.D.
 RICHARD J. SEGURA, JR., B.A., J.D.
 DAVID A. SHEPPARD, B.A., J.D.
 JUDGE ERIC M. SHEPPERD, B.A., J.D.
 ALISSA R. SHERRY, B.A., M.A., Ph.D.
 NADAV SHOKED, LL.B., S.J.D.
 RONALD J. SIEVERT, B.A., J.D.
 AMBROSIO A. SILVA, B.S., J.D.
 BEA A. SMITH, B.A., M.A., J.D.
 LYDIA N. SOLIZ, B.B.A., J.D.
 KATHERINE J. SOLON, B.A., M.S., J.D.
 JUSTICE ROSE B. SPECTOR, B.A., J.D.
 DAVID B. SPENCE, B.A., J.D., M.A., Ph.D.
 LAURA L. STEIN, B.A., M.A., Ph.D.
 JAMES B. STEINBERG, B.A., J.D.
 CHARLES H. STILL, B.B.A., J.D.

WILLIAM F. STUTTS, B.A., J.D.
MATTHEW J. SULLIVAN, B.S., J.D.
BRADLEY P. TEMPLE, B.A., J.D.
SHERINE E. THOMAS, B.A., J.D.
TERRY O. TOTTENHAM, B.S., LL.M., J.D.
JEFFREY K. TULIS, B.A., M.A., Ph.D.
ROBERT W. TURNER, B.A., LL.B.
TIMOTHY J. TYLER, B.A., J.D.
SUSAN S. VANCE, B.B.A., J.D.
ERIC M. VANHORN, B.A., J.D.
LANA K. VARNEY, B.J., J.D.
SRIRAM VISHWANATH, B.S., M.S., Ph.D.

DEBORAH M. WAGNER, B.A., M.A., J.D.
CLARK C. WATTS, B.A., M.D., M.A., M.S., J.D.
JANE M. WEBRE, B.A., J.D.
WARE V. WENDELL, A.B., J.D.
DARA J. WHITEHEAD, B.A., M.S.
RANDALL B. WILHITE, B.B.A., J.D.
DAVID G. WILLE, B.S.E.E., M.S.E.E., J.D.
MARK B. WILSON, B.A., M.A., J.D.
JUDGE PAUL L. WOMACK, B.S., J.D.
LUCILLE D. WOOD, B.A., J.D.
LARRY F. YORK, B.B.A., LL.B.
DANIEL J. YOUNG, B.A., J.D.

VISITING PROFESSORS

OWEN L. ANDERSON, B.A., J.D.
ANTONIO H. BENJAMIN, LL.B., LL.M.
AARON-ANDREW P. BRUHL, B.A., M.Phil., J.D.
PETER F. CANE, B.A., LL.B., D.C.L.
VICTOR FERRERES, J.D., LL.M., J.S.D.

FRANCESCO FRANCONI, J.D., LL.M.
ROBERT M. O'NEIL, A.B., A.M., LL.B., LL.D., LL.D.
GUILLERMO A. PADILLA, J.D., Ph.D.
ADAM H. ROSENZWEIG, B.A., LL.M., J.D.
KEVIN G. TOH, A.B., M.A., Ph.D., J.D.

* * *

Texas Law Review

Volume 89

Number 5

April 2011

OMAR OCHOA
Editor in Chief

ANTHONY F. ARGUIJO
Managing Editor

DAVID A. SNYDER
Chief Articles Editor

ANDREA L. FAIR
Administrative Editor

SARAH A. HUNGER
Chief Notes Editor

GRAYSON E. MCDANIEL
Book Review Editor

TRACEY A. BAMBERGER
Online Content Editor

KATHERINE N. DOORLEY
Research Editor

SARAH E. COBLE
CHRISTOPHER G. HORNIG
MICAH R. KEGLEY
ANDREW B. LANGWORTHY
Articles Editors

SERINE R. CONSOLINO
JAMIE E. FRANCE
CHRISTOPHER G. GRANAGHAN
Notes Editors

JAMES R. LLOYD
ZACHARY T. PADGETT
MICHAEL P. PARMERLEE
CATHERINE E. WAGNER
Articles Editors

CEDRIC L. ALLEN
SHANE C. ANDERSON
LAUREN E. BARROWS
LAUREN D. CORBEIL
MICHAEL A. CUMMING

JONAH D. JACKSON
JEFFREY LIANG
ELIZABETH C. ROWLAND
JOHN F. SUMMERS

AMANDA M. SUZUKI
CHRISTINE M. TAMER
KRIS Y. TENG
MARK F. WILES
J. T. WILLIAMS

Associate Editors

Members

F. GIBBONS ADDISON
DAVID E. ARMENDARIZ
BETH N. BEAURY
ERIC S. BERELOVICH
JUSTIN L. BERNSTEIN
CHRISTOPHER M. BEZEG
ROBERT W. BROWN
ROBERT T. CARMAN
JONATHAN L. CHALTAİN
DANIEL C. CLEMONS
NEAL A. COLEMAN
TANIA M. CULBERTSON
LEONARD J. DENBINA
DARIUS A. DILLS
ADRIAN L. DUNBAR
LESLIE S. DYKE
MAXIM FARBEROV
SHEILA FORJUOH
CHARLES E. FOWLER
STEPHEN A. FRASER
SCOTT J. FULFORD
NEIL K. GEHLAWAT
RYAN M. GOLDSTEIN

MATTHEW A. GREENBERG
ALDEN G. HARRIS
ANDREW T. INGRAM
MICHELLE J. JACOBSON
PAIGE M. JENNINGS
DON J. KAHN
DAVID J. KOHTZ
W. LAWSON KONVALINKA
DANIEL LENHOFF
ERIC M. LEVENTHAL
KRISTIN M. MALONE
GABRIEL H. MARKOFF
BLAIR K. MCCARTNEY
MYRIAH J. MELTON
LINDSEY A. MILLS
ADAM D. MOSES
ADAM J. MOSS
KATHLEEN L. NANNEY
BRIAN C. OSIMIRI
JOSHUA H. PACKMAN
GEORGE M. PADIS
DREW D. PENNEBAKER

ATHENA J. PONCE
JEFFREY T. QUILICI
MICHAEL T. RAUPP
JEREMY M. REICHMAN
AMELIA C. RENDEIRO
RAFE A. SCHAEFER
KATHRYN E. SEMMLER
SIDNEY K. SMITH
NICHOLAS D. STEPP
KASYN M. STEVENSON
CHRIS S. STEWART
ARIELLE B. SWARTZ
KEITH P. SYSKA
REID A. TEPPER
KARSON K. THOMPSON
GEORGE D. VALLAS
ERIC T. WERLINGER
SARAH K. WEYNAND
JEFFREY M. WHITE
WRIGHT T. WOMMACK
KRISTEN A. WONG
CATHARINE S. WRIGHT
JOYCE Y. YOUNG

PAUL N. GOLDMAN
Business Manager

MITCHELL N. BERMAN
JOHN S. DZIENKOWSKI
Faculty Advisors

TERI GAUS
Editorial Assistant

* * *

Texas Law Review

Volume 89, Number 5, April 2011

ARTICLES

Patentable Subject Matter and Institutional Choice
John M. Golden 1041

Share Repurchases, Equity Issuances, and the Optimal Design
of Executive Pay
Jesse M. Fried 1113

BOOK REVIEW

But I Thought the Earth Belonged to the Living
Mark L. Ascher 1149

reviewing Lawrence M. Friedman's
DEAD HANDS: A SOCIAL HISTORY OF WILLS,
TRUSTS, AND INHERITANCE LAW

NOTES

A Proposed Wealth Redistribution System Based on the
Underlying Premise of Revenue Sharing in American Pro
Sports
F. Gibbons Addison 1179

Third-Party Assistance in Determining Obviousness
Gregory R. Baden 1203

The Effect of the Parliamentary Voting System Act on the
British Constitution
Joshua H. Packman 1229

* * *

Articles

Patentable Subject Matter and Institutional Choice

John M. Golden*

Given the existence of fine-grained requirements for patentability such as nonobviousness, the utility of a separate requirement of patentable subject matter has sometimes been questioned. The courts' fumbling efforts to regulate patentable subject matter have helped stain the enterprise with suspicion and even disrepute. This Article first defends limitations on subject-matter eligibility by showing that they provide a categorical filter that can improve patent-system performance. The Article then argues that the enterprise of regulating patentable subject matter should be primarily entrusted to the USPTO, rather than, as it is now, to the courts. Two mathematical models illustrate (1) how more individualized tests for patentability can fail to ensure that patents improve social welfare and (2) how a particular form of subject matter—fundamental principles having a very high number of potential uses—can generate particularly high social costs and thus qualify as a form of subject matter that the patent system would best filter out. With respect to the proper locus for rulemaking authority, the USPTO's capacity and incentive to respond promptly and meaningfully to questions of subject-matter eligibility make it the best candidate. Moreover, giving the USPTO rulemaking authority with respect to subject-matter eligibility does not require giving it rulemaking authority on all matters of patent-law substance. Just as other regimes of U.S. law have divided tasks of adjudication and enforcement between different institutions, the patent system can divide areas of primary interpretive authority between the USPTO and Article III courts. Such an institutional innovation appears the best way to leverage the relative institutional competences of the USPTO, the courts, and Congress.

* Professor, The University of Texas School of Law. For helpful comments and conversations, I thank David Adelman, Ronen Avraham, Bernie Black, Oren Bracha, Tom Cotter, Mark Gergen, Todd Rakoff, Dan Rodriguez, Joshua Sarnoff, Matt Spitzer, participants in The University of Texas School of Law's Drawing Board workshop and the University of Akron School of Law's Third Annual IP Professors Forum, and the editors of the *Texas Law Review*.

I.	Introduction.....	1042
II.	Patent Law’s Current Regime of Hybrid Interpretive Authority	1044
	A. Power Sharing Across a Procedure–Substance Divide.....	1045
	B. Efforts to Expand USPTO Authority and Their Critique.....	1051
III.	The Case for Binding USPTO Authority over Subject-Matter Eligibility	1054
	A. Nature and Justification of Patent Law’s Subject-Matter Inquiry	1055
	1. <i>The Patent Bargain and Its Finer Grained Requirements</i>	1055
	2. <i>The Coarse-Grained Filter for Patentable Subject Matter</i>	1058
	3. <i>The Value of a Coarse-Grained Filter</i>	1064
	B. The Choice of Institutional Actor	1074
	1. <i>Courts’ Struggles with Defining Patentable Subject Matter</i>	1075
	2. <i>Undesirability of Reliance on Congress</i>	1091
	3. <i>The USPTO’s Comparative Institutional Competence</i>	1095
IV.	Conclusion	1111

I. Introduction

In recent years, a wide range of substantive and institutional aspects of U.S. patent law have received intense scrutiny.¹ Aspects subjected to legislative, administrative, or judicial review have included questions of whether patentable subject matter properly encompasses business methods² and human genetic sequences,³ as well as questions of how much rulemaking authority should be possessed by U.S. patent law’s primary administrative agency, the United States Patent & Trademark Office (USPTO).⁴ This

1. See generally JAMES BESSEN & MICHAEL J. MEURER, *PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK* (2008) (criticizing U.S. patent law’s performance as a means of increasing net incentives for innovation); DAN L. BURK & MARK A. LEMLEY, *THE PATENT CRISIS AND HOW THE COURTS CAN SOLVE IT* (2009) (contending that U.S. patent law needs to be better tailored to fit distinct technologies and their associated industries).

2. See, e.g., *Bilski v. Kappos*, 130 S. Ct. 3218, 3223 (2010) (examining “whether a patent can be issued for a claimed invention designed for the business world”).

3. See, e.g., *Ass’n for Molecular Pathology v. U.S. Patent & Trademark Office*, 702 F. Supp. 2d 181, 185 (S.D.N.Y. 2010) (responding to the question, “Are isolated human genes and the comparison of their sequences patentable?”).

4. See, e.g., Craig Allen Nard, *Legal Forms and the Common Law of Patents*, 90 B.U. L. REV. 51, 56–57 (2010) (contending that “judicial primacy acts as a bulwark against the more politicized legislative process or capture-prone administrative rulemaking”); Arti K. Rai, *Growing Pains in the Administrative State: The Patent Office’s Troubled Quest for Managerial Control*, 157 U. PA. L. REV. 2051, 2080 (2009) (arguing that Congress should “giv[e] the [USPTO] significantly greater authority over fee setting”). The USPTO has been officially renamed twice in the last few decades. Jeffrey A. Lefstin, *The Constitution of Patent Law: The Court of Customs and Patent Appeals and the Shape of the Federal Circuit’s Jurisprudence*, 43 LOY. L.A. L. REV. 843, 847 n.17 (2010) (“The Patent Office was renamed the Patent and Trademark Office in 1975 . . . and renamed the U.S. Patent and Trademark Office in 1999 . . .” (internal citations omitted)). For purpose of simplicity,

Article develops a specific connection between these lines of inquiry by arguing that Congress should expand the USPTO's rulemaking authority so that it encompasses substantive questions of subject-matter eligibility.

By giving the USPTO this prescribed dose of substantive rulemaking authority, the proposed reform would sharpen the already chimeric character of the U.S. patent system, a legal hybrid that features (1) the U.S. Court of Appeals for the Federal Circuit, a semispecialized circuit court of appeals that, absent Supreme Court reversal, provides authoritative national interpretations of *substantive* patent law;⁵ and (2) the USPTO, an administrative agency that currently has rulemaking authority with respect to questions of office *procedure*.⁶ By explicitly splitting interpretive authority over substantive patent law between the USPTO and the federal judiciary, the proposal would further hybridize patent law's legal regime and break from a common paradigm under which primary interpretive authority for the substance of a statutory regime lies either wholly with the courts or wholly with an administrative agency.⁷ In this sense, this Article builds on an earlier generation of administrative law literature discussing split-authority models in which authority for an administrative regime is shared by multiple administrative agencies.

Part II describes the already peculiarly hybrid nature of U.S. patent law's legal regime. Part III then presents the case for giving the USPTO binding interpretive authority over subject-matter eligibility as a means toward patent law's commonly accepted utilitarian ends. In particular, Part III first describes the nature of subject-matter inquiries and develops mathematical models that help justify their continued use as patentability filters. Part III then argues that the generally categorical nature of questions of subject-matter eligibility—their tendency toward resolution on grounds that are comparatively generic, rather than substantially case specific—suggests that issues of subject-matter eligibility are especially appropriate for agency rulemaking. This conclusion is bolstered by an analysis of comparative

this Article refers to the USPTO or its predecessors uniformly, albeit sometimes anachronistically, as the USPTO.

5. See John M. Golden, *The Federal Circuit and the D.C. Circuit: Comparative Trials of Two Semi-Specialized Courts*, 78 GEO. WASH. L. REV. 553, 554 (2010) (“The D.C. and Federal Circuits both provide examples of relatively new experiments in semi-specialization.”).

6. 35 U.S.C. § 2(b)(2)(A) (2006) (giving the USPTO authority to make rules “govern[ing] the conduct of proceedings in the Office”).

7. See *United States v. Mead Corp.*, 533 U.S. 218, 229 (2001) (stating that, when it is “apparent from the agency’s generally conferred authority and other statutory circumstances that Congress would expect the agency to be able to speak with the force of law when it addresses ambiguity . . . , a reviewing court . . . is obliged to accept the agency’s position if Congress has not previously spoken to the point . . . and the agency’s interpretation is reasonable”); cf. Michael J. Burstein, *Rules for Patents*, 52 WM. & MARY L. REV. (forthcoming 2011) (manuscript at 1), available at <http://ssrn.com/abstract=1718695> (arguing for a general grant of substantive rulemaking authority to the USPTO); Jonathan S. Masur, *Regulating Patents* 5 (Univ. of Chi. Pub. Law & Legal Theory, Working Paper No. 334, 2011), available at <http://ssrn.com/abstract=1709222> (same).

institutional competence. The federal judiciary's historic struggles with subject-matter issues suggest a lack of judicial facility for resolving problems in this area. Congress's long silence on most significant questions of patentable subject matter comports with a sense that the national legislature lacks the sustained interest, time, and knowledge to competently clarify subject-matter eligibility's bounds. Indeed, among potential institutional candidates for this task, the USPTO appears the only one likely to have the expertise and the incentive—generated by the constant crush of new applications that it must process—to deal with subject-matter eligibility questions promptly, effectively, and with adequate protection of developed expectations.

Of course, with any administrative agency, there are the usual concerns of capture and bias. Part III argues that, with respect to the USPTO, these concerns (1) are not as great as is often contended and (2) can be further diluted through additional institutional reforms, such as provision for administrative challenges to subject-matter eligibility in U.S.-style reexaminations or European-style oppositions. As Part III points out, the USPTO already has an established record of developing nonbinding but influential interpretive rules on matters of substance. With a grant of primary interpretive authority over issues of subject-matter eligibility, the USPTO might be able to repeat these successes by clearing doctrinal tangles generated by the courts. The 2010 decision by the U.S. Supreme Court in *Bilski v. Kappos*,⁸ which essentially threw back to the U.S. Court of Appeals for the Federal Circuit the problem of developing “limiting criteria” for patentable subject matter,⁹ provides no indication to the contrary.

II. Patent Law's Current Regime of Hybrid Interpretive Authority

U.S. patent law has a distinct institutional structure. It features both (1) a non-comprehensive statute that leaves much room for gap-filling and interpretation¹⁰ and (2) an appellate court of national jurisdiction, the U.S. Court of Appeals for the Federal Circuit, that has a virtually exclusive hold on patent appeals taken as of right.¹¹ Also, U.S. patent law features (3) an administrative agency, the USPTO, that lacks binding interpretive authority on matters of substantive patent law but that possesses binding rulemaking authority with respect to procedural aspects of USPTO activities, such as patent examination, issue, and reexamination.¹² The result is an odd legal

8. 130 S. Ct. 3218 (2010).

9. *Id.* at 3231.

10. See Nard, *supra* note 4, at 53 (“[T]he common law has been the dominant legal force in the development of U.S. patent law for over two hundred years.”).

11. See John M. Golden, *The Supreme Court as “Prime Percolator”: A Prescription for Appellate Review of Questions in Patent Law*, 56 UCLA L. REV. 657, 664 (2009) (discussing the Federal Circuit's appellate jurisdiction).

12. Nard, *supra* note 4, at 88 (describing the USPTO as “provided with substantive guidance as to when a patent application should (or should not) issue, but left with plenary authority to establish

chimera—an administrative law regime in which Congress has effectively split interpretive responsibility¹³ between a semispecialized Article III court and an executive agency having nearly 10,000 employees¹⁴ and an annual budget of about \$2 billion.¹⁵

Of course, the U.S. Supreme Court and Congress can also intervene to fill legal gaps or to override Federal Circuit or USPTO interpretations. But these institutions' involvement with patent law is purely discretionary, and they frequently choose to take a back seat. Caught between opposing lobbies,¹⁶ Congress has repeatedly failed to pass patent reform legislation in the past decade. Further, even when the Supreme Court is viewed as substantially involved with patent law, such involvement still tends to entail no more than a couple decisions on the merits each year.¹⁷ Thus, as a practical matter, the Supreme Court's role in developing substantive patent law is unlikely to dwarf that of the Federal Circuit, which annually decides hundreds of patent cases on the merits.¹⁸

In any event, Supreme Court review of Federal Circuit decisions at most results in substitution of one judicial panel's formulation of legal doctrine for another's. Regardless of how the Supreme Court and the Federal Circuit share their judicial responsibilities, U.S. patent law remains distinctive as a peculiarly hybrid regime in which interpretive authority is divided between the federal government's Article III judiciary and an executive agency.

A. Power Sharing Across a Procedure–Substance Divide

A key aspect of patent law's distinctiveness is the USPTO's lack of substantive rulemaking power.¹⁹ As a result of this limitation, the USPTO

regulations that 'govern the conduct of proceedings in the Office'" (quoting 35 U.S.C. § 2(b)(2)(A) (2006)).

13. See generally Joseph Scott Miller, *Substance, Procedure, and the Divided Patent Power*, 63 ADMIN. L. REV. (forthcoming 2011), available at <http://ssrn.com/abstract=1578542> (describing Congress as having "spli[t] the patent power's substantive and procedural parts").

14. U.S. PATENT & TRADEMARK OFFICE, PERFORMANCE AND ACCOUNTABILITY REPORT: FISCAL YEAR 2009, at 11 [hereinafter USPTO REPORT] ("At the end of FY 2009, the USPTO work force was composed of 9,716 federal employees (including 6,243 patent examiners, and 388 trademark examining attorneys).").

15. See *id.* at 48 (reporting expenditures of \$1.98 billion in fiscal year 2009).

16. See John M. Golden, *Principles for Patent Remedies*, 88 TEXAS L. REV. 505, 507 (2010) (describing lobbying battles between "Orwellian-named entities grouped along industry lines").

17. In the ten years from 2001 through 2010, the Supreme Court issued a total of ten patent decisions on the merits. See *Bilski v. Kappos*, 130 S. Ct. 3218 (2010) (addressing the scope of patentable subject matter); Golden, *supra* note 11, at 670 fig. (counting nine Supreme Court decisions on the merits in patent cases from October Term 2001 through October Term 2007).

18. Golden, *supra* note 11, at 667 n.51.

19. See *Merck & Co. v. Kessler*, 80 F.3d 1543, 1550 (Fed. Cir. 1996) ("Congress has not vested the Commissioner with any general substantive rulemaking power . . ."); see also *Cooper Techs. Co. v. Dudas*, 536 F.3d 1330, 1336 (Fed. Cir. 2008) ("We have . . . previously held that 35 U.S.C. § 2(b)(2) does not authorize the Patent Office to issue 'substantive' rules."); Nard, *supra* note 4, at 76 ("Contrast the Federal Circuit's extraordinary powers with the relatively minor role played by

lacks authority to issue presumptively binding rules on the substantive legal questions such as patentability, even though the USPTO must routinely rule on such issues in deciding whether to grant patents. The USPTO can provide guidance, both for its employees and for others, regarding how the USPTO plans to interpret and to apply existing case law and statutory provisions. The USPTO cannot, however, independently develop interpretations or gap-filling rules or legal doctrines that legally bind non-USPTO parties on matters of patent-law substance. USPTO guidance on substantive issues—including, for example, the definition of what actions or omissions by a patent applicant constitute inequitable conduct that can render a patent unenforceable—has no more than persuasive authority in the courts.²⁰

On the other hand, the USPTO is far from a merely ministerial agency. Patent examiners and the USPTO's internal Board of Patent Appeals and Interferences (BPAI) regularly rule on difficult substantive questions, such as subject-matter eligibility and nonobviousness, whose governing principles and proper manner of explication have commonly confounded Article III courts. The BPAI periodically issues "Precedential" opinions that are binding precedent for the BPAI itself²¹ and that look very much like ordinary judicial opinions.²² More frequently, the BPAI issues shorter opinions that do not constitute binding precedent for the BPAI but might provide useful instruction to third parties, patent examiners, and members of future BPAI panels.²³

Further, Congress has explicitly granted the USPTO the power to make rules "govern[ing] the conduct of proceedings in the Office."²⁴ The Federal Circuit has ruled that, with respect to rules promulgated under that authority, the USPTO receives *Chevron*²⁵ deference, a high level of formal deference requiring courts to accept reasonable agency interpretations of statutory

the [USPTO] in the context of substantive engagement."). The USPTO's lack of substantive rulemaking authority is distinctive but not unique. See M. Elizabeth Magill, *Agency Choice of Policymaking Form*, 71 U. CHI. L. REV. 1383, 1387 (2004) ("The Equal Employment Opportunity Commission . . . does not have the authority to issue legislative rules that further define the prohibitions on discrimination contained in Title VII of the Civil Rights Act.").

20. See Golden, *supra* note 11, at 665 ("According to Federal Circuit precedent, the [USPTO] lacks substantive rulemaking power and thus receives only weak deference for its interpretations of the Patent Act.").

21. See U.S. PATENT & TRADEMARK OFFICE, BOARD OF PATENT APPEALS & INTERFERENCES (BPAI), STANDARD OPERATING PROCEDURE 2 (REVISION 7): PUBLICATION OF OPINIONS AND BINDING PRECEDENT 5-6 (2008) [hereinafter BPAI Operating Procedure] (specifying that BPAI opinions designated as "Precedential" are "binding upon the Board").

22. See, e.g., *Ex parte Lundgren*, 76 U.S.P.Q.2d 1385 (2005).

23. BPAI Operating Procedure, *supra* note 21, at 6 ("Informative opinions are not binding, but illustrate norms of Board decision-making for the public, the patent examining corps, and future Board panels.").

24. 35 U.S.C. § 2(b)(2)(A) (2006).

25. *Chevron U.S.A., Inc. v. Natural Res. Def. Council, Inc.*, 467 U.S. 837 (1984).

law.²⁶ Recent battles over ultimately rescinded USPTO rules regulating patent applications have highlighted the nontriviality of this authority: the line between procedural rules and substantive rules is not always clear, and even a relatively narrow definition of the category of “procedural rules” likely leaves substantial room for rules that significantly affect practical legal outcomes.²⁷

Moreover, the USPTO plays a nontrivial part in patent law’s development and practical effect even when, as on matters of patent-law substance, it dutifully acts in the roles of subordinate and hopeful adviser to the courts. The USPTO’s massive *Manual of Patent Examining Procedure* (MPEP) provides page after page of painstaking instruction on how substantive patent law doctrines should be applied in the context of patent examination.²⁸ No matter how incoherent or tortured relevant judicial precedent is, the USPTO must try to distill it into a set of comprehensible guidelines for several thousand patent examiners,²⁹ each of whom must ultimately rule on the patentability of claims in a sample of the hundreds of thousands of applications that the USPTO receives annually.³⁰ The result of the USPTO’s efforts at distillation can sometimes be comic: certain complicated, fact-specific analyses simply do not seem readily reducible to flowchart form.³¹ But given the common reliance of examiners and patent applicants on MPEP guidance,³² as well as that guidance’s reasonably

26. See *Cooper Techs. Co. v. Dudas*, 536 F.3d 1330, 1337 (Fed. Cir. 2008) (“Because the Patent Office is specifically charged with administering statutory provisions relating to ‘the conduct of proceedings in the Office,’ we give *Chevron* deference to its interpretations of those provisions.” (citation omitted)).

27. See *Tafas v. Doll*, 559 F.3d 1345, 1355 (Fed. Cir. 2009) (quoting the D.C. Circuit’s observation that “procedure impacts on outcomes and thus can virtually always be described as affecting substance” (internal quotation marks omitted)), *vacated and petition for reh’g en banc granted*, 328 F. App’x 658, *dismissed as moot sub nom.* *Tafas v. Kappos*, 586 F.3d 1369 (Fed. Cir. 2009).

28. See, e.g., U.S. PATENT & TRADEMARK OFFICE, MANUAL OF PATENT EXAMINING PROCEDURE § 2106, at 2100-16 (rev. 8th ed. 2010) [hereinafter MPEP], available at <http://www.uspto.gov/web/offices/pac/mpep/index.htm>.

29. USPTO REPORT, *supra* note 14, at 11.

30. *Id.* at 112 tbl.1 (reporting that, in fiscal year 2009, the USPTO received 457,966 standard utility-patent applications and issued 165,212 standard utility patents). See generally Kelly Casey Mulally, *Legal (Un)certainty, Legal Process, and Patent Law*, 43 LOY. L.A. L. REV. 1109, 1126-27 (2010) (“Although patent examiners apply the same legal rules that courts apply, they are not lawyers The MPEP accordingly attempts to distill the law in an algorithmic fashion . . . so that examiners may apply the law to a given patent application using a step-by-step rubric.”).

31. See, e.g., MPEP, *supra* note 28, ch. 2100, at 2100-1 to 2100-247 (giving hundreds of pages of instruction on how to examine issues relating to patentability).

32. See, e.g., *Molins PLC v. Textron, Inc.*, 48 F.3d 1172, 1180 n.10 (Fed. Cir. 1995) (“The MPEP [is] commonly relied upon as a guide to patent attorneys and patent examiners on procedural matters.” (internal quotation marks omitted)); David B. Gornish, *A Patent Law Primer for Health Law Practitioners*, HEALTH LAW., Oct. 2008, at 1, 3 (“Although it does not carry the force of law, the MPEP sets forth the USPTO’s interpretation of how the law is to be applied by the agency, and is therefore regularly consulted by patent attorneys.”); Theresa Stadheim, *How KSR v. Teleflex Will*

frequent citation in court opinions,³³ the significance of USPTO guidance activity as a source of influence and practical meaning should not be understated.

On the other hand, in providing guidance, the USPTO is significantly more handicapped, practically speaking, than the enforcement agencies for a statutory regime with which patent law is often compared: U.S. antitrust law. Both patent law and antitrust law are frequently, albeit somewhat oxymoronic, characterized as common-law statutory regimes. This characterization is meant to indicate that these legal regimes are governed, indeed created, by statutory provisions but are nevertheless common-law-like because their statutory underpinnings ultimately provide no more than a basic legal framework that leaves the task of filling large statutory lacunae primarily to the courts.³⁴ Accordingly, just as the USPTO lacks rulemaking authority on substantive matters of patent law, the two administrative agencies with primary enforcement responsibilities for U.S. antitrust law, the Department of Justice (DOJ) and the Federal Trade Commission (FTC), appear to lack general rulemaking authority on substantive matters of antitrust law.³⁵

On the other hand, the DOJ and FTC have a significant power that the USPTO lacks: discretion to set policy by deciding when to act. The USPTO is subject to an extreme form of action forcing. By statute, the USPTO must examine any application that is filed, and must grant a patent for an application that “appears” to meet legal requirements.³⁶ Further, these

Affect Patent Prosecution in the Electrical and Mechanical Arts, 91 J. PAT. & TRADEMARK OFF. SOC'Y 142, 148 (2009) (“The MPEP is the main reference for patent examiners, and is used by patent prosecution attorneys when deciding how to traverse various examiner rejections.”).

33. *E.g.*, *Eli Lilly & Co. v. Teva Pharm. USA*, 619 F.3d 1329, 1343 (Fed. Cir. 2010) (“As the district court acknowledged, the [MPEP] explains that the initiation of a clinical trial has a significant impact on the [USPTO’s] utility inquiry”); *In re ’318 Patent Infringement Litig.*, 583 F.3d 1317, 1326 (Fed. Cir. 2009) (“Although no case has been called to our attention where utility was established simply by analytic reasoning, the PTO’s [MPEP] has recognized that ‘arguments or reasoning’ may be used to establish an invention’s therapeutic utility.” (footnote omitted)).

34. *Nard*, *supra* note 4, at 53 (“[T]he patent code, much like [antitrust law’s] Sherman Act, is a common law enabling statute, leaving ample room for courts to fill in the interstices or to create doctrine emanating solely from Article III’s province.”).

35. *See* Daniel A. Crane, *Technocracy and Antitrust*, 86 TEXAS L. REV. 1159, 1199 (2008) (“Like the [DOJ’s] Antitrust Division, the FTC has little power to create antitrust norms but merely enforces the norms created by the generalist Article III courts that review FTC decisions.”); Hillary Greene, *Guideline Institutionalization: The Role of Merger Guidelines in Antitrust Discourse*, 48 WM. & MARY L. REV. 771, 841 (2006) (“The DOJ enjoys no express or implicit grant of rulemaking authority within the antitrust realm. Though in theory one could argue that the FTC enjoys rulemaking authority in the competition context, it is not clear that position would prevail, and as a practical matter it is equally unlikely the FTC would advance such a position.”).

36. 35 U.S.C. § 131 (2006) (“The Director shall cause an examination to be made of the application and the alleged new invention; and if on such examination it appears that the applicant is entitled to a patent under the law, the Director shall issue a patent therefor.”).

required substantive decisions are subject to judicial review. In the event of a rejection, an applicant for patent rights can appeal directly to the Federal Circuit³⁷ or can file a civil challenge in the U.S. District Court for the District of Columbia.³⁸ In the event of an allowance in an *inter partes* reexamination, a challenger to patent rights can appeal to the Federal Circuit.³⁹ Alternatively, a challenger who can establish standing may challenge the validity of issued patent claims in the district courts.⁴⁰

In contrast, the DOJ and FTC can make significant contributions to antitrust law's effective substance and doctrinal development by deciding when not to act and by also signaling in advance the circumstances that are likely to trigger enforcement action.⁴¹ Even though private parties can bypass the enforcement agencies' judgments not to act by bringing civil suits to enforce the antitrust laws,⁴² the antitrust agencies can provide significant comfort to potential targets of antitrust suits by delineating safe harbors in which those potential targets will not face the prospect of criminal liability.⁴³ And discrepancies between the antitrust laws' apparent reach and these agencies' enforcement practices and guidance can help inspire courts to rethink prior judicial understandings.⁴⁴

On the other hand, this distinction between the situations of the USPTO and U.S. antitrust-enforcement agencies should not be overplayed. Although the USPTO does not have a true analog of the enforcement agencies' discretionary enforcement authority, the open-textured nature of various aspects of U.S. patent law can, even with an overlay of judicial precedent, leave the USPTO with substantial opportunities to be an influential first mover, particularly with respect to extension of existing doctrine to new technologies.

Thus, for example, in the wake of the Supreme Court's 1980 determination in *Diamond v. Chakrabarty*⁴⁵ that a "human-made, genetically

37. *Id.* § 141.

38. *Id.* § 145 (providing for a "civil action against the Director").

39. *Id.* § 315.

40. *See id.* § 282 (providing that invalidity is a defense "in any action involving the validity or infringement of a patent").

41. *Cf.* HERBERT HOVENKAMP, *THE ANTITRUST ENTERPRISE: PRINCIPLE AND EXECUTION* 60 (2005) ("The Justice Department does not enforce the price discrimination provisions of the Robinson-Patman Act at all, and the FTC does so only rarely.").

42. 15 U.S.C. § 15.

43. *See* Jeffrey Manns, *Private Monitoring of Gatekeepers: The Case of Immigration Enforcement*, 2006 U. ILL. L. REV. 887, 969 n.354 ("[T]he Antitrust Division of the Department of Justice routinely issues enforcement guidelines that effectively create a safe harbor from enforcement.").

44. *See, e.g.*, *Ill. Tool Works Inc. v. Indep. Ink, Inc.*, 547 U.S. 28, 45 (2006) (observing, in support of overruling a prior judicial presumption that a patent confers market power, that in 1995 guidelines, the DOJ and FTC "stated that in the exercise of their prosecutorial discretion they will not presume that a patent, copyright, or trade secret necessarily confers market power upon its owner" (internal quotation marks omitted)).

45. 447 U.S. 303 (1980).

engineered bacterium” is patentable subject matter,⁴⁶ the USPTO concluded that there is no subject-matter bar to the patenting of any particular type of living thing outside of a human being.⁴⁷ The courts have effectively affirmed the USPTO’s conclusion with respect to the patentability of plants.⁴⁸ Congress, meanwhile, has repeatedly reinforced the USPTO’s policy regarding human beings by enacting appropriations riders that forbid the use of federal funds for the granting of a patent directed to such subject matter.⁴⁹

Another area in which USPTO guidance has been influential has been with respect to the meaning of patent law’s utility requirement. The late 1990s generated a flood of applications for patent rights on “expressed sequence tags” (ESTs), fragments of DNA that had been isolated and were generally known to hybridize with complementary DNA sequences, but whose specific biological purpose and significance were often unknown.⁵⁰ The USPTO responded with utility guidelines that interpreted and applied a 1966 Supreme Court decision⁵¹ in a way that gave teeth to the utility requirement in the realm of biological- or chemical-substance inventions.⁵² In particular, the utility guidelines instructed that to satisfy patent law’s utility requirement, a claimed biological or chemical substance needs, at the time of application, to have a known utility that is “specific and substantial”—i.e., that is specific to the substance in question and not a “‘throw-away,’ ‘insubstantial,’ or ‘nonspecific’ utilit[y], such as the use of a complex invention as landfill.”⁵³ When rejections of EST claims based on these

46. *Id.* at 305, 318.

47. See MPEP, *supra* note 28, § 2105, at 2100-3 (“It is clear from the Supreme Court decision and opinion that the question of whether or not an invention embraces living matter is irrelevant to the issue of patentability.”); *id.* § 2106, at 2100-5 (reporting a 1987 notice indicating “that the Patent and Trademark Office would now consider nonnaturally occurring, nonhuman multicellular living organisms, including animals, to be patentable subject matter”).

48. *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int’l, Inc.*, 534 U.S. 124, 127 (2001) (“We hold that utility patents may be issued for plants.”).

49. See John M. Golden, *WARF’s Stem Cell Patents and Tensions Between Public and Private Sector Approaches to Research*, 38 J.L. MED. & ETHICS 314, 318 (2010) (“Congress has repeatedly enacted appropriations bill riders that forbid the USPTO to use funds to issue patent claims directed to or encompassing a human organism.” (internal quotation marks omitted)). The USPTO’s conclusion regarding the subject-matter eligibility of the vast array of organisms that are not bacteria, plants, or human beings—including genetically altered oysters, mice, and cows, see MPEP, *supra* note 28, § 2105, at 2100-4 to 2100-5 (explaining that nonnaturally occurring organisms are patentable, including a polyploid Pacific coast oyster); *Transgenic Non-Human Mammals*, U.S. Patent No. 4,736,866 col. 1 (filed June 22, 1984); *Transgenic Bovines and Milk from Transgenic Bovines*, U.S. Patent No. 6,013,857 (filed June 5, 1995)—apparently have not yet been subjected to decisive scrutiny, at least explicitly, by either the U.S. Congress or the federal judiciary.

50. See ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, *PATENT LAW AND POLICY: CASES AND MATERIALS* 250–53 (3d ed. 2002) (discussing the controversy over ESTs).

51. *Brenner v. Manson*, 383 U.S. 519 (1966).

52. Utility Examination Guidelines, 66 Fed. Reg. 1092 (Jan. 5, 2001).

53. *Id.* at 1098.

guidelines were appealed, a Federal Circuit panel not only affirmed the rejections, but found the USPTO's guidelines to comport with judicial precedent, thereby essentially adopting the USPTO's reasoning as the circuit's own.⁵⁴

As these examples demonstrate, with respect to questions of patentability, the USPTO can and does play a significant role both in setting the legal agenda and in developing approaches to applying the law that can stick. Nonetheless, the USPTO remains handicapped relative to agencies that have substantive rulemaking authority or, at least, enforcement discretion. The USPTO must always work within the bounds of preexisting judicial precedent; generally speaking, the USPTO cannot override Supreme Court or Federal Circuit decisions or adopt an explicit policy of under-enforcement. The result is a sharp limitation on the range of interpretive options that the USPTO can explore, and thus on the aid that the USPTO can provide to courts or legislators willing and able to consider updating the law.

B. Efforts to Expand USPTO Authority and Their Critique

Limits on the USPTO's rulemaking authority have not gone unnoticed. On August 21, 2007, the USPTO initiated a test of these limits by issuing rules that, among other things, would have regulated the extent to which patent applicants could file serial continuing applications or requests for continued examination, mechanisms by which U.S. patent applicants can prolong examination—for example, so that they can continue to contest patent-examiner rejections without taking an appeal to the BPAI or the Federal Circuit.⁵⁵ Citing statutory provisions and judicial precedent discussing applicants' statutorily provided capacities to file continuation applications or requests for continued examination, third parties promptly challenged the USPTO's new rules as beyond the agency's authority.⁵⁶

The U.S. House of Representatives took note. On September 7, 2007, the House passed a patent-reform bill that, among other things, explicitly

54. See *In re Fisher*, 421 F.3d 1365, 1372 (Fed. Cir. 2005) (“The [US]PTO’s standards for assessing whether a claimed invention has a specific and substantial utility comport with this court’s interpretation”); *id.* at 1374 (“We agree with the [BPAI] that the facts here are similar to those in *Brenner*.”).

55. Changes to Practice for Continued Examination Filings, Patent Applications Containing Patentably Indistinct Claims, and Examination of Claims in Patent Applications, 72 Fed. Reg. 46,716, 46,716 (Aug. 21, 2007) (“The Office is revising the rules of practice to require that any third or subsequent continuing application . . . , and any second or subsequent request for continued examination in an application family, . . . be supported by a showing as to why the amendment, argument, or evidence sought to be entered could not have been previously submitted.”).

56. See *Tafas v. Dudas*, 541 F. Supp. 2d 805, 810 (E.D. Va. 2008) (describing complaints against the USPTO filed by Triantafyllos Tafas and a pharmaceutical company), *affirmed in part and vacated in part sub nom.* *Tafas v. Doll*, 559 F.3d 1345 (Fed. Cir. 2009), *vacated and petition for reh’g en banc granted*, 328 F. App’x 658, *dismissed as moot sub nom.* *Tafas v. Kappos*, 586 F.3d 1369 (Fed. Cir. 2009).

stated that the preexisting grant of rulemaking authority to the USPTO encompassed a grant of “the authority to promulgate regulations to ensure the quality and timeliness of applications and their examination, including specifying circumstances under which an application for patent may claim the benefit . . . of the filing date of a prior filed application”⁵⁷ pursuant to U.S. Patent Act provisions regarding continuing, divisional, and international applications.⁵⁸ A House Judiciary Committee report dated September 4, 2007, explained that “the USPTO has always had authority to promulgate rules that place limitations or conditions on patent applications,” and indicated its members’ expectation that the USPTO would receive deference from the courts with respect to such rules.⁵⁹

The House bill did not become law, however. Moreover, a district court sided with the challengers to the USPTO rules, declaring that they were “substantive in nature and exceed[ed] the scope of the USPTO’s rulemaking authority.”⁶⁰ A divided panel of the Federal Circuit substantially disagreed, finding the challenged rules to be procedural, rather than substantive.⁶¹ Nonetheless, the panel affirmed the district court’s judgment that one particular rule was invalid because, in the panel’s view, that rule conflicted with “clear and unambiguous” language in the U.S. Patent Act.⁶² The en banc Federal Circuit subsequently vacated the panel decision⁶³ and then dismissed the case as moot after the USPTO, under a new director and a new President, rescinded the rules.⁶⁴

The spat over the USPTO’s 2007 rules package thus yielded neither lasting Federal Circuit precedent nor enacted legislation. But the push to clarify or to expand the bounds of USPTO rulemaking authority was not an isolated incident.

Even before the USPTO issued its rules package in August 2007, a number of members of the House had contemplated giving the USPTO the sort of comprehensive rulemaking authority that many contemporaneous administrative agencies enjoy. An earlier version of the 2007 House bill had

57. Patent Reform Act of 2007, H.R. 1908, 110th Cong. § 14 (as passed by House, Sept. 7, 2007).

58. 35 U.S.C. §§ 1–376 (2006).

59. H.R. REP. NO. 110-314, at 87 (2007) (“Typically, courts will give deference to the rules promulgated by agencies if an Act which charges the agency with responsibility doesn’t speak to the issue that the rule pertains to.”).

60. *Tafas*, 541 F. Supp. at 811.

61. *Tafas*, 559 F.3d at 1356 (“[W]e conclude that the Final Rules challenged in this case are procedural.”).

62. *Id.* at 1360.

63. *Tafas v. Doll*, 328 F. App’x 658, 658 (Fed. Cir. 2009), *dismissed as moot sub nom.* *Tafas v. Kappos*, 586 F.3d 1369 (Fed. Cir. 2009).

64. *Tafas*, 586 F.3d at 1371 (dismissing the appeal as moot but declining to vacate the district court’s judgment because “it was the USPTO (the losing party in the district court action) that acted unilaterally to render the case moot”).

proposed enactment of the following provision on USPTO rulemaking authority:

(5) Regulatory Authority.—In addition to the authority conferred by the other provisions of this title, the Director may promulgate such rules, regulations, and orders that the Director determines appropriate to carry out the provisions of this title or any other law applicable to the United States Patent and Trademark Office or that the Director determines necessary to govern the operation and organization of the Office.⁶⁵

To the extent the original House bill intended to give the USPTO rulemaking authority with respect to all portions of the Patent Act, its proposed grant of authority was wildly disproportionate to that necessary for the USPTO to implement rules regulating continuing applications and requests for continued examinations. Further, as Dan Burk and Mark Lemley have pointed out, such a comprehensive delegation would have given the USPTO interpretive authority with respect to parts of the U.S. Patent Act with which the courts have substantially greater expertise and responsibility.⁶⁶ For example, the USPTO has historically had no direct involvement with determinations of whether an accused infringer's conduct in fact constitutes infringement under Section 271 of the Act.⁶⁷ Instead, federal courts have been exclusively charged with the task of determining questions of infringement, as well as deciding on remedies for infringement in accordance with Sections 282 through 289.⁶⁸ Because the original House bill did not propose to give the USPTO any new responsibility for determinations specific to the enforcement context, its apparent proposal to provide rulemaking authority with respect to such determinations might have been unintended but, at the very least, appears difficult to justify. In any event, the broad language of subsection 5 vanished in later versions of the House bill, giving way to the far narrower provisions passed by the House in September 2007.

The failure of the 2007 House bill to win enactment by the Senate does not mean efforts to enhance the USPTO's rulemaking authority are entirely dead. A Senate version of the Patent Reform Act of 2007⁶⁹ and both House⁷⁰

65. Patent Reform Act of 2007, H.R. 1908, 110th Cong. § 11 (as introduced Apr. 18, 2007).

66. BURK & LEMLEY, *supra* note 1, at 107 (describing the USPTO as “by design see[ing] only one piece of the patent puzzle—the question of whether a patent should issue in the first place”).

67. 35 U.S.C. § 271 (2006) (defining forms of conduct that can constitute infringement of U.S. patent rights).

68. *Id.* §§ 283–89 (providing and limiting remedies for infringement of utility and design patents).

69. Patent Reform Act of 2007, S. 1145, 110th Cong. § 9 (2008).

70. Patent Reform Act of 2009, H.R. 1260, 111th Cong. § 11 (2009).

and Senate⁷¹ versions of a later effort at patent-reform legislation, the Patent Reform Act of 2009, all contained a provision to grant the USPTO broader fee-setting authority. In identical language, these bills proposed to give the Director

authority to set or adjust by rule any fee established or charged by the Office under sections 41 and 376 . . . for the filing or processing of any submission to, and for all other services performed by or materials furnished by, the Office, provided that such fee amounts are set to reasonably compensate the Office for the services performed.⁷²

To try to ensure USPTO fee setting would be well-considered, the bills set in place various procedural requirements. Before setting fees through the new rulemaking authority, the USPTO would have to consult with the Patent Public Advisory Committee,⁷³ a body with “nine voting members who . . . serve at the pleasure of the Secretary of Commerce.”⁷⁴ Further, proposed fee changes would be subjected to both (1) public comment following publication in the Federal Register of the proposed change and its “specific rationale and purpose,” and (2) congressional comment after notification of “the Chair and Ranking Member of the Senate and House Judiciary Committees.”⁷⁵

Thus, members of Congress have continued to consider giving the USPTO greater rulemaking authority than exists under current law. Part III argues that the retreat from the 2007 House bill’s overly broad proposal to give the USPTO comprehensive rulemaking authority has gone too far. Congress should provide the USPTO with rulemaking authority on issues relating to at least one particular branch of patentability questions—namely, questions of subject-matter eligibility that look to resolve the categories of innovations that can properly be the subject of patent rights.

III. The Case for Binding USPTO Authority over Subject-Matter Eligibility

This Part argues that Congress should give the USPTO primary interpretive authority for questions of subject-matter eligibility. The argument proceeds in two stages. First, subpart III(A) explores the nature of inquiries into subject-matter eligibility; determines that, under existing law and as generally conceived, such inquiries are comparatively categorical, rather than richly innovation specific; and shows how use of such a coarse-grained filter can improve the functioning of a patent system even when finer

71. Patent Reform Act of 2009, S. 515, 111th Cong. § 9 (2009); *accord* Patent Reform Act of 2009, S. 610, 111th Cong. § 9 (2009).

72. S. 515 § 9(a)(1); *accord* S. 610 § 9(a)(1); H.R. 1260 § 11(a)(1); S. 1145 § 9(a)(1).

73. S. 515 § 9(a)(2)–(3); *accord* S. 610 § 9(a)(2)–(3); H.R. 1260 § 11(a)(2)–(3); S. 1145 § 9(a)(2)–(3).

74. 35 U.S.C. § 5(a)(1) (2006).

75. *See* S. 515 § 9(a)(3)(E)–(5) (providing for public and congressional notice and comment for fee changes); *accord* S. 610 § 9(a)(3)(E)–(5); H.R. 1260 § 11(a)(3)(E)–(5); S. 1145 § 9(a)(3)(E)–(5).

grained filters for “patent-worthiness” are also present. Subpart III(B) then contends that the USPTO is the best institution to have primary interpretive authority for such a coarse-grained screen.

A. Nature and Justification of Patent Law’s Subject-Matter Inquiry

1. *The Patent Bargain and Its Finer Grained Requirements.*—Patent law is frequently described as offering a rights-for-disclosure bargain in which society provides limited-term rights in exchange for public disclosure of an invention and means of its creation and use.⁷⁶ Consistent with this description, patent law features a series of requirements for patentability that, at least according to a utilitarian understanding, are meant to ensure that the average benefits from disclosure and innovative effort stimulated by patents are greater than the average social costs—including administrative costs—that patents generate.⁷⁷ These requirements include the relatively fine-grained requirements of novelty, nonobviousness, adequate disclosure, and utility, which are discussed in this subsection. They also include the coarser-grained requirement of patentable subject matter, or subject-matter eligibility, which the next subsection discusses.

Patent law’s requirements that an invention be novel, nonobvious, and adequately disclosed commonly demand detailed, fact-intensive inquiries into the state of the prior art; the level of knowledge of one of ordinary skill in the art; and, in the United States, the date of invention itself. Consequently, assessments of novelty, nonobviousness, and adequacy of disclosure typically rely on a rich set of “adjudicative facts”—“facts that help the decision-maker establish what happened at a particular time and place.”⁷⁸ Such facts contrast with so-called “legislative facts”—facts less specific to a particular case or party “that help a decision-maker decide questions of law and policy.”⁷⁹ Although the boundaries between legislative and adjudicative facts are necessarily fuzzy, a decision maker’s predominant concern with one or the other of these categories of facts is commonly thought to suggest the

76. See, e.g., John M. Golden, “Patent Trolls” and Patent Remedies, 85 TEXAS L. REV. 2111, 2117 (2007) (“In general, patents are believed to serve their constitutional purpose . . . by providing for an exchange: in exchange for public disclosure . . . , the public provides a limited-term ‘right to exclude’ . . .”).

77. Cf. STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW 152 (2004) (“An economic rationale for the novelty requirement is that products or processes that do not satisfy it are either ones that society already enjoys or will soon possess . . . so that society has no reason to suffer the losses associated with patent monopoly.”).

78. RICHARD J. PIERCE, JR., SIDNEY A. SHAPIRO & PAUL R. VERKUIL, ADMINISTRATIVE LAW AND PROCESS § 3.6.1, at 69–70 (3rd ed. 1999).

79. *Id.* at 69; see also *id.* § 6.4.3d, at 306 (“Legislative facts are not unique to a particular dispute but are general facts that are useful in permitting informed interpretation and application of statutory and constitutional provisions.”).

kinds of institutions and processes—e.g., legislative or adjudicatory—that are likely to be best for the task.⁸⁰

Of course, the tendency of questions of novelty, nonobviousness, and adequacy of disclosure to turn on a detailed set of adjudicative facts is not absolute. For example, the U.S. Supreme Court's recent decision on nonobviousness in *KSR International Co. v. Teleflex Inc.*⁸¹ might be understood as both emphasizing and relying on a perceived "legislative fact"—namely, that a person skilled in a technological or scientific art generally possesses not only book learning that likely appears in publicly available documents, but also common sense of the trade that likely extends beyond what has been publicly documented.⁸² Similarly, in determining whether a patent's written description provides sufficient disclosure, the courts appear frequently to rely on relatively generic characterizations of the level of development of an art and its artisans' average skill level.⁸³

But such exceptions should not obscure the general rule. Inquiries into novelty and nonobviousness typically require painstaking review of prior-art materials that are specifically related to the claimed invention.⁸⁴ Hence, even after the Supreme Court's *KSR* opinion reasserts the role of common sense in the nonobviousness analysis, the opinion engages in detailed analysis of the contents of multiple prior-art documents before presenting the Court's conclusion that, at the time of invention, the claimed invention was obvious as a matter of law.⁸⁵ Likewise, questions about the adequacy of a patent's

80. *Cf. id.* § 6.4.1, at 279 ("Adjudication is best-suited to . . . resolution of factual disputes between individuals or between an agency and an individual. Informal rulemaking is best-suited to . . . establishment of rules applicable to groups of people.").

81. 550 U.S. 398 (2007).

82. *See id.* at 419 ("In many fields it may be that there is little discussion of obvious techniques or combinations, and it may be the case that market demand, rather than scientific literature, will drive design trends.").

83. *See* Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1650 (2003) ("Overwhelming evidence indicates that the application of the [person-having-ordinary-skill-in-the-art] standard varies by industry, leading for example to fewer, but broader, software patents and more, but narrower, biotechnology patents.").

84. *See* 35 U.S.C. § 103(a) (2006) (stating that the nonobviousness requirement prohibits granting patent rights when "the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the [relevant] art"); KIMBERLY A. MOORE ET AL., *PATENT LITIGATION AND STRATEGY* 505–06 (3d ed. 2008) ("To succeed in demonstrating that a claim [lacks novelty], the accused infringer must show (1) that each and every limitation of the claim is found in a single prior art reference (such as a different patent or a scientific article) and (2) that the reference . . . discloses the invention in a fashion sufficient to allow a person of ordinary skill in the relevant technological art to create the invention without undue experimentation.").

85. *See KSR*, 550 U.S. at 422–26 (explaining why the claimed invention was correctly found obvious based on the prior art).

disclosure tend to demand detailed study of that disclosure and its relation to specific patent claims.⁸⁶ Testimony by qualified experts is often vital.⁸⁷

The patentability requirement of utility seems a more intermediate case. Generally speaking, the constraints from the utility requirement are so minimal that often very few specific facts are needed to satisfy it. U.S. patent law requires that an invention have “specific and substantial utility”—in other words, that it be “useful for [some] particular practical purpose.”⁸⁸ Relatively generic attributes of an overwhelming percentage of claimed inventions—e.g., a new form of mousetrap, eyeglass lens, or elevator—will suffice to establish such usefulness. Indeed, the utility requirement presents virtually no obstacle to the patenting of most mechanical, electrical, or computer-related inventions because their possession of the required utility is obvious on their face.⁸⁹ Even cutting-edge technologies tend to produce utility questions that are relatively generic. As illustrated by recent debates relating to patent applications on genetic sequences such as ESTs, new technologies or products of new technologies can generate widespread assertion of new forms of utility that lend themselves to consideration in a categorical

86. See, e.g., *ALZA Corp. v. Andrx Pharm., LLC*, 603 F.3d 935, 941 (Fed. Cir. 2010) (“In arguing that the disclosure in the ’373 patent specification does enable a person of ordinary skill to make and use the claimed dosage forms, ALZA directs us to ten lines of the specification, which mention non-osmotics and refer to a textbook discussing how to make and use various types of non-osmotic sustained-release dosage forms.”); *Ariad Pharm., Inc. v. Eli Lilly & Co.*, 598 F.3d 1336, 1355–56 (Fed. Cir. 2010) (en banc) (“The specification of the ’516 patent hypothesizes three classes of molecules potentially capable of reducing NF-κB activity We review the specification’s disclosure of each . . . to determine whether there is substantial evidence to support the jury’s verdict that the written description evidenced that the inventor possessed the claimed invention.”).

87. See, e.g., *ALZA*, 603 F.3d at 942 (“We find no clear error in the district court crediting the testimony of Lam over Dr. Davies and Ayer [in relation to questions of enablement].”); *Sitrick v. Dreamworks, LLC*, 516 F.3d 993, 1000–01 (Fed. Cir. 2008) (finding that “[d]efendants’ two experts explained that one skilled in the art would not be able to take the teachings regarding video games and apply them to movies,” and that a contrary expert opinion “did not raise a triable issue of fact because it was: (1) ‘conclusory’ . . . and (2) presented by a person who ‘admitted to not being skilled in the art’”).

88. MPEP, *supra* note 28, § 2107, at 2100–20 (instructing examiners that “if the applicant has asserted that the claimed invention is useful for any particular purpose (i.e., it has a ‘specific and substantial utility’) and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility”).

89. See F. SCOTT KIEFF ET AL., *PRINCIPLES OF PATENT LAW: CASES AND MATERIALS* 740 (4th ed. 2008) (“[T]he utility requirement is not usually an obstacle for mechanical and electrical applications,” but “poses a greater concern for chemical and biological inventions”); MOORE ET AL., *supra* note 84, at 504 (observing that “the issue of utility is unlikely to arise in an electrical or mechanical invention, whose uses are readily apparent”). Exceptions can occur when an inventor claims a utility that is incredible—for example, because it would violate well-established understandings of physics. *MERGES & DUFFY*, *supra* note 50, at 208 (“An applicant for a patent cannot satisfy the utility requirement merely by advancing wonderful allegations.”). Thus, the USPTO has rejected applications for patents involving “an invention asserted to change the taste of food using a magnetic field . . . , a perpetual motion machine[,] . . . and a method of controlling the aging process.” MPEP, *supra* note 28, § 2107.01, at 2100-24 to -25. The USPTO instructs examiners, however, that such “examples are fact specific and should not be applied as a *per se* rule.” *Id.* § 2107.01, at 2100-25.

manner.⁹⁰ It should perhaps come as no surprise that, where a generic form of utility is asserted—such as the utility of an EST sequence in detecting a complementary sequence with which it will hybridize—the response can be a generic form of legal analysis.

Nevertheless, the test of utility is generally designed to be individualized. Indeed, a key aspect of the legal response to ESTs was emphasis on the need for an asserted utility that is specifically associated with the claimed invention.⁹¹ Thus, for purposes of patent law's utility requirement, it is not enough to assert that claimed ESTs can be used to detect complementary sequences. Any genetic sequence has this property. But only some sequences, for example, have the specific utility of being useful for forensic purposes because they are associated with portions of the genetic code that differ significantly between individuals.⁹²

In sum, although the relatively minimal nature of the utility requirement might often obscure the point, the requirement of a *specific* and substantial utility fundamentally demands a particularized inquiry that can turn on quite detailed adjudicative facts.⁹³ Hence, at least for purposes here, the utility requirement seems properly grouped with the novelty, nonobviousness, and adequate-disclosure requirements as an adjudicative-fact-oriented inquiry.

2. *The Coarse-Grained Filter for Patentable Subject Matter.*—Questions of patentable subject matter tend to have a different flavor. Discussion of relevant adjudicative facts is commonly thin, even when the issue is difficult. A basic explanation for this is not hard to find.⁹⁴ Unlike requirements of novelty, nonobviousness, and utility, subject-matter eligibility does not require anything distinctive or specific to the claimed invention. Instead, subject-matter eligibility requires that the claimed invention belong to one or more broadly drawn categories of things deemed potentially patentable.⁹⁵ As the USPTO has recently emphasized, the requirement of subject-matter eligibility contrasts with other patentability requirements in

90. See *supra* text accompanying notes 50–53.

91. See *supra* text accompanying note 46.

92. See Sonia M. Suter, *All in the Family: Privacy and DNA Familial Searching*, 23 HARV. J.L. & TECH. 309, 314 (2010) (“Although humans are more genetically alike than different—we share over 99.9% of our genetic material with one another—a fair amount of genetic variability, or polymorphism, exists in certain portions of the genome.”).

93. Cf. *In re Brana*, 51 F.3d 1560, 1563–64 (Fed. Cir. 1995) (discussing the significance of in vivo tests using mice and in vitro tests involving “human tumor cells” in the process of assessing a claimed invention’s utility).

94. See *infra* notes 215–19 and accompanying text.

95. Cf. MERGES & DUFFY, *supra* note 50, at 65 (“Our inquiry in this chapter [on patentable subject matter] will focus on general classes of inventions.”); 1 WILLIAM CALLYHAN ROBINSON, THE LAW OF PATENTS FOR USEFUL INVENTIONS § 69, at 105 (1890) (entitling a section on subject-matter eligibility “No Invention Patentable unless Embraced within one of the Prescribed Classes” (emphasis omitted)).

being “merely a coarse filter,” one that involves “only a threshold question for patentability.”⁹⁶

For purposes of determining subject-matter eligibility, the focus is on a patent’s claim language.⁹⁷ To the extent claim construction is required before subject-matter eligibility can be decided, there might be a need to consider (a) “intrinsic evidence” appearing in the remainder of the patent or its prosecution history, and possibly also (b) “extrinsic evidence” that can come in such forms as dictionaries, scientific treatises, or expert testimony.⁹⁸ But under current law, claim construction is formally considered to be a “purely legal” exercise for a judge, not a task involving true adjudicative fact-finding.⁹⁹ Moreover, and perhaps more to the point, subject-matter determinations tend not to be so sensitive to the fine problems of claim construction that arise from efforts to encompass or to distinguish specific pieces of prior art or specific accused products or processes. Just as a micrometer will rarely be needed to help determine whether the distance between two cities is less than 500 miles, claim construction often need occur at only a crude, perhaps largely inchoate level in order to establish whether a claim passes the requirement of subject-matter eligibility.

Once a claim has been construed to whatever extent necessary, the primary subject-matter eligibility question becomes whether the subject matter encompassed by the patent claim falls within at least one of four statutory categories of patentable subject matter: those for a “process, machine, manufacture, or composition of matter.”¹⁰⁰ A second question, presuming the statutory-category requirement is met on its face, is whether the

96. Interim Guidance for Determining Subject Matter Eligibility for Process Claims in View of *Bilski v. Kappos*, 75 Fed. Reg. 43,922, 43,926 (July 27, 2010); see also Thomas F. Cotter, *A Burkean Perspective on Patent Eligibility, Part II: Reflections on the (Counter) Revolution in Patent Law*, 11 MINN. J.L. SCI. & TECH. 365, 379 (2010) (describing subject-matter eligibility as “a crude filter for carrying out social policy”).

97. Pamela Samuelson & Jason Schultz, “Clues” for Determining Whether Business and Service Innovations Are Unpaten[t]able Abstract Ideas, 15 LEWIS & CLARK L. REV. (forthcoming) (manuscript at 23), available at <http://ssrn.com/abstract=1690882> (“Unlike rejections under [35 U.S.C.] § 102 (lack of novelty) or § 103 (obviousness), § 101 determinations are made on the claim language alone and thus do not require extensive research into the state of the art or the details of other documents.”).

98. See *Phillips v. AWH Corp.*, 415 F.3d 1303, 1317 (Fed. Cir. 2005) (en banc) (“Although we have emphasized the importance of intrinsic evidence in claim construction, we have also authorized district courts to rely on extrinsic evidence, which consists of all evidence external to the patent and prosecution history, including expert and inventor testimony, dictionaries, and learned treatises.” (internal quotation marks omitted)).

99. *Cybor Corp. v. FAS Techs., Inc.*, 138 F.3d 1448, 1456 (Fed. Cir. 1998) (en banc) (holding that claim construction is a “purely legal question” reviewed de novo on appeal).

100. 35 U.S.C. § 101 (2006); see *Bilski v. Kappos*, 130 S. Ct. 3218, 3225 (2010) (“Section 101 . . . specifies four independent categories of inventions or discoveries that are eligible for protection”); *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470, 483 (1974) (reasoning on the basis of the proposition that “no patent is available for a discovery . . . unless it falls within one of the express categories of patentable subject matter of 35 U.S.C. § 101”).

encompassed subject matter overlaps with “three specific exceptions to § 101’s broad patent-eligibility principles: ‘laws of nature, physical phenomena, and abstract ideas.’”¹⁰¹ If a claim encompasses matter within the scope of an exception, the claim fails the requirement of subject-matter eligibility.

Approaches to determining subject-matter eligibility are fundamentally categorical. The inquiry tends to proceed by classifying the claimed invention as a member of a type—for example, a method of tax planning,¹⁰² a business-method patent,¹⁰³ a method for “organizing human activity,”¹⁰⁴ an isolated but otherwise naturally occurring genetic sequence,¹⁰⁵ a genetically engineered bacterium,¹⁰⁶ a computer program,¹⁰⁷ or a general-purpose computer loaded with a computer program.¹⁰⁸ The decisive question then tends to be whether the identified type falls within a statutory category or overlaps with an exception. Hence, for the Supreme Court in *Diamond v. Chakrabarty*, the key point was that the claimed invention was an artificially engineered bacterium that qualified as a human-made “manufacture” or “composition of matter.”¹⁰⁹ In *Diamond v. Diehr*,¹¹⁰ the claimed invention was an industrial process of curing rubber that was the sort of “process” contemplated by the Patent Act.¹¹¹ In *Bilski v. Kappos*, on the other hand, the claimed invention was “the concept of hedging risk and the application of

101. *Bilski*, 130 S. Ct. at 3225; see also Efthimios Parasidis, *A Uniform Framework for Patent Eligibility*, 85 TUL. L. REV. 323, 326 (2010) (“[A]n invention is patent-eligible if (1) it corresponds to a statutory category outlined in section 101 of the Patent Act . . . and (2) does not violate the product of nature doctrine, which precludes eligibility for laws of nature, natural phenomenon, mental processes, and abstract ideas.”).

102. *E.g.*, U.S. Patent No. 7,693,769 (filed Dec. 24, 2004).

103. *E.g.*, U.S. Patent No. 7,865,392 (filed July 14, 2004).

104. Brief for the Respondent at 8, *Bilski v. Kappos*, 130 S. Ct. 3218 (2010) (No. 08-964), 2009 WL 3070864, at *8 (arguing that the U.S. Patent Act’s “term ‘process’ encompasses all technological and industrial processes,” but “does not extend patent-eligibility . . . to methods of organizing human activity that are untethered to technology”).

105. *E.g.*, U.S. Patent No. 7,863,501 (filed Sept. 30, 2005).

106. *E.g.*, U.S. Patent No. 4,259,444 (filed June 7, 1972).

107. *E.g.*, U.S. Patent No. 7,870,612 (filed Sept. 11, 2006).

108. *E.g.*, U.S. Patent No. 5,007,101 (filed Dec. 28, 1982).

109. See *Diamond v. Chakrabarty*, 447 U.S. 303, 313 (1980) (holding that “the relevant distinction [is] not between living and inanimate things, but between products of nature . . . and human-made inventions,” and observing that Chakrabarty’s “micro-organism is the result of human ingenuity and research”).

110. 450 U.S. 175 (1981).

111. See *id.* at 184 (describing the claimed invention as “a physical and chemical process for molding precision synthetic rubber,” and stating that “[i]ndustrial processes such as this are the types which have historically been eligible to receive the protection of our patent laws”).

that concept to energy markets,”¹¹² and this concept was an “abstract idea”—a member of the classes of objects excluded from subject-matter eligibility.¹¹³

The subject-matter inquiry thus tends to be predominantly coarse-grained and categorical—a far cry from the highly individualized comparison of claims to prior art that dominates inquiries such as novelty and nonobviousness. On the other hand, the subject-matter inquiry’s tendency toward coarse-grained inquiry is less than absolute. The U.S. Supreme Court has added wrinkles to the subject-matter inquiry that can make the inquiry adjudicative-fact rich: (1) a “no preemption” rule that a patent claim, even if otherwise avoiding the exceptions, cannot “pre-empt substantially all uses of [a] fundamental principle”;¹¹⁴ and (2) a “significance filter,” implementing a rule that a patent claim for a process cannot satisfy the requirements of subject-matter eligibility simply as a result of its incorporation of one or more steps involving “insignificant extra-solution activity.”¹¹⁵ The no-preemption rule can trigger a detailed inquiry into two related matters: (a) the scope of conceivable uses of a “fundamental principle” that the claimed invention employs, and (b) the capacity of others to exploit the fundamental principle in a way not encompassed by the patent claims. Likewise, because “significance” is apparently associated with concerns of novelty and nonobviousness, the significance filter can entangle the subject-matter inquiry in novelty and nonobviousness inquiries that are commonly adjudicative-fact rich.¹¹⁶

Do these wrinkles mean that the facially categorical nature of subject-matter inquiries is no more than a façade? My sense is that the proper answer is “No.” Both the courts and the USPTO have indicated that these wrinkles are largely intended to provide doctrinal safety valves that provide official decision makers with some room—but substantially limited room—to ensure that the substance of subject-matter restrictions is not gutted through strategic claim drafting.¹¹⁷ Because these safety valves appear

112. *Bilski v. Kappos*, 130 S. Ct. 3218, 3229 (2010).

113. *See id.* at 3231 (holding that the “concept of hedging . . . is an unpatentable abstract idea” and reaffirming that “limiting an abstract idea to one field of use or adding token postsolution components [does] not make the concept patentable”).

114. *See In re Bilski*, 545 F.3d 943, 953 (Fed. Cir. 2008) (en banc) (concluding that Supreme Court precedent “drew a distinction between those claims that ‘seek to pre-empt the use of’ a fundamental principle . . . and claims that seek only to foreclose others from using a particular ‘application’ of that fundamental principle”), *aff’d on other grounds sub nom.* *Bilski v. Kappos*, 130 S. Ct. 3218 (2010).

115. *Bilski*, 545 F.3d at 962 (citing Supreme Court precedent for the proposition that, under a machine-or-transformation test, “the involvement of the machine or transformation in the claimed process must not merely be insignificant extra-solution activity”).

116. *See supra* text accompanying notes 77–78.

117. *Cf.* John R. Thomas, *The Patenting of the Liberal Professions*, 40 B.C. L. REV. 1139, 1152 (1999) (“To the extent that the prohibition against patenting ideas presents sound policy, allowing applicants to avoid these limitations through artful claim drafting appears unwise.”).

intended for application only in extreme cases that are not likely to require detailed analysis of facts specific to the claimed invention, their existence seems an insufficient basis for rejecting characterization of subject-matter eligibility as a fundamentally coarse-grained and categorical inquiry.¹¹⁸

Understanding the significance filter and no-preemption rule as safety valves enables practical reconciliation of crosscurrents in the judicially developed law on subject-matter eligibility. The continuing force of such crosscurrents was evident in the Supreme Court's 2010 opinion in *Bilski v. Kappos*. In that opinion, the Court both acknowledged the continuing viability of the significance filter¹¹⁹ and re-emphasized another principle: "the need to consider the invention as a whole, rather than 'dissect[ing] the claims into old and new elements'" when assessing subject-matter eligibility.¹²⁰ Although the Court's significance filter seems capable of effecting just such a dissection, the Court provided no explanation of the potential contradiction. A reason for such silence might be an understanding that the significance filter has only a safety-valve role: the claim-as-a-whole principle establishes a strong default approach, and the significance filter kicks in only where two conditions hold: (1) a claim limitation can be severed from the remainder of the claim without depriving that remainder of "life and meaning,"¹²¹ and (2) there is no plausible argument that the severable limitation has a role, either separately or through its relation to other claim limitations, in establishing the claimed invention's novelty or nonobviousness.¹²²

118. Cf. Henry E. Smith, *Institutions and Indirectness in Intellectual Property*, 157 U. PA. L. REV. 2083, 2129–30 (2009) (acknowledging that "traditional equitable analysis . . . provides a narrow safety valve" for denying injunctions against continuing encroachment of property rights, but contending that "the traditional equitable test . . . is not an invitation to engage in a case-by-case policy analysis").

119. *Bilski*, 130 S. Ct. at 3230 ("*Flook* stands for the proposition that the prohibition against patenting abstract ideas 'cannot be circumvented by attempting to limit the use of [such an idea] to a particular technological environment' or adding 'insignificant postsolution activity.'" (quoting *Diamond v. Diehr*, 450 U.S. 175, 191–92 (1981))).

120. *Id.* (quoting *Diehr*, 450 U.S. at 188).

121. The "life and meaning" language is borrowed from Federal Circuit jurisprudence holding that a claim's preamble is generally not viewed as limiting the scope of the claim unless the preamble "recites essential structure or steps, or . . . is necessary to give life, meaning, and vitality to" the claim. *Am. Med. Sys., Inc. v. Biolitec, Inc.*, 618 F.3d 1354, 1358 (Fed. Cir. 2010) (internal quotation marks omitted); see also *Innova/Pure Water, Inc. v. Safari Water Filtration Sys., Inc.*, 381 F.3d 1111, 1118 (Fed. Cir. 2004) ("Language in a preamble limits a claim where it breathes life and meaning into the claim, but not where it merely recites a purpose or intended use of the invention." (citation omitted)).

122. Cf. Robert Plotkin, *A History of Software Patents*, in THE HISTORY OF INFORMATION SECURITY: A COMPREHENSIVE HANDBOOK 141, 153 (Karl de Leeuw & Jan Bergstra eds., 2007) (discussing "two different approaches that courts have taken in analyzing software patent claims: (1) analyzing the claim as a whole to determine whether it qualifies as patentable subject matter . . . ; and (2) dissecting the claim and analyzing part of it to determine whether it qualifies as patentable subject matter").

The sort of extreme circumstances that might give rise to the significance filter's application are perhaps best illustrated through the facts of the case that firmly ensconced the filter in U.S. patent law, *Parker v. Flook*.¹²³ In this case, the Supreme Court found a failure to satisfy subject-matter eligibility where a method claim recited (1) a series of steps for calculating a threshold value above which an alarm would be triggered, and (2) a step consisting of "[a]djusting said alarm limit to said updated alarm limit value."¹²⁴ The Court apparently believed that there was no plausible basis for arguing that the latter adjustment step, either alone or through combination with the other steps, helped establish the claimed invention's novelty or nonobviousness.¹²⁵ After all, why would one calculate a new alarm limit value without doing such updating? The Court did not feel a need to engage in detailed inquiry before concluding that the adjustment step was properly viewed as "insignificant post-solution activity [that could] not transform an unpatentable principle into a patentable process."¹²⁶ The seeming triviality of this conclusion likely explains why the Court made no explicit effort to reconcile it with a simultaneous assertion that subject-matter inquiries should precede and be separate from inquiries into novelty and nonobviousness.¹²⁷ In this context, reconciliation of the two must have seemed straightforward: subject-matter inquiries should precede and be separate from inquiries into novelty and nonobviousness *in all substantial details*.

The Court's no-preemption principle might similarly be understood as a limited safety valve that protects against extreme forms of strategic claim drafting. The Court felt that grounds for enforcing the principle arose in *Gottschalk v. Benson*.¹²⁸ In the Court's view, this case's patent claims for "a method for converting binary-coded decimal . . . numerals into pure binary numerals . . . purported to cover any use of the claimed method in a general-

123. 437 U.S. 584 (1978).

124. *Id.* at 596–97 (internal quotation marks omitted).

125. The Commissioner of Patents and Trademarks contended that "[a]utomatic adjustment of alarm limits according to a programmed definition is conventional" and, more generally, that "[i]t is not disputed that, aside from the novel algorithm for computing updated alarm-limit values, every aspect of respondent's claimed invention is conventional and well-known in the art." Brief for Petitioner, *Parker v. Flook*, 437 U.S. 584 (1978) (No. 77-642), 1978 WL 206636, at *7; *see also* Reply Brief for Petitioner, *Parker v. Flook*, 437 U.S. 584 (No. 77-642), 1978 WL 206638, at *5 ("[R]espondent claims a mathematical formula, the solution of which is conventionally applied to conventional technology."). Flook's brief responded that "the applicant never conceded anything with respect to novelty," but provided no basis for believing that the final step of adjusting the alarm limit contributed in any way to the claimed invention's novelty or nonobviousness. Brief for the Respondent, *Parker v. Flook*, 437 U.S. 584 (No. 77-642), 1978 WL 206637, at *10 n.11.

126. *Diamond v. Diehr*, 450 U.S. 175, 191–92 (1981) (paraphrasing language in *Flook*, 437 U.S. at 590).

127. *See Flook*, 437 U.S. at 593 ("The obligation to determine what type of discovery is sought to be patented must precede the determination of whether that discovery is, in fact, new or obvious.").

128. 409 U.S. 63 (1972).

purpose digital computer of any type.”¹²⁹ Further, because “[t]he mathematical formula . . . ha[d] no substantial practical application except in connection with a digital computer . . . , the patent would wholly pre-empt the mathematical formula and in practical effect would be a patent on the algorithm itself.”¹³⁰ Thus, the Court applied *Benson*’s no-preemption principle in what it viewed as a very extreme set of circumstances—circumstances likely to implicate concerns about applicants’ capacity to undermine the substance of a subject-matter limitation by strategically adding trivial limitations to otherwise formally ineligible claims.¹³¹

In sum, neither the Court’s significance filter nor its no-preemption rule alters the conclusion that the question of subject-matter eligibility is fundamentally coarse-grained and categorical. The existence of safety valves need not overcome our sense of this doctrinal area’s normal operating conditions.

3. *The Value of a Coarse-Grained Filter.*—The relatively coarse-grained nature of the requirement for subject-matter eligibility makes it straightforward to attack. Unlike typically more fine-grained inquiries into utility, novelty, and nonobviousness, such a subject-matter inquiry does not generally involve an individualized assessment of a form of value. Why, then, should a pro-patentability result from patent law’s other, more fundamentally particularized inquiries ever be trumped by a determination that an innovation does not fall within a permitted category of subject matter? Why not follow the suggestion of various commentators and offer patent rights for essentially all useful, novel, and nonobvious innovations regardless of type?¹³²

129. *Id.* at 64.

130. *Id.* at 71–72; see also *Diehr*, 45 U.S. at 185–86 (“The sole practical application of the algorithm [in *Benson*] was in connection with the programming of a general purpose digital computer.”).

131. The USPTO’s *Manual of Patent Examining Procedure* provides two examples of situations in which application of the no-preemption principle would supposedly be appropriate—“a claim that recites a computer that solely calculates a mathematical formula (see *Benson*) or a computer disk that solely stores a mathematical formula,” MPEP, *supra* note 28, § 2106, at 2100–13—but these examples do not seem necessarily to involve a problem of preemption at all. In any event, other than a formalist’s sense that a preemptive claim seems equivalent to patenting forbidden subject matter itself, the rationale for the no-preemption principle seems questionable. Why should a patent applicant be denied the ability to patent a practical application of a principle simply on the grounds that nobody else has yet developed an alternative practical application of that principle? What if the applicant or another party develops an alternative application? In light of the vexing nature of such questions, *Benson* might be best reinterpreted as an early application of a *Flook*-like significance filter. See *Benson*, 409 U.S. at 67 (“The mathematical procedures can be carried out in existing computers long in use, no new machinery being necessary. And, as noted, they can also be performed without a computer.”).

132. See John F. Duffy, *Rules and Standards on the Forefront of Patentability*, 51 WM. & MARY L. REV. 609, 623 (2009) (“[I]t should be a rare situation in which an entire class of patents complies with the nonobviousness requirement and yet still somehow discourages or impedes the development and spread of useful knowledge.”); Kristen Osenga, *Ants, Elephant Guns, and*

The basic answer proffered here is that even more fundamentally particularized inquiries are no more than crude proxies for the question of whether any individual patent grant will further overall social goals.¹³³ Both practically speaking and as a matter of virtually any plausible theory, those inquiries are imperfect filters for “good” patent grants. Consequently, an additional imperfect filter—a requirement that patent claims be directed to a permissible category of subject matter—can improve patent-system performance even if that filter is relatively coarse grained.¹³⁴ To demonstrate this point, the following discussion makes the conventional assumption that the aim of U.S. patent law is to optimize social welfare to the extent possible by promoting innovation,¹³⁵ and then shows how a subject-matter eligibility requirement can add value.

A preliminary point is that patents produce social costs as well as social benefits.¹³⁶ From a utilitarian perspective, patentability requirements for

Statutory Subject Matter, 39 ARIZ. ST. L.J. 1087, 1092 (2007) (“[T]he question of subject-matter eligibility for any invention is essentially pro forma, and whether a patent is granted . . . should be based on the application meeting the requirements of patentability provided by 35 U.S.C. §§ 102, 103, and 112.”); Michael Risch, *Everything Is Patentable*, 75 TENN. L. REV. 591, 658 (2008) (“[T]he PTO and courts should focus on answering specific questions about how to best apply rigorous standards of novelty, nonobviousness, utility, and specification with a scalpel rather than simply eliminating broad swaths of innovation with a machete.”); Aaron J. Zakem, Note, *Rethinking Patentable Subject Matter: Are Statutory Categories Useful?*, 30 CARDOZO L. REV. 2983, 2988 (2009) (describing Federal Circuit efforts to restrict subject-matter eligibility as “distracting the court from analyzing other requirements for patentability such as utility, novelty, non-obviousness, and enablement”).

133. Cf. Tun-Jen Chiang, *The Rules and Standards of Patentable Subject Matter*, 2010 WIS. L. REV. 1353, 1407 (“[A]ll of these [patentability] doctrines are proxies for the same ‘ultimate’ question . . . : whether granting a patent . . . would be socially beneficial in terms of providing more incentive benefits than monopoly costs.”). Defense of a coarse-grained subject-matter filter could alternatively rely on, for example, the potentially greater ease of enforcing subject-matter limitations as opposed to patentability requirements involving more fact-intensive inquiries Samuelson & Schultz, *supra* note 97 (manuscript at 22–24) (arguing that subject-matter inquiries can increase the efficiency of USPTO and court proceedings); see also John R. Thomas, *The Responsibility of the Rulemaker: Comparative Approaches to Patent Administration Reform*, 17 BERKELEY TECH. L.J. 727, 734 (2002) (“Until recently, subject matter limitations provided a less time-consuming mechanism [than novelty and nonobviousness requirements] for the USPTO to reject applications.”), or possible benefits from freeing resources for greater scrutiny of patent claims not excluded by the subject-matter filter, Chiang, *supra*, at 1403 (“[L]ayers of screening make sense in order to reserve detailed scrutiny for a subset of cases.”).

134. Cf. Cotter, *supra* note 96, at 374 (contending that “some degree of redundancy or doctrinal overlap may actually be a virtue”).

135. See Golden, *supra* note 16, at 509 (discussing various potential goals for patent rights, but “generally assum[ing] a utilitarian goal that is standard in modern accounts”). In theory at least, calculation of social welfare could reflect concern for “freedom of speech, personal autonomy, and other liberty interests” that an overly robust patent system might infringe. Thomas F. Cotter, *A Burkean Perspective on Patent Eligibility*, 22 BERKELEY TECH. L.J. 855, 895 (2007); see also John R. Thomas, *Liberty and Property in Patent Law*, 39 HOUS. L. REV. 569, 570 (2002) (“[T]he potential impact of the patent law upon personal liberties is becoming more apparent and more worthy of concern.”).

136. See, e.g., Golden, *supra* note 16, at 517–18 (describing static-efficiency and dynamic-efficiency costs of patent rights, as well as “administrative, compliance, and enforcement costs”).

utility, novelty, nonobviousness, and adequate disclosure of the claimed invention are imperfect filters because their satisfaction does not ensure that the social benefits from issuing a particular set of patent rights exceeds the associated social costs. At best, satisfaction of these requirements indicates that a set of potential patent rights has many of the hallmarks of a set of patent rights that has a substantial likelihood of being socially desirable overall. Relative to an inquiry going directly to the question of whether a particular grant of patent rights promotes or harms social welfare, these facially individualized inquiries ultimately only serve to classify a claimed invention as a member or nonmember of a particular “type”—the class of innovations and associated descriptions that satisfy patent law’s requirements of utility, novelty, nonobviousness, and adequate disclosure.

Moreover, in a world of limited resources and capabilities, perfect administration of patentability requirements is itself no more than a dream. Inescapably, courts and the USPTO will reject some percentage of patent claims that an infallible administrator would have granted, and will allow some percentage of patent claims that an infallible administrator would have rejected.¹³⁷ Because there is no reason to believe that these administrative errors will exactly counterbalance U.S. patent law’s doctrinal imperfections,¹³⁸ the result is a system that (1) rejects claims for patent rights in some percentage, e_1 , of situations in which issue and legal enforceability would advance overall social welfare—Type I errors or false negatives;¹³⁹ and (2) grants and fails to invalidate claims for patent rights in some percentage, e_2 , of situations in which the issue and legal enforceability of patent rights is contrary to overall social welfare—Type II errors or false positives.¹⁴⁰

What do these errors mean for overall social welfare? Although somewhat crude, an additive model of patent benefits and costs helps provide a response.

Let us assume the following: (1) the average social cost of a “bad patent claim,” a patent claim whose issue and enforceability is contrary to overall

137. See FED. TRADE COMM’N, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY 38 (2003) (“All legal regimes should consider the extent to which they are subject to error—that is, false negatives and false positives. . . . In the patent context, this translates into denying a patent that should have been granted versus granting an unwarranted patent.”).

138. Administrative errors might instead compound doctrinal errors if patentability requirements have been set too low relative to the ultimate social goal, USPTO examiners have inadequate incentives to enforce patentability requirements as they exist, and juries are too likely to defer to the USPTO’s decision to issue a patent.

139. See RONALD E. WALPOLE & RAYMOND H. MYERS, PROBABILITY AND STATISTICS FOR ENGINEERS AND SCIENTISTS 290 (4th ed. 1989) (“Rejection of the null hypothesis when it is true is called a type I error.” (emphasis omitted)).

140. See *id.* (“Acceptance of the null hypothesis when it is false is called a type II error.” (emphasis omitted)).

social welfare, is C ; (2) the average social benefit from a “good patent claim,” a patent claim whose issue and enforceability promotes overall social welfare, is B ; (3) the average social cost of a failed effort to secure a good patent claim—a variant of rent-seeking cost—is R_1 ; (4) the average social cost of a failed effort to obtain a bad patent claim is R_2 ; (5) the number of applications for good patent claims is n_1 ; and (6) the number of applications for bad patent claims is n_2 . Then, the patent system’s contribution—or detriment—to social welfare is given by the formula:

$$(1 - e_1)n_1B - e_2n_2C - e_1n_1R_1 - (1 - e_2)n_2R_2 \quad (\text{Eq. 1}).^{141}$$

Further, because the last two terms in this formula are negative, we can recognize that, under this model, a threshold condition for the patent system’s success in promoting social welfare is given by the following, simpler inequality:

$$(1 - e_1)n_1B - e_2n_2C > 0 \quad (\text{Eq. 2}).$$

In words, a threshold condition for the patent system’s increasing social welfare under this model is the intuitive requirement that the net value added by good patent claims exceed the net cost from bad patent claims.¹⁴²

Now, consider what happens if we restrict our attention to innovations of a particular type T . After indicating this focus by adding the subscript T to the relevant values for e_i , n_i , B , and C , we obtain the following threshold condition for the social desirability of extending patents to innovations of type T :

$$(1 - e_{1,T})n_{1,T}B_T - e_{2,T}n_{2,T}C_T > 0 \quad (\text{Eq. 3}).$$

If the values of $e_{i,T}$, $n_{i,T}$, B_T , and C_T differ across innovation types T , then the inequality can be satisfied for some types of innovations but not others. If, for example, the patent system has difficulty weeding out bad patent claims of type T , and if bad patent claims of that type are particularly likely to be privately useful for purposes of “extorting” large licensing fees or mistaken court awards, the quantities $e_{2,T}$, $n_{2,T}$, and C_T might all be especially large for that type of innovation, with the result being that the inequality is not satisfied even if $e_{1,T}$, $n_{1,T}$, and B_T all have values comparable to those for types of innovation that the patent system serves well. Such might be the case for significant sectors of information technology, for example, if one

141. For simplicity of later reference, mathematical expressions such as formulas, inequalities, and equations, as well as actual equations, have a numerical “equation designation” (e.g., “Eq. 1”) associated with them, even though, for example, an *inequality* is not technically an *equation*.

142. The patent system might, as a practical matter, also be subject to a political constraint requiring that the ratio of issued “bad patent claims” to issued “good patent claims”—i.e., the ratio $n_2e_2/n_1(1 - e_1)$ —fall below some critical value. Otherwise, loss of confidence in the patent system might force the system’s abandonment. In any event, replacement of the regular text’s cost-benefit constraint with this relative-frequency constraint does not alter the ultimate conclusion that, if parameters such as e_1 , e_2 , n_1 , and n_2 have values that differ for different types of subject matter, certain types of subject matter might fail to satisfy even minimal requirements for patents to be a desirable means for promoting social welfare.

credits accounts suggesting that the complexity of many modern software products generates an abnormally high risk of “extortive” settlements or mistakenly high court awards.¹⁴³ In any event, the point of the inequality in Equation 3 is that a coarse-grained subject-matter filter can be used to exclude patents on types of innovations for which patenting is not socially beneficial.¹⁴⁴

Of course, one response to such complaints is that, to the extent patenting does not work for certain types of innovation, the patent system simply needs to be reformed. Hence, a reformer might seek to raise the nonobviousness bar and to dilute remedies for patent infringement in hopes that such changes will enable the USPTO and courts to work together to decrease socially erroneous patent grants and their cost (decreasing $e_{2,T}$ and C_T). Patent applicants might further enhance the positive effects of such reforms by reducing applications for bad patent claims (decreasing $n_{2,T}$) in accordance with the reduced prospects for profit from such claims. But these reforms might also increase erroneous patent-claim rejections (increasing $e_{1,T}$), diminish the private value and social benefits from good patent claims (decreasing B_T), and thus reduce applications for good patent claims (decreasing $n_{1,T}$). At least in the abstract, therefore, the reforms could make matters worse rather than better.

Moreover, this ambiguity becomes deeper once one recognizes that there will be substantial pressure for such reforms to be applied system wide, rather than to specific innovation types only. The international TRIPS agreement, to which the United States and more than a hundred other nations are bound, includes a nondiscrimination principle mandating that signatories make patent rights “available and . . . enjoyable without discrimination as to the place of invention, the field of technology and whether products are imported or locally produced.”¹⁴⁵ TRIPS’s nondiscrimination principle provides positive-law reinforcement for a common, albeit nonuniversal, preference for keeping the basic form of patent law constant even as applied to widely divergent technologies.¹⁴⁶ If this unitarian thrust forces reforms to be applied system wide, a reform to make the patent system work better for one

143. See, e.g., Mark A. Lemley & Carl Shapiro, *Patent Holdup and Royalty Stacking*, 85 TEXAS L. REV. 991, 992 (2007) (“In the information technology sector in particular, modern products such as microprocessors, cell phones, or memory devices can easily be covered by dozens or even hundreds of different patents.”).

144. See Chiang, *supra* note 133 (observing that a “benefit of [35 U.S.C. § 101] categorical rejections is that [they] can at least be directed to categories that, on the whole, are socially detrimental” for purposes of patenting).

145. Agreement on Trade-Related Aspects of Intellectual Property Rights, Including Trade in Counterfeit Goods art. 27, Apr. 15, 1994, 33 I.L.M. 1197 [hereinafter TRIPS].

146. Cf. R. Polk Wagner, *Of Patents and Path Dependency: A Comment on Burk and Lemley*, 18 BERKELEY TECH. L.J. 1341, 1360 (2003) (concluding that “there seem to be strong reasons to conclude that we should . . . work diligently to reduce the very technological-specificity that [an article by Dan Burk and Mark Lemley] advocates”).

type of innovation might decrease system effectiveness with respect to other innovation types.

Patent-law unitarianism makes the importance of properly delineating patent law's subject-matter boundaries all the more crucial. To the extent different types of innovation respond differently to the patent system's incentives, enforcing limits on patentable types of innovation—in TRIPS lingo, adopting a less than all-inclusive definition of “technology” against which one cannot discriminate¹⁴⁷—becomes vital to the system's overall social desirability.

For an example of a type of innovation that a patent system might not serve well, consider the sorts of very general scientific or mathematical ideas, such as Newton's law of universal gravitation¹⁴⁸ or Einstein's energy–mass equation,¹⁴⁹ that the Supreme Court has emphatically declared not to be patentable subject matter.¹⁵⁰ Commentators have suggested that this exclusion makes sense on grounds that the distance between such general ideas and specific practical applications means that patent rights directed to such ideas are particularly likely to be difficult to define and to enforce,¹⁵¹ might impede development or exploitation of a wide array of uses,¹⁵² will decrease incentives and increase costs for follow-on innovations that develop the ultimately desired practical applications,¹⁵³ and will not provide a

147. TRIPS, *supra* note 145.

148. See DOUGLAS C. GIANCOLI, PHYSICS: PRINCIPLES WITH APPLICATIONS 72 (1980) (“Every body in the universe attracts every other body with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between them. This force acts along the line joining the two bodies.”).

149. See *id.* at 655 (discussing “Einstein's equation $E = mc^2$ ”).

150. See *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980) (“Einstein could not patent his celebrated law that $E=mc^2$; nor could Newton have patented the law of gravity.”); see also *Bilski v. Kappos*, 130 S. Ct. 3218, 3225 (2010) (“The Court's precedents provide three specific exceptions to § 101's broad patent-eligibility principles: ‘laws of nature, physical phenomena, and abstract ideas.’” (quoting *Chakrabarty*, 447 U.S. at 309)).

151. See SHAVELL, *supra* note 77, at 165 (concluding that “the definition and enforcement of property rights in basic research results [is] impractical”); see also WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 305–06 (2003) (contending that patent rights on “fundamental scientific (including mathematical) and technological principles” would generate huge transaction costs “because the scope of [the] idea . . . often is extremely difficult to pin down”).

152. Cf. SHAVELL, *supra* note 77, at 151 (citing, as a potential justification for refusing patents on “abstract ideas,” “the great loss that society would suffer were very general and useful ideas to be restricted by giving a party a monopoly in them”); Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 882 (1990) (“Particularly when the technology is in its early stages, the grant of a broad-gauged pioneer patent . . . may preclude other inventors from making use of their inventions . . .”).

153. See MERGES & DUFFY, *supra* note 50, at 89 (“Granting a broad initial patent increases the incentives to the pioneer, and decreases them for the improver.”).

substantial and reliable basis for effective funding of research likely to lead to the discovery of such general ideas.¹⁵⁴

These intuitions can be given a mathematical formulation as follows. Let us suppose that an initial innovative idea can be characterized as having some number, N , of distinct kinds of practical uses, where a “kind of practical use” refers to a distinctive form of technological embodiment. For example, if the initial innovative idea is a steam engine, different kinds of practical uses might include its use to pump water from a mine, its use to propel a vehicle such as a locomotive, or its use to provide power to manufacturing machinery such as a loom. A basic scientific principle, such as Newton’s second law of motion, $F = ma$,¹⁵⁵ or the law of conservation of energy¹⁵⁶ might have significant use in the development and operation of an even wider variety of technological embodiments. Such a general principle might be particularly susceptible to fruitful combination with strains of knowledge and technology completely unfamiliar to the principle’s original developer. Hence, the general principle might be likely to be associated with a very large value of N . The number N of kinds of practical uses can thus be viewed as a technological measure of an initial idea’s generality of application.

Three other parameters necessary for the desired mathematical formulation relate to the effect that patent rights are likely to have on the development of such practical uses. I call these patent rights’ acceleration value α , coordination value β , and drag value λ .

According to conventional incentive theories for patent rights, the availability of patent rights for the idea should accelerate the time of the idea’s development and public disclosure.¹⁵⁷ Moreover, to the extent distinct practical uses of the initial idea are separately patentable, patent law might accelerate their follow-on development and disclosure as well. Presuming that applications of the idea have an overall benign effect on social welfare, patent-law-induced acceleration of such development and disclosure should generate an increase in social welfare. Call the average increase in social value per kind of use of the initial idea the “acceleration value,” α , of patent rights.

In accordance with the “prospect theory” famously expounded by Edmund Kitch, a patent might also add social value by giving the patentee a special ability or incentive to coordinate or otherwise to promote the idea’s

154. See LANDES & POSNER, *supra* note 151, at 307 (“[A] great deal of [basic] research has no near-term commercial application and so could not be financed by patenting.”).

155. See GIANCOLI, *supra* note 148, at 55 (discussing Newton’s second law of motion).

156. See *id.* at 133 (describing the law of conservation of energy as “one of the most important principles in physics”).

157. See, e.g., Golden, *supra* note 16, at 520–21 (discussing ways in which patents can stimulate innovation and public disclosure).

application or further refinement through follow-on innovations.¹⁵⁸ Call the resulting average value added per kind of use the “coordination value,” β , of patent rights.

Finally, consider that a patent also generates social costs—for example, deadweight loss associated with limitations on consumption, and dynamic-efficiency losses from impediments to, or taxes on, follow-on innovation.¹⁵⁹ Call the average social cost per kind of use the “drag value,” λ , of patent rights.

The net effect on social welfare from issuing patent rights for the idea is then given by the formula:

$$\Delta W = \alpha N + \beta N - \lambda N \quad (\text{Eq. 4}).$$

Some additional points can be made regarding Equation 4. First, the average patent-right values α , β , and λ are probably not independent of N . The acceleration and coordination values α and β are measures of average, per-kind-of-use benefits that likely reflect a law of diminishing returns: they likely decrease in size as N becomes very large. For both α and β , such decrease could result, for example, because there are only a finite number of truly high-value kinds of uses. Alternatively, some kinds of uses might be high-value but might require complementary developments or, at least, significant amounts of time before they can be appreciated or developed.¹⁶⁰ If such complementary developments or gestation times are not themselves accelerated by the existence of the initial idea, this latter set of uses might not benefit as much as others—if at all—from patent law’s acceleration of the building-block idea’s development and disclosure. Likewise, the initial innovator might not benefit as much from such uses because the innovator will likely either have to share value with later innovators or might be unable, due to patent expiration, to demand a sharing of the value at all. The natural result is a value for α that decreases as N grows to encompass more and more uses that benefit less and less from the initial acceleration. Somewhat similarly, an individual patentee’s limited capacity for coordinating or pursuing all potential uses of a basic idea suggests the average coordination value β should decrease as N increases, with the individual patentee’s capacity becoming more and more strained. As a consequence of such reasons to expect diminishing returns with increasing N , the acceleration value α and coordination value β will be illustratively modeled as decreasing functions of

158. See Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265, 266 (1977) (contending that the patent system “tends to assure efficient allocation of . . . resources among” technological opportunities “by awarding exclusive and publicly recorded ownership of a prospect shortly after its discovery”).

159. See, e.g., Golden, *supra* note 16, at 517 (discussing static- and dynamic-efficiency costs of patent rights).

160. See Kitch, *supra* note 158, at 271–72 & tbl.1 (using a table of inventions to illustrate the fact that “the first patentable invention frequently occurs years before the first significant commercial product”).

N : $\alpha(N) = \alpha_0 N^\gamma$ and $\beta(N) = \beta_0 N^\varepsilon$, where α_0 , γ , β_0 , and ε are all positive real numbers.

By comparison, the expected effect of increasing N on the average drag cost, λ , seems substantially ambiguous. For increasing values of N , an increasing percentage of kinds of uses of the original idea might not be realized, or at least not commercially realized, until long after any patent on the original idea has expired. Such delayed commercial uses might experience little drag from patent rights. The result could be a drag value λ that is a decreasing function of N . On the other hand, the average drag cost, λ , could, instead, increase with N as an individual patentee finds personally developing or licensing all prospective kinds of uses increasingly difficult. Such increased difficulty with increasing N can result from simple limitations on the patentee's capacity to spend time learning about potential uses of the initial idea, from increased difficulty in acquiring adequate knowledge about technologies more remote from the patentee's core areas of expertise, or from increased difficulty in developing relationships of understanding and trust with technological communities more and more remote from the patentee's core areas of expertise. Alternatively, λ might tend to grow with N because larger values of N might correlate with a higher degree of difficulty in designing a technology that competes with applications of the initial idea but does not infringe a patent on that idea. By leaving little if any space for such competing "design-arounds," a patent on a very general idea might be associated with more severe deadweight losses and greater impediments to follow-on innovation.

Because the above-described effects have potentially opposite implications for λ 's N -dependence, we will, for illustrative purposes, treat the drag value, λ , as constant, taking $\lambda = \lambda_0$ in the analysis that follows. The implications of that analysis will only be strengthened if the drag value λ in fact grows with the value of N .

Plugging the above-described approximate forms for α , β , and λ into Equation 4, we obtain:

$$\Delta W = \alpha_0 N^{1-\gamma} + \beta_0 N^{1-\varepsilon} - \lambda_0 N \quad (\text{Eq. 5}).$$

Figure 1 provides a plot of ΔW as a function of N for certain assumed values of α_0 , β_0 , λ_0 , γ , and ε . The figure illustrates a key point. This is that, as N becomes very large, the two positive terms on the right-hand side of Equation 5, $\alpha_0 N^{1-\gamma}$ and $\beta_0 N^{1-\varepsilon}$, ultimately become small relative to the negative term, $-\lambda_0 N$. Thus, Equation 5 yields a result comporting with the common intuition that very generally applicable scientific or mathematical ideas should not be patentable.¹⁶¹ In the model's terms, such ideas tend to be

161. On the other hand, Equation 5 can also be understood to support the notion that the patent system might be particularly effective in improving social welfare to the extent it spurs development of pioneering inventions that have a wide range of practical applications (albeit a range of practical

characterized by a high number of potential uses and, across this large spectrum of uses, relatively low average acceleration and coordination values. As a result, the negative drag effects from patents on such ideas can be expected to outweigh their positive benefits.

Figure 1

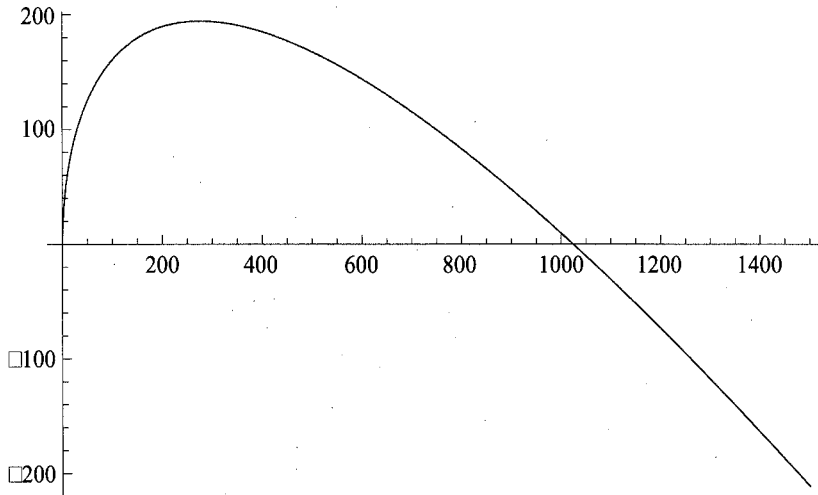


Fig. 1: This is a plot of the expected patentability-induced social-welfare increment ΔW as a function of N , the number of kinds of uses for an innovation, where the Greek-lettered parameters in Equation 5 have the following values: $\alpha_0 = 4$, $\beta_0 = 16$, $\lambda_0 = 1$, $\gamma = 0.3$, $\varepsilon = 0.5$.

More generally, the above argument provides a plausible account of how truly fundamental innovations—innovations associated with an unusually high number N of potential uses—can be socially inappropriate for patent protection even when specific, real-world utilities for them are known

applications not so large that patents' negative drag effects dominate). See A. Samuel Oddi, *Un-Unified Economic Theories of Patents—The Not-Quite-Holy Grail*, 71 NOTRE DAME L. REV. 267, 285 (1996) (speaking of “elegant (basic/revolutionary/pioneer) inventions generally considered the most valuable to society”). As Figure 1 indicates, the peak value of ΔW as a function of N might occur at a relatively large value of N . Consequently, Equation 5 captures some of the tension implicit in the common feeling that the patent system is meant to reward significant innovations, see MERGES & DUFFY, *supra* note 50, at 615 (“[G]ranting obvious patents may create a proliferation of economically insignificant patents that are expensive to search and license.”), but not necessarily to provide rights in developments so fundamental that they are properly viewed as “part of the storehouse of knowledge of all men,” *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 130 (1948). The fact that ΔW rises with increasing N , reaching a peak at a potentially quite substantial value of N , might be understood to correspond to the notion that patents for seminal—i.e., relatively high- N —inventions can be among the most socially valuable patents that a government issues.

and when they also satisfy patent law's requirements of novelty, nonobviousness, and adequate disclosure. Because fundamental "laws of nature" like Newton's second law of motion, $F = ma$,¹⁶² are a form of "abstract idea" that tends to have an extremely high number of different kinds of uses, the cost, C_T , of improperly granting patents to this type of innovation could be extraordinarily large. Further, the relative toothlessness of other patentability requirements in preventing socially undesirable grants for this type of innovation—because, for example, it is clearly novel and nonobvious—is likely to correspond to a high Type II-error rate, $e_{2,T}$, associated with the granting and lack of invalidation of socially undesirable patent rights. Moreover, when alternative means to motivate the discovery of such fundamental principles are considered, there might be only a negligible set of circumstances in which a grant of a patent on such an idea is expected to be beneficial. Meanwhile, allowance for patent protection plus the potentially enormous windfall from owning patent rights for such a fundamental principle suggest that the number, $n_{2,T}$, of applications for socially undesirable patent claims for this sort of innovation will be large. In short, this second mathematical model for high- N innovations corroborates the intuition from Equation 3 that patent rights might well be substantially desirable for certain types of innovations while almost wholly undesirable for others.

B. The Choice of Institutional Actor

Section III(A)(3) provided justification for retaining a patentability requirement that seeks to exclude from patent protection whole categories of innovation. This subpart considers the remaining institutional question: What institutional actor should play the primary, day-to-day role in determining the limits of the patent system's covered categories?

This subpart contends that, given the coarse-grained nature of the subject-matter inquiry, the USPTO seems the best place to locate primary responsibility for the inquiry's scope and structure. The courts appear inherently ill-suited to the task, and their long-standing struggles with subject-matter inquiries corroborate this appearance. Likewise, Congress appears unable to give this area the attention it needs, and is subject to barer interest-group pressures and the demands of alternative institutional responsibilities that likely make socially undesirable its sustained involvement in the provision of detailed and diligently updated patent-law guidance. Meanwhile, executive actors other than the USPTO lack any pressing need to maintain close familiarity or concern with issues of claim drafting, patent scope, and application processing that are likely crucial to sensible regulation of patentable subject matter. Although the USPTO is an imperfect locus for

162. See *supra* note 155.

and invention classification in the process of examining patent applications. Through both regularly updated guidelines and the individual work of thousands of examiners, the USPTO addresses concerns of claim breadth and form on a daily basis. A different agency supposedly better insulated from patent-community “capture” would likely suffer more than counterbalancing deficiencies of understanding, urgency, and effective supervisory authority.

1. *Courts’ Struggles with Defining Patentable Subject Matter.*—One reason to consider giving the USPTO primary interpretive authority for questions of patentable subject matter is that courts’ performance in forming and applying tests for patentable subject matter has been fairly poor.¹⁶³ Decades-long difficulties in handling computer-related innovation highlight the courts’ problems.¹⁶⁴ Further, the persistence of these difficulties suggests that the courts’ problems are fundamentally institutional, rather than chance by-products of an unfortunate constellation of individual judges or judgments. The courts appear to face at least four institutional difficulties in defining patentable subject matter: (1) justified discomfort with addressing what are in effect broad questions of policy; (2) systematic difficulty, given Article III constraints and layers of appellate review, in addressing subject-matter eligibility questions in a timely and definitive manner; (3) relative inflexibility in dealing with the disruption costs that a substantial change of course on subject-matter eligibility can create; and (4) stare decisis-mediated attachment to a substantially incoherent body of precedent that could benefit from a systematic overhaul. The general and perhaps inevitable result of such institutional problems is a case law that tilts—sometimes alternately—toward either of two suboptimal approaches: one in which there are no effective limits on subject-matter eligibility and another in which there are limits but only ones of highly uncertain, inarticulate, and even variable scope.

Chronicles of U.S. courts’ efforts to regulate subject-matter eligibility commonly start at least as early as the U.S. Supreme Court’s 1854 decision of *O’Reilly v. Morse*.¹⁶⁵ But no doubt in part because pre-1952 U.S. Patent Acts did not establish sharp distinctions between patentability questions of subject-matter eligibility, novelty, utility, and nonobviousness, judicial opinions that predate the Patent Act of 1952 tend to mix concerns of subject-

163. See Duffy, *supra* note 132, at 623 (characterizing the history of efforts to limit patentable subject matter as a “march of failures”).

164. See Donald S. Chisum, Patenting Intangible Methods: Revisiting *Benson* (1972) After *Bilski* (2010) 1 (Oct. 27, 2010) (unpublished manuscript), available at <http://ssrn.com/abstract=1698724> (“The persistence of controversy over the patentability of software for 38 years and the inability of a majority in the Supreme Court to provide significant guidance on the patentability of intangible methods verify the weakness of *Benson*’s reasoning.”).

165. See 56 U.S. (15 How.) 62, 112–20 (1854) (limiting the extent to which an idea can be patented and forbidding the inventor from patenting improvements to the idea reached through alternative means).

matter eligibility, claim overbreadth, claim indefiniteness, or lack of cognizable “invention” in ways that obscure those cases’ present meaning.¹⁶⁶

The first major post-1952 effort by the Supreme Court to clarify limits on patentable subject matter came decades after the 1952 Act’s passage, in a series of decisions stretching from 1972 to 1981.¹⁶⁷ Almost perforce, the results of such Supreme Court review were legally significant. But with respect to performance in explicating background policy or doctrine, they were underwhelming. The Court’s decisions revealed it to be fully capable of opening the floodgates to patenting biotechnological inventions¹⁶⁸ but—as detailed below—substantially inarticulate and incoherent when attempting to delineate limits to the patentability of software-related inventions.

The Supreme Court’s recent involvement with questions of patentable subject matter has generated a similar impression. From 2001 to 2010, the Supreme Court has heard three oral arguments on such questions, has issued two precedential opinions, and has dismissed one of the petitions for which it heard oral argument as improvidently granted:

- In *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred International, Inc.*,¹⁶⁹ the Court confirmed something many, including the USPTO,¹⁷⁰ had long suspected—namely, that one may obtain a valid utility patent for a new form of plant despite the U.S. Code’s separate provision for plant patents and plant-variety protection.¹⁷¹

166. Cf. MERGES & DUFFY, *supra* note 50, at 87 (discussing how, after *Morse*, patent claims have commonly been “declared void not for being too broad in the abstract, but rather for being broader than the invention disclosed in the patent specification” (emphasis omitted)); Parasidis, *supra* note 98, at 348 (“Early Supreme Court opinions excluded products of nature from patent-eligible subject matter primarily through use of the novelty and disclosure requirements.”).

167. See *Diamond v. Diehr*, 450 U.S. 175, 185–87 (1981) (“This Court has undoubtedly recognized limits to § 101 and every discovery is not embraced within the statutory terms.”); *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980) (“This is not to suggest that § 101 has no limits or that it embraces every discovery.”); *Parker v. Flook*, 437 U.S. 584, 598 (1978) (Stewart, J., dissenting) (“It is a commonplace that laws of nature, physical phenomena, and abstract ideas are not patentable subject matter.”); *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972) (“Phenomena of nature, though just discovered, mental processes, and abstract intellectual concepts are not patentable, as they are the basic tools of scientific and technological work.”).

168. *Chakrabarty*, 447 U.S. at 318 (declaring that Congress could “amend § 101 so as to exclude from patent protection organisms produced by genetic engineering,” but that, absent amendment, “[t]he language of that section fairly embraces” such subject matter).

169. 534 U.S. 124 (2001).

170. See *id.* at 144–45 (“We also note that the PTO has assigned utility patents for plants for at least 16 years . . .”).

171. See *id.* at 145 (“[W]e hold that newly developed plant breeds fall within the terms of § 101, and that neither the [Plant Patent Act] nor the [Plant Variety Protection Act] limits the scope of § 101’s coverage.”); see also *id.* at 147 (Scalia, J., concurring) (finding crucial the fact that, in light of *Chakrabarty*, 447 U.S. 303, “[s]tare decisis . . . prevents us from any longer regarding as an open question . . . whether ‘composition of matter’ includes living things”).

- In *Laboratory Corp. of America Holdings v. Metabolite Laboratories, Inc.*,¹⁷² the Court dismissed the case after oral argument.¹⁷³ Nonetheless, in a dissent from the dismissal that apparently spurred further judicial activity in this area,¹⁷⁴ three Justices joined a dissent that not only opposed the dismissal, but also contended that the Court should declare unpatentable a diagnostic-method claim that, in the three Justices' view, consisted only of "instruct[ions] . . . to (1) obtain test results and (2) think about them."¹⁷⁵
- In *Bilski v. Kappos*, the Court held that a claim for a method of hedging commodity risk lacked patentable subject matter because it constituted "an unpatentable abstract idea."¹⁷⁶ But the Court did little to clarify the meaning of the term *abstract idea*.¹⁷⁷ Instead, the Court largely contented itself with adding an essentially conclusory statement of its holding to a brief recitation of "guideposts" from its prior string of post-1952 decisions.¹⁷⁸

The decades-old guideposts cited in *Bilski* are essentially proven failures. Indeed, their inability to provide the basis for a coherent and stable jurisprudence appears largely to account for the Federal Circuit's pre-2010 efforts to generate new verbal formulations—whether the expansive "useful, concrete, and tangible result" test adopted in 1998¹⁷⁹ or the more restrictive

172. 548 U.S. 124 (2006).

173. *Id.* at 125.

174. See Golden, *supra* note 11, at 707–08 ("It seems no coincidence that, in the wake of *Laboratory Corp.*, Federal Circuit panels experimented with more vigorous enforcement of requirements for subject matter eligibility, and that the Federal Circuit subsequently used an en banc decision to reject as overly permissive a test for subject matter eligibility that the Circuit's decisions had generated in the late 1990s." (citation omitted)).

175. *Lab. Corp.*, 548 U.S. at 136 (Breyer, J., dissenting).

176. *Bilski v. Kappos*, 130 S. Ct. 3218, 3231 (2010).

177. Peter S. Menell, *Forty Years of Wandering in the Wilderness and No Closer to the Promised Land: Bilski's Superficial Textualism and the Missed Opportunity to Ground Patent Law Interpretation and Return Patent Law to Its Technology Mooring*, STANFORD L. REV. (forthcoming 2011) (manuscript at 3), available at <http://ssrn.com/abstract=1722422> ("[T]he Supreme Court's *Bilski* decision has left the patent community in the wilderness."); cf. Tony Dutra, *Michel, Federal Circuit's Bilski Author, Says High Court Review 'Too Soon, Wrong Case'*, 80 PAT., TRADEMARK & COPYRIGHT J. 323, 324 (2010) (reporting former Federal Circuit Chief Judge Paul Michel's belief that the Supreme Court's failure to provide "additional definition to 'abstractness'" left the patent community with inadequate guidance).

178. *Bilski*, 130 S. Ct. at 3231 ("The Court . . . need not define further what constitutes a patentable 'process,' beyond pointing to the definition of that term provided in § 100(b) and looking to the guideposts in *Benson*, *Flook*, and *Diehr*.").

179. *State St. Bank & Trust Co. v. Signature Fin. Grp., Inc.*, 149 F.3d 1368, 1375 (Fed. Cir. 1998) (quoting *In re Alappat*, 33 F.3d 1526, 1544 (Fed. Cir. 1994) (en banc)) (concluding that a claimed machine's "admitte[d] produc[tion of] a 'useful, concrete, and tangible result'" made it patentable subject matter), *overruled in relevant part by In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008) (en banc), *aff'd sub nom. Bilski v. Kappos*, 130 S. Ct. 3218 (2010).

“machine-or-transformation test” adopted ten years later.¹⁸⁰ Perhaps the most definitive doctrinal outcome of the Supreme Court’s decision in *Bilski* was its effective sweeping aside of such later formulations without the provision of any replacement, the result being a return to the general state of law on computer-related inventions largely as it had stood in the early 1980s.¹⁸¹

In the resulting doctrinal framework, the Supreme Court’s clearest point of guidance is the oft-repeated declaration that “laws of nature, natural phenomena, and abstract ideas” are unpatentable.¹⁸² But the Court has never given much guidance regarding the meanings of these terms.¹⁸³ This is problematic because much of what the Court seems to take as the apparent clarity of these exclusions is no more than superficial.¹⁸⁴

The excluded category of “natural phenomena” might be the least problematic. The excluded matter apparently consists substantially of things that are understood to exist independent of any conscious human action.¹⁸⁵ One might contend that there seems little need for such a subject-matter exclusion given the modern U.S. Patent Act’s separate requirements of novelty and nonderivation.¹⁸⁶ The former demands distinction from prior art, of which much natural phenomena is likely to be a part.¹⁸⁷ The latter establishes that an invention may not be patented if the claimed inventor “did not

180. See *In re Bilski*, 545 F.3d at 955 (en banc) (responding to arguments “that the Supreme Court did not intend the machine-or-transformation test to be the sole test governing [subject-matter] analyses” for process claims), *aff’d sub nom.* *Bilski v. Kappos*, 130 S. Ct. 3218 (2010).

181. See *Bilski*, 130 S. Ct. at 3231 (“[N]othing in today’s opinion should be read as endorsing interpretations of § 101 that the Court of Appeals for the Federal Circuit has used in the past.”).

182. *Diamond v. Diehr*, 450 U.S. 175, 185 (1981).

183. Parasidis, *supra* note 101, at 333 (“Notwithstanding frequent citation . . . , the concepts encompassed by the [off-cited subject matter exceptions] have not been clearly defined by the United States Supreme Court or the United States Court of Appeals for the Federal Circuit.”); *cf.* *Bilski*, 130 S. Ct. at 3236 (Stevens, J., concurring) (“The Court . . . never provides a satisfying account of what constitutes an unpatentable abstract idea. . . . This mode of analysis (or lack thereof) . . . means that the Court’s musings on this issue stand for very little.”).

184. *Cf.* Menell, *supra* note 177 (manuscript at 11–12) (describing the Supreme Court as having “collapsed the rich historical development of patentable subject matter doctrine into three amorphous, static, and ill-defined exceptions”).

185. See MERRIAM-WEBSTER’S COLLEGIATE DICTIONARY 826 (11th ed. 2006) (defining *natural* as meaning, *inter alia*, “not artificial” or “in a state of nature untouched by the influences of civilization and society”); *cf.* Ruloff F. Kip, Jr., *The Patentability of Natural Phenomena*, 20 GEO. WASH. L. REV. 371, 385 (1952) (arguing that matter should be patentable if it “has a new effect which the public either previously did not know or did not enjoy, and the new effect is the result of a positive human act operating upon natural phenomena”); Parasidis, *supra* note 98, at 370 (“Natural phenomena include any state or process in the universe that occurs or exists independent of man’s knowledge of its existence, or that arises without man’s assistance.”).

186. MERGES & DUFFY, *supra* note 50, at 358 (“The novelty requirements [of the U.S. Patent Act] are set forth in a section of the statute, 35 U.S.C. § 102, that also encompasses two other very important issues—the prohibition on derivation and the so-called ‘statutory bars’ to patenting.”).

187. See 35 U.S.C. § 102 (2006) (including various provisions requiring distinction between a claimed invention and prior art); see also *id.* § 101 (“Whoever invents or discovers any *new* and useful process, machine, manufacture, or composition of matter, or any *new* and useful improvement thereof, may obtain a patent therefor” (emphasis added)).

himself invent the subject matter sought to be patented.”¹⁸⁸ But any resulting concern with redundancy differs in kind from concerns with clarity and coherence. The natural-phenomena bar seems to score relatively well with respect to the latter even though a recent district court decision has highlighted some ambiguity about the extent to which matter must differ “markedly” from natural phenomena to be subject-matter eligible.¹⁸⁹

The exclusions of “laws of nature” and “abstract ideas,” on the other hand, cannot be taken at anything near face value. The Supreme Court has explicitly indicated that the term “laws of nature” encompasses human-developed descriptions of physical phenomena such as Newton’s “law of gravity.”¹⁹⁰ But according to modern science, Newton’s theories are not really laws of nature at all; they are no more than approximate descriptions or models, limiting forms of behavior described more precisely and accurately by modern theories of quantum mechanics and relativity.¹⁹¹ The scope of the “laws of nature” exclusion thus does not seem tied to whether the description at issue is truly accurate, and thus truly “of nature.” Instead, the real concern seems to be that the so-called laws of nature cited by the Supreme Court are “abstract ideas”—generalized descriptions untethered to any particular, practical ends.¹⁹²

Given (a) the potential redundancy of the natural-phenomenon exclusion with patent law’s novelty and nonderivation requirements, and (b) the collapsibility of the natural-laws exclusion into the abstract-ideas exclusion, the category of “abstract ideas” seems the true centerpiece of judge-made limitations on patentable subject matter. Unfortunately, the

188. *Id.* § 102(f).

189. *Ass’n for Molecular Pathology v. U.S. Patent & Trademark Office*, 702 F. Supp. 2d 181, 229 (S.D.N.Y. 2010) (“In light of DNA’s unique qualities as a physical embodiment of information, none of the structural and functional differences cited . . . between native *BRCA1/2* DNA and the isolated *BRCA1/2* DNA claimed in the patents-in-suit render the claimed DNA ‘markedly different.’”).

190. *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980).

191. *See, e.g.*, GERALD HOLTON & STEPHEN BRUSH, *INTRODUCTION TO CONCEPTS AND THEORIES IN PHYSICAL SCIENCE* 491 (2d ed. 1985) (“With this chapter [on quantum mechanics] we come to the end of Newtonian mechanics as a fundamental basis for physical theory, even though almost all of its practical applications remain valid at the present time.”). Recognition that so-called “laws of nature” are merely human theories—and always subject to falsification—might help explain why Einstein’s mass-energy formula has not generally been described as a “law,” but rather as a corollary of his theories of relativity. *See, e.g.*, Andrew T. Domondon, *Bringing Physics to Bear on the Phenomenon of Life: The Divergent Positions of Bohr, Debrück, and Schrödinger*, 37 *STUD. HIST. & PHIL. BIOLOGY & BIOMED. SCIS.* 433, 453–54 (2006) (noting Karl Popper’s “suggest[ion] that one of the defining traits of scientific activity is the process of consciously attempting to falsify propositions through experimental tests” (citing K. POPPER, *THE LOGIC OF SCIENTIFIC DISCOVERY* (1959))).

192. *See* Kevin Emerson Collins, *Propertizing Thought*, 60 *S.M.U. L. REV.* 317, 348 (2007) (observing that the prohibition against patenting “laws of nature” requires “distinguish[ing] claims to ‘laws of nature’ in the abstract from those that recite applications”).

facial simplicity of this category's name belies the difficulty in articulating its substance.

As with "laws of nature," plain meaning appears to fail. Potentially relevant ordinary meanings of the word "abstract" appear to be (1) nonspecific or (2) lacking direct connection to the material world, in either mode of operation or effect.¹⁹³ But neither of these ordinary meanings appears to match the meaning of the category "abstract ideas" as developed in actual practice.

For purposes of a subject-matter exclusion, the term *abstract* cannot simply mean that the relevant idea is nonspecific. Patent claims declared to be "abstract and sweeping" in cases such as *Gottschalk v. Benson* have included recitations of steps that, according to ordinary meaning, would be likely to be considered very specific.¹⁹⁴ For example, a claim held unpatentable in *Benson* made the following painstaking description of how to manipulate entries in a "reentrant shift register" to do a mathematical calculation:

The method of converting signals from binary coded decimal form into binary which comprises the steps of (1) storing the binary coded decimal signals in a reentrant shift register, (2) shifting the signals to the right by at least three places, until there is a binary "1" in the second position of said register, (3) masking out said binary "1" in said second position of said register, (4) adding a binary "1" to the first position of said register, (5) shifting the signals to the left by two positions, (6) adding a "1" to said first position, and (7) shifting the signals to the right by at least three positions in preparation for a succeeding binary "1" in the second position of said register.¹⁹⁵

The second suggested ordinary meaning for "abstract"—namely, lacking direct connection to the material world—fares even worse under current case law. The Supreme Court's recent decision in *Bilski* suggests that the term "abstract" cannot simply mean "disconnected from tangible matter." Otherwise, the Court would presumably have embraced something along the lines of the Federal Circuit's "machine-or-transformation" test. Instead, the Court explicitly indicated that use of this test as an exclusive determinant of

193. MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 5 (11th ed. 2006) (defining "abstract" to mean, *inter alia*, "disassociated from any specific instance" or "dealing with a subject in its abstract aspects: theoretical"); *cf.* Parasidis, *supra* note 101, at 387 ("Over the years, courts have defined abstract ideas in various ways: as reciting a law of nature, referring to a broad application of a law of nature without limitation, something that is vague and lacking specificity, or something that refers to the manipulation of intangible entities." (citations omitted)).

194. *Cf.* Samuelson & Schultz, *supra* note 97 (manuscript at 3) (concluding from *Bilski v. Kappos*, 130 S. Ct. 3218 (2010), that "the term 'abstract idea' as a disqualification from patent protection is not limited to very high-level abstractions").

195. *Gottschalk v. Benson*, 409 U.S. 63, app. at 73–74 (1972) (internal quotation marks omitted).

the patentability of a process claim would be “inconsistent with [statutory] text and the statute’s purpose and design.”¹⁹⁶

Given the lack of an applicable plain meaning for the term “abstract idea,” a legal realist might hazard that characterization of patent claims as involving “abstract ideas” really is just formal cover for a court’s conclusion that those claims are excessively “abstract” in the sense that they are too broad to be socially justified by whatever innovative contribution has been made.¹⁹⁷ This reasoning returns us to the underlying policy concerns that can justify subject-matter exclusions.¹⁹⁸ But it does nothing to suggest that the doctrinal categories used by the courts are well-defined, transparent, or optimal for advancing social welfare. The apparent subterfuge involved in adopting an obfuscatory doctrine designed to effectively reserve for judges substantially arbitrary power seems a far from ideal way to achieve patent law’s utilitarian goals.

Unfortunately, the problems with the jurisprudence on subject-matter eligibility do not end here. In addition to giving a list of ill-defined primary exclusions, the Supreme Court has provided secondary exclusions that are at least as obscure as the category of “abstract ideas.” These are the no-preemption rule and the significance filter discussed in section III(A)(2). In *Gottschalk v. Benson*, the Court held that, even if a patent claim ostensibly escapes the reach of the “abstract idea” exclusion by limiting itself to a specified set of applications, the patent claim will still be invalid if its scope of coverage “wholly pre-empt[s]” use of the formula and therefore amounts “in practical effect” to “a patent on the algorithm itself.”¹⁹⁹ In *Parker v. Flook*, the Court added another supplementary exclusion. Even when a patent claim does not fully preempt use of an abstract idea, the claim will be unpatentable if any limitations on preemption are in some way trivial: for example, because they are achieved by restricting claim coverage to particular fields of use such as “the petrochemical and oil-refining industries,” or because they are achieved by including in the claim “conventional or obvious” “post-solution activity” such as resetting an “alarm limit” after mathematical calculation of a new alarm-limit value.²⁰⁰

196. *Bilski v. Kappos*, 130 S. Ct. 3218, 3226 (2010).

197. Cf. MORTON J. HORWITZ, *THE TRANSFORMATION OF AMERICAN LAW, 1870–1960*, at 170 (1994) (“[A]bove all, [Legal] Realism is a continuation of the Progressive attack on the attempt of late-nineteenth-century Classical Legal Thought to create a sharp distinction between law and politics and to portray law as neutral, natural, and apolitical.”).

198. See *supra* text accompanying notes 186–93.

199. *Benson*, 409 U.S. at 71–72.

200. *Parker v. Flook*, 437 U.S. 584, 589–90 (1978). In *Diamond v. Diehr*, 450 U.S. 175 (1981), the Court’s current final word on the subject-matter eligibility of software-related inventions, the Court reaffirmed the existence of the secondary exclusions but did little to clarify their scope. See *Diehr*, 450 U.S. at 191–92 (instructing that the unpatentability of mathematical formulas “cannot be circumvented by attempting to limit the use of the formula to a particular technological environment,” and that, likewise, “insignificant post-solution activity will not transform an

It should come as no surprise then that, for decades, the Federal Circuit and USPTO have struggled to make sense of the Supreme Court's exclusions.²⁰¹ The Federal Circuit's original efforts to address the patentability of software-related patents grew almost predictably tortuous,²⁰² and, with the Supreme Court remaining silent, the circuit ultimately developed an approach to subject-matter eligibility that directed attention away from prior judge-formulated exclusions. The new focus became an alternative inquiry—namely, whether the claimed invention “produce[d] a ‘useful, concrete and tangible result.’”²⁰³ But the “result” test threatened to make the requirement of subject-matter eligibility a mere corollary to patent law's separate requirement of utility. After three members of the Supreme Court called the test into question,²⁰⁴ the circuit wheeled about, abandoning the “result” test²⁰⁵ and declaring that the Supreme Court had in fact already “enunciated a definitive test to determine whether a process claim is tailored narrowly enough to encompass only a particular application of a fundamental principle”:²⁰⁶ “A claimed process is surely patent-eligible . . . if: (1) it is tied to a particular machine or apparatus, or (2) it transforms a particular article into a different state or thing.”²⁰⁷

unpatentable principle into a patentable process”). Indeed, *Diehr* likely only added confusion through a footnote suggesting that *Flook* had instructed that “a mathematical algorithm must be assumed to be within the ‘prior art’” for purposes of assessing patentability. *Id.* at 189 n.12. Such an assumption appears to have been made in *Flook* only because the patent applicant had not disputed a prior finding that Dale Flook's mathematical formula was “the only novel feature” of his claimed invention. *Flook*, 437 U.S. at 588.

201. *Cf.* BURK & LEMLEY, *supra* note 1, at 157 (“Courts have repeatedly sought to draw lines between software inventions that involved physical transformation and those that represented merely mental steps or mathematical algorithms. At each turn, those lines quickly eroded or had to be abandoned as unworkable.”); Nard, *supra* note 4, at 94 (“Over the years, the courts have constructed various tests [for claims for process inventions] These tests have proven to be too unpredictable and unruly—the standards therein offered very little guidance to the bar and inventors.”).

202. *See, e.g., In re Alappat*, 33 F.3d 1526, 1545 (Fed. Cir. 1994) (en banc) (rejecting one basis for a USPTO finding of no patentable subject matter “because a general purpose computer in effect becomes a special purpose computer once it is programmed to perform particular functions pursuant to instructions from program software”), *partially overruled by In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008) (en banc), *aff'd on other grounds sub nom. Bilski v. Kappos*, 545 F.3d 943 (Fed. Cir. 2008) (en banc); MERGES & DUFFY, *supra* note 50, at 154 (criticizing the Federal Circuit's “‘new machine’ doctrine” as “akin to arguing that, every time a person's brain holds a new thought, it is a new brain”).

203. *State St. Bank & Trust Co. v. Signature Fin. Grp., Inc.*, 149 F.3d 1368, 1374 (Fed. Cir. 1998) (quoting *In re Alappat*, 33 F.3d at 1544).

204. *Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc.*, 548 U.S. 124, 136–37 (2005) (Breyer, J., dissenting).

205. *Bilski*, 545 F.3d at 959–60 (holding that the “‘useful, concrete and tangible result’ inquiry is inadequate and reaffirm[ing] that the machine-or-transformation test . . . is the proper test to apply”).

206. *Bilski*, 545 F.3d at 954.

207. *Id.*

The Supreme Court has since declared that there is no such definitive test for patentable subject matter. According to the Court, the machine-or-transformation test is no more than “a useful and important clue, an investigative tool, for determining whether some claimed inventions are processes under § 101.”²⁰⁸ Four Justices further suggested that this clue might be conclusive for “processes similar to those in the Industrial Age,” a time period of undefined scope that apparently preceded a similarly undefined “Information Age.”²⁰⁹ Four other Justices emphasized their belief that the Court was “merely restoring the law to its historical state of rest”²¹⁰—by which they apparently meant the law’s unstable condition at the point when the Court previously exited the stage in 1981.²¹¹

The end result of this to-and-fro has been not only a remarkable lack of jurisprudential progress, but also a great deal of confusion and uncertainty. Undoubtedly, the courts could do better. In section III(A)(2), I suggest how the courts could make better sense of the Supreme Court’s secondary exclusions by explicitly characterizing them as limited safety valves that protect against potential abuse of general norms. Further, prior discussion in this section indicates that, for purposes of clarity, the primary exclusion for “laws of nature” should be collapsed into the exclusion for “abstract ideas.”

But realistically, there seems a limited amount that courts can do to bring clarity, rationality, and predictability to substantial bounds on the categories of patentable subject matter. The malleability of technology and of techniques of patent claim drafting mean that the policing of such bounds requires not only continuous vigilance, but also continual updating of guidelines for examiners and courts alike.²¹² Moreover, the task of defining

208. *Bilski v. Kappos*, 130 S. Ct. 3218, 3227 (2010).

209. *Id.* See generally Ben Cameron, Keynote Address, Symposium, *How the Show Goes On: Law and Theater in the Twenty-First Century*, 29 COLUM. J.L. & ARTS 385, 389–90 (2006) (observing that the author Daniel Pink has “posit[ed] that the American economy is already leaving the Information Age behind and moving into the Conceptual Age[.] . . . a new arts-informed age spurred by people who are hungry for meaning and definition”); John D. Feerick, Welcoming Remarks, Symposium, *The Future of Law and Financial Services*, 6 FORDHAM J. CORP. & FIN. L. 5, 5 n.3 (2001) (“While the use of the term ‘Information Age’ is prevalent, its exact definition has not been established.”).

210. *Bilski*, 130 S. Ct. at 3235 (Stevens, J., concurring).

211. *Cf. Diamond v. Diehr*, 450 U.S. 175, 219 (1980) (Stevens, J., dissenting) (opining that the Court’s decision would “aggravate” concerns that “the cases considering the patentability of program-related inventions do not establish rules that enable a conscientious patent lawyer to determine with a fair degree of accuracy which, if any, program-related inventions will be patentable”).

212. See Zakem, *supra* note 132, at 3007 (“[A]ny line excluding innovations based on their tangible attributes must constantly be redrawn to accommodate our ever-evolving understanding of technological progress.”). The social-welfare calculus for patents of a particular type might change with time even without substantial developments in techniques of claim drafting. For example, institutional and doctrinal reforms or the emergence of complementary technologies might alter the inputs into a technology-specific analog of Equation 1, see *supra* text accompanying notes 141–43. The result could be a shift in the social desirability of patents for the relevant technology type. To

included and excluded categories of subject matter necessarily implicates a substantial set of legislative facts—whether as a result of a formalist concern with how to classify new types of innovation relative to old ones, or as a result of a pragmatic concern with the present or future capacity of government officials to administer a patent system that is socially beneficial as applied to a particular category of innovation.²¹³

One might hope that a robust adversary process will cause parties appearing before courts to present and filter relevant legislative facts in a way that facilitates effective categorical decision making. But there is little reason to believe that such hope is well-founded. Even with relatively robust participation by amici curiae, a court system focused on resolving individualized disputes through the decisions of less than one thousand Article III judges²¹⁴ faces severe practical limitations on its capacity to canvass the relevant spectrum of legislative facts, to command the production of necessary additional data, and to sensibly digest whatever legislative facts are presented.²¹⁵ Further, there are additional concerns of legitimacy and democratic governance when unelected, life-tenured judges are left to devise broad-based policy in an area where Congress has been essentially silent²¹⁶ and the Executive Branch lacks policy-making authority.²¹⁷ Under such circumstances, sustained judicial fiddling with the bounds of patentable subject matter seems likely to continue to confirm a point that prominent jurists have asserted for decades: namely, that courts are essentially incompetent to address such matters in a way that takes proper account of a host of corollary scientific, economic, and policy issues.²¹⁸

the extent the scope of patentable subject matter should respond to such a shift, there might be additional cause for continual updating of relevant legal doctrines. Cf. Cotter, *supra* note 135, at 883 (“Times do change, after all, and legal doctrines must change with them.”).

213. Cf. Duffy, *supra* note 132, at 617 (“[T]he more careful judicial opinions recognize that any act, judicial or legislative, creating rules of patentable subject matter should be based on a policy judgment about the efficacy and desirability of patents in the area covered by the proposed rule.”).

214. Admin. Office of the U.S. Courts, *Judges and Judgeships*, U.S. COURTS, <http://www.uscourts.gov/JudgesAndJudgeships/FederalJudgeships.aspx>.

215. See *Diamond v. Chakrabarty*, 447 U.S. 303, 317 (1980) (“The choice we are urged to make is a matter of high policy for resolution within the legislative process after the kind of investigation, examination, and study that legislative bodies can provide and courts cannot.”).

216. See 35 U.S.C. § 101 (2006) (indicating that patentable subject matter includes “any . . . process, machine, manufacture, or composition of matter”); see also *id.* § 101 (defining “process” as “process, art or method, . . . includ[ing] a new use of a known process, machine, manufacture, composition of matter, or material”).

217. See *supra* text accompanying notes 18–21.

218. See, e.g., *Chakrabarty*, 447 U.S. at 317 (asserting that the Supreme Court was “without competence to entertain” arguments that patents on microorganisms should be restricted because of risks from modern biotechnology); *Parke-Davis & Co. v. H.K. Mulford & Co.*, 189 F. 95, 115 (S.D.N.Y. 1911) (highlighting “the extraordinary condition of the law which makes it possible for a man without any knowledge of even the rudiments of chemistry to pass upon such questions” as the patentability of purified forms of adrenaline), *aff’d in part, rev’d in part*, 196 F. 496 (2d Cir. 1912).

Moreover, at least partly because of concerns of legitimacy, democracy, and separation of powers, present-day U.S. courts tend to handcuff themselves to ways of approaching subject-matter eligibility that are unlikely to produce optimal social results. When confronted with a question of subject-matter eligibility, a U.S. court almost reflexively declares itself legally and practically incompetent to process arguments based on the practical benefits or costs expected to result from patenting a particular type of subject matter.²¹⁹ The court then tends to issue an opinion that resolves the legal question of patentability in either one of two ways: (1) the *Chakrabarty* approach of declining to enforce or to recognize substantial limits with respect to the subject matter in question,²²⁰ or (2) the *Diehr* and *Bilski* approach of (a) citing rickety doctrinal formulas from court opinions stretching back to the mid-1800s²²¹ and then (b) pointing to a smattering of facts that favor holding that the present invention falls on one or another side of those formulas.²²² Whichever path is taken, the result is a legal judgment that has little apparent basis in any carefully calibrated analysis of how, in light of technological and economic developments, limits to subject matter eligibility serve the patent system's constitutional purpose to "promote the Progress of . . . useful Arts."²²³

A realist might contend that the courts' visible legal reasoning is substantially a charade and that, in fact, the judges' rough policy judgments control. But even if this were true, and even if direct objections to the presumed non-transparency of judges' opinions were ignored, the staging of the charade would itself exacerbate concerns about judicial competence. Parties, as well as the judges, would engage in formal legal argument to help provide cover for the judges' ultimate policy choices. Precious briefing space and argument time would be devoted to formal arguments, rather than underlying

219. See *Chakrabarty*, 447 U.S. at 317 ("[W]e are without competence to entertain these arguments—either to brush them aside as fantasies generated by fear of the unknown, or to act on them.").

220. See *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int'l, Inc.*, 534 U.S. 124, 130 (2001) ("[W]e hold that newly developed plant breeds fall within the terms of § 101 . . ."); *Chakrabarty*, 447 U.S. at 308 ("[R]espondent's micro-organism plainly qualifies as patentable subject matter.").

221. See *Bilski v. Kappos*, 130 S. Ct. 3218, 3225 (2010) (citing *Le Roy v. Tatham*, 55 U.S. (14 How.) 156, 174–75 (1853), to support a statement that specified "exceptions to § 101's broad patent-eligibility principles . . . have defined the reach of the statute . . . going back 150 years"); *Diamond v. Diehr*, 450 U.S. 175, 182–85 (1980) (quoting *Le Roy*, 55 U.S. (14 How.) at 185, for the proposition that "'a principle, in the abstract, . . . cannot be patented'").

222. See *Bilski*, 130 S. Ct. at 3231 (using less than 300 words to explain the Court's specific holding that all of the petitioners' nine patent claims, which themselves used about 600 total words, see Interim Guidance for Determining Subject Matter Eligibility for Process Claims in View of *Bilski v. Kappos*, 75 Fed. Reg. 43,922, 43,926–27 (July 27, 2010) (listing the claims), lacked patentable subject matter); *Diehr*, 450 U.S. at 176 (explaining a conclusion that patent claims for a process using the Arrhenius equation did "not seek to patent a mathematical formula" but instead sought "patent protection for a process of curing synthetic rubber").

223. U.S. CONST. art. I, § 8, cl. 8.

substance. Defense of the courts' performance on grounds that they are in fact merely masking policy judgments with legal formalism thus seems no defense at all. Even a formal charade imposes constraints on the information and arguments that courts receive, substantially impairing judges' capacity to check inchoate intuitions against actual fact, counterargument, and criticism. Even if the courts are engaged in a charade to the extent they claim not to be engaging issues of social or economic policy, the charade itself does not improve their competence in resolving questions of "social-cost accounting."²²⁴

Additional problems arise from the courts' substantial potential for—and track record of—tardiness in addressing well-established questions about patentable subject matter. Although appeal from a final USPTO decision can be taken to a panel of Federal Circuit judges as of right, appeals from USPTO proceedings other than *inter partes* reexaminations generally occur only at the behest of a party seeking a patent grant.²²⁵ Consequently, patent applicants and patent holders disappointed by a USPTO ruling on subject-matter eligibility can often delay a date of judicial reckoning by abandoning rejected claims, perhaps after cutting a deal with the USPTO by pursuing narrower claims that the USPTO—or, at least, one USPTO examiner—views as acceptable.

Moreover, even when a question of subject-matter eligibility gets to the courts, definitive answers might require review by the Federal Circuit sitting en banc or by the Supreme Court. Merits review at both of these levels is discretionary and, in practice, relatively rare.²²⁶ The Supreme Court's recent five-to-four affirmation that the Patent Act encodes no general exclusion of

224. See Jerry L. Mashaw, *Administrative Due Process as Social-Cost Accounting*, 9 HOFSTRA L. REV. 1423, 1452 (1981) ("This exploration of analyzing administrative due process issues from the now-dominant social-cost accounting perspective does not provide a very substantial basis for supporting that technique's continued use."); see also Colin S. Diver, *The Optimal Precision of Administrative Rules*, 93 YALE L.J. 65, 108 (1983) ("Not only do courts lack the administrator's presumed investigative resources, analytic competence, and technical literacy, but they view social policy issues through the refracting prism of judicial review.").

225. See 35 U.S.C. § 141 (2006) (providing rights to appeal to the Federal Circuit to "[a]n applicant"; "[a] patent owner, or a third-party requester in an inter partes reexamination proceeding"; or "[a] party to an interference"); cf. *id.* § 145 (giving an applicant the option of suing in the U.S. District Court for the District of Columbia); *id.* § 145 (likewise giving an option of civil suit to "[a]ny party to an interference").

226. See Lee Petherbridge, *Patent Law Uniformity?*, 22 HARV. J.L. & TECH. 421, 424 n.23 (2009) (listing thirty-two patent-related cases decided en banc by the Federal Circuit from 1984 through 2008); *supra* text accompanying note 17; cf. *The Statistics*, 124 HARV. L. REV. 411, 418 tbl.II (2010) (reporting that the Supreme Court granted only 4.3% of paid petitions for certiorari (69 out of 1607) during October Term 2009); U.S. Court of Appeals for the Federal Circuit, Panel and En Banc Petitions for Rehearing (Feb. 8, 2011), http://www.cafc.uscourts.gov/images/stories/the-court/statistics/Panel_and_En_Banc_Petitions_for_Rehearing_2001-2010.pdf (showing that, from fiscal year 2001 through fiscal year 2010, the Federal Circuit took 14 cases en banc sua sponte and granted 28 of 1266 petitions for en banc rehearing).

patents for business methods²²⁷ is a case in point. This affirmation came in the Court's *Bilski* decision of 2010, more than a decade after a Federal Circuit panel had so ruled and Congress had responded by enacting a statutory provision specifically designed to limit business-method patents' reach.²²⁸

In addition to impeding development of the sort of legal certainty that facilitates private planning and ordering, judicial delays in definitive review can distort doctrinal outcomes. When a court explores limiting subject-matter eligibility after a lengthy period during which hundreds or even thousands of patents on the relevant subject matter have issued,²²⁹ the societal deck might be stacked against a course correction. In addition to the reliance interests developed by private parties, other strands of law might have developed in a way that reflects an assumption of subject-matter eligibility.²³⁰ Unless courts are immune to concerns about disrupting private planning and statutory schemes, the result seems likely to be an increased probability that the patent system will lock itself into a suboptimal legal approach that more timely review could have avoided or mitigated.²³¹

Of course, an administrative agency could also temporize and equivocate. But the USPTO cannot issue patents for a particular type of subject matter without at least implicitly taking a stance on subject-matter eligibility. And in a world where thousands of patent examiners need to know how to deal with hundreds of thousands of pending patent applications, the USPTO faces great pressure to provide guidance promptly. Hence, on the same day as the U.S. Supreme Court's June 2010 decision on patentable subject matter in *Bilski v. Kappos*, the USPTO issued a memorandum to examiners on how they should approach questions of subject-matter eligibility affected by the Supreme Court's decision.²³² Less than one month later, the USPTO published in the Federal Register further interim guidelines, inviting

227. See *Bilski*, 130 S. Ct. at 3228 ("Section 101 similarly precludes the broad contention that the term 'process' categorically excludes business methods.").

228. See *id.* at 3232 (Stevens, J., concurring) ("Congress quickly responded to a Federal Circuit decision with a stopgap measure . . .").

229. Cf. *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int'l, Inc.*, 534 U.S. 124, 145 (2001) (observing that "some 1,800 utility patents for plants" had already issued by the time the Supreme Court was addressing whether restrictions on patentable subject matter forbade regular-utility patents for plants).

230. See MERGES & DUFFY, *supra* note 50, at 173 (arguing that enactment of a statutory provision limiting the rights of owners of business-method patents "had the effect of entrenching" *State Street's* determination that there is no business-method exception to patentability).

231. Cf. Liza Vertinsky, *Comparing Alternative Institutional Paths to Patent Reform*, 61 ALA. L. REV. 501, 536-37 (2010) ("Where specificity is important, but flexibility to deviate from past practices is also important, agency rulemaking offers an alternative to judicial decision-making.").

232. Memorandum from Robert W. Bahr, Acting Assoc. Comm'r for Patent Examination Policy, USPTO, to Patent Examining Corps 2 (June 28, 2010) [hereinafter Memorandum to Patent Examining Corps], available at http://www.uspto.gov/patents/law/exam/bilski_guidance_28jun2010.pdf.

public comment even while making the guidelines immediately effective.²³³ The stark contrast between the USPTO's prompt post-*Bilski* actions and the Supreme Court's several-months delay in issuing its *Bilski* decision²³⁴ highlights the comparative urgency of subject-matter concerns for the USPTO.

Moreover, even if the likelihood of administrative delays and vacillation were comparable to judicial delays and vacillation, the suboptimality resulting from judicial delay or vacillation might be worse.²³⁵ Somewhat counterintuitively in light of conventional wisdom that courts can be better trusted to ensure legal stability,²³⁶ this follows because U.S. law tends to provide substantial protections against "surprises" due to administrative-agency action, but provides relatively little protection—at least in non-criminal contexts—from disruptive statutory (re)interpretation by courts.

Administrative agencies have great capacity to restrict the retroactive effect of rulemaking that adopts a new statutory interpretation. Most obviously, the agency can often choose to delay the effective date of a new rule or can limit the subject matter to which the new rule applies. Judges have occasionally invoked nonretroactivity norms or even the Administrative Procedure Act to require an administrative agency to implement such restrictions.²³⁷ Generally speaking, Congress also has ready means to tailor legislation in accordance with nonretroactivity norms. Recent legislative proposals to prohibit patents for methods of tax planning have explicitly included provisions specifying that the prohibition will have only prospective effect.²³⁸

233. Interim Guidance for Determining Subject Matter Eligibility for Process Claims in View of *Bilski v. Kappos*, 75 Fed. Reg. 43,922, 43,922 (July 27, 2010).

234. See *Bilski v. Doll*, 129 S. Ct. 2735, 2735 (2009) (granting certiorari on June 1, 2009); *Bilski v. Kappos*, 130 S. Ct. 3218 (2010) (issuing opinion on June 28, 2010).

235. Cf. Burstein, *supra* note 7 (manuscript at 5) (illustrating "the uncertainty created by judicial decision-making" by noting how, "[i]n the case of gene patents, the vagaries of litigation prevented a core question of patentability from being litigated before the industry assumed significant reliance costs").

236. Matthew C. Stephenson, *Legislative Allocation of Delegated Power: Uncertainty, Risk, and the Choice Between Agencies and Courts*, 119 HARV. L. REV. 1035, 1038 (2006) ("As a rough generalization, agency interpretive decisions tend to be ideologically consistent across issues but variable over time, while judicial interpretations tend to be ideologically heterogeneous across issues but stable over time."); see also Daniel A. Farber, *Modeling Coherence, Stability, and Risk Aversion in Legislative Delegation Decisions*, 119 HARV. L. REV. F. 157, 157 (2006) (characterizing as a "reasonable assumption" the proposition that "courts are more reluctant to depart from previous decisions," but "agencies are more likely to be consistent in their decisions at any given time").

237. See, e.g., *Bowen v. Georgetown Univ. Hosp.*, 488 U.S. 204, 208 (1988) ("Retroactivity is not favored in the law."); *id.* at 216 (Scalia, J., concurring) (concluding from the language of the Administrative Procedure Act "that rules have legal consequences only for the future"); see also Magill, *supra* note 19, at 1435 ("Under certain circumstances, a court will treat the legal rule established in an [agency] adjudication as prospective only.").

238. See Patent Reform Act of 2007, H.R. 1908, 110th Cong. § 10(b) (2007) (providing that a ban on tax-planning patents would "take effect on the date of the enactment," would apply to pending or later-filed applications, and should "not be construed as validating any patent issued

In contrast, a judicial change of course will necessarily come through adjudication, commonly treated by U.S. law as a characteristically backward-looking rather than forward-looking process.²³⁹ Consistent with judges' tendency to view themselves as interpreters, rather than makers, of law, the judicial course change will likely be packaged as a foreordained result of preexisting law. This packaging makes acknowledgment and mitigation of disruptive change difficult to manage.²⁴⁰ Hence, if, for example, the Supreme Court had decided in 2001 that utility patents could not issue for plant inventions, it would have been considered extraordinary to excuse from the reach of its ruling any of "some 1,800 utility patents for plants" that had issued over the course of "at least 16 years."²⁴¹ Provision of exemptions for previously issued patents on the grounds of parties' reasonable expectations might be thought fair but might also seem too much like an admission that the Court's decision had effectively altered applicable policy rather than merely applied preexisting law.

This analysis of problems with the judicial development of rules for subject-matter eligibility resonates with past debates over the wisdom of administrative agencies' developing broadly applicable legal doctrine through relatively formal processes of adjudication rather than more informal processes of rulemaking.²⁴² As suggested by those debates, the types of

before the date of the enactment"); S. 2369, 110th Cong. § 1(b) (2007) (same); Stop Tax Haven Abuse Act, S. 681, 110th Cong. § 303 (2007) (providing that a prohibition of patents for inventions "designed to minimize, avoid, defer, or otherwise affect [tax] liability" would "take effect on the date of the enactment . . . and apply to any application for a patent that has not been granted by that date").

239. See, e.g., *Bowen*, 488 U.S. at 221 (Scalia, J., concurring) ("Adjudication deals with what the law was; rulemaking deals with what the law will be." (emphasis omitted)); Shyamkrishna Balganes, *The Pragmatic Incrementalism of Common Law Intellectual Property*, 63 VAND. L. REV. 1543, 1587 (2010) (discussing how "the backward-looking nature of adjudication" interacts with "[l]egal pragmatism[s]'s insist[ence] that adjudication be forward-looking" (internal quotation marks omitted)); Christopher J. Peters, *Assessing the New Judicial Minimalism*, 100 COLUM. L. REV. 1454, 1501 (2000) ("[A]djudication for the most part is retrospective; it assesses events that already have occurred.").

240. Cf. J. Nicholas Bunch, Note, *Takings, Judicial Takings, and Patent Law*, 83 TEXAS L. REV. 1747, 1762 (2005) ("[A]lthough courts have recognized that Takings Clause concerns arise when settled rules of decision are departed from in the context of property rights, they have also consistently refused to find that a judicial taking requiring payment of just compensation has occurred.").

241. *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred Int'l, Inc.*, 534 U.S. 124, 144–45 (2001). See generally PETER L. STRAUSS ET AL., GELLHORN AND BYSE'S ADMINISTRATIVE LAW: CASES AND COMMENTS 131 (rev. 10th ed. 2003) (observing that, in connection with a 1982 decision holding that the Bankruptcy Act violated Article III, the Supreme Court "employ[ed] the extraordinary remedy of staying its judgment for more than three months" to give Congress an opportunity to enact substitute legislation (emphasis added)).

242. See Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 HARV. L. REV. 353, 398 (1978) (suggesting "managerial direction and contract" as alternatives to adjudication for the solution of "sufficiently polycentric" issues (emphasis omitted)); Peter L. Strauss, *Revisiting Overton Park: Political and Judicial Controls over Administrative Actions Affecting the Community*, 39 UCLA L. REV. 1251, 1257 (1992) (arguing that "judges' strengths lie in resolving discrete

issues for which development of legal doctrine through adjudication looks advisable are often distinct from the types of issues for which rulemaking appears to have a comparative advantage.²⁴³ In the patent-law context, courts' characteristically case-specific and backward-looking approaches to resolving disputes might work passing well for such characteristically case-specific and backward-looking questions as novelty and nonobviousness, the adequacy of the patentee's disclosure, and patent claims' meaning—all of which are tied to past times of invention, patent application, or patent issue. Further, the heavy adjudicative-fact flavor of these issues tends to limit (a) demands for precise *ex ante* guidance and (b) disruptions that might result from a court decision's retroactive effects: the doctrine enunciated tends naturally to be both less outcome restrictive and more narrowly pronounced. Accordingly, in the Supreme Court's leading opinions on the nonobviousness requirement, the 1966 opinion in *Graham v. John Deere Co. of Kansas City*²⁴⁴ and the 2007 opinion in *KSR International Co. v. Teleflex Inc.*, the Court primarily enunciated forms of facts and factors that courts either should or must consider, and then exemplified application of described modes of analysis through decisions based on individualized sets of facts.²⁴⁵

As subpart III(A) argued, questions of subject-matter eligibility are comparatively nonindividualized. At least under existing understandings, the outcome for such questions does not seem generally to depend on a specific past outlook, past state of knowledge, or past disclosure. Instead, questions of subject-matter eligibility appear intended to reflect relatively broad categorical distinctions not tied to a particular temporal reference point. Such categorical judgments are especially likely to have strong effects on a variety of parties and to implicate polycentric policy concerns best addressed after wide-ranging fact-finding and perhaps even coordinated readjustment of multiple doctrines or practices.²⁴⁶ In combination, the substantially categori-

controversies between individuals, in which . . . broad social adjustments are secondary to the outcome of their concrete dispute").

243. See David L. Shapiro, *The Choice of Rulemaking or Adjudication in the Development of Administrative Policy*, 78 HARV. L. REV. 921, 927–29 (1965) (describing situations “in which it would be a mistake to attempt much more than the resolution of each case as it comes”).

244. 383 U.S. 1 (1966).

245. See, e.g., *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 415 (2007) (describing *Graham* as “set[ting] forth a broad inquiry” and reaffirming the need for “an expansive and flexible approach” to considering facts and theories relevant to an assessment of nonobviousness); *Graham*, 383 U.S. at 17 (instructing that, in an assessment of nonobviousness, “the scope and content of the prior art are to be determined; differences between the prior art and the claims at issue are to be ascertained; and the level of ordinary skill in the pertinent art resolved”).

246. Creation of the United States Court of Appeals for the Federal Circuit has not solved this problem. Although the Federal Circuit's semi-specialized docket enables it to develop and to apply *legal expertise* through frequent decision of questions of patent law, the court remains fundamentally limited in its capacity to develop *factual or policy expertise* relating to the nature, development, or dissemination of present-day innovations. The Federal Circuit's twelve active judges cannot be expected to be masters of the technical and economic details of all the new

cal nature of determinations of patentable subject matter and a history of judicial struggle to make such determinations suggest that applicable rules for patentable subject matter are likely to be better developed through legislative or quasi-legislative processes, rather than precedent-bound, case-by-case adjudication.

2. *Undesirability of Reliance on Congress.*—If legislative or quasi-legislative rulemaking is desirable, why not avoid the complications of the latter and just demand that Congress go to work?

The question is almost too easy to answer. First, Congress appears to be too slow moving to provide substantially useful, updated instruction for patent law's application to ever-evolving technologies.²⁴⁷ Second, even if Congress can act quickly enough, it tends to be too uninformed and too liable to special-interest manipulation to be trusted to craft statutory language that adequately advances the public interest on questions that are likely too technical to garner widespread public attention.²⁴⁸ Congress's defunding of its Office of Technology Assessment in the 1990s²⁴⁹ suggests that Congress itself has lost faith in its capacity to conduct the sort of thorough, careful fact-

technologies—or possible technologies—that their decisions can affect. Nor has the circuit been given the broad powers of independent “investigation, examination, and study” that legislatures typically have but U.S. courts have typically lacked. *Diamond v. Chakrabarty*, 447 U.S. 303, 317 (1980). Finally, although some aspects of the doctrinal edifice erected by the Federal Circuit might be characterized as reflecting forms of policy tradeoffs, the circuit's ability to make optimal tradeoffs is limited by the circuit's restriction to deciding issues only as they arise in individual cases.

247. Cf. BURK & LEMLEY, *supra* note 1, at 105 (“The prospect of the legislature continually revisiting the circumstances of each industry and passing appropriate new legislation for each situation is equally bleak.”); Nard, *supra* note 4, at 90 (“Faced with the non-linear path of technological innovation and diversity of inventions, it is understandable and desirable that Congress maintained [35 U.S.C.] § 101's broad standard, which implicitly signaled to the courts to fill in the statutory interstices.”); R. Polk Wagner, *The Supreme Court and the Future of Patent Reform*, FED. LAW., Feb. 2008, at 35, 38–39 (contending that, “[a]s the patent law becomes more politicized and the stakes rise, the opportunities for substantial [legislative] reform of the system narrow”); George Packer, *The Empty Chamber: Just How Broken Is the Senate?*, NEW YORKER, Aug. 9, 2010, at 38, 49 (predicting that “the Senate will remain a sclerotic, wasteful, unhappy body”).

248. See David S. Olson, *Taking the Utilitarian Basis for Patent Law Seriously: The Case for Restricting Patentable Subject Matter*, 82 TEMP. L. REV. 181, 237 (2009) (indicating likely drawbacks of congressional assumption of active gatekeeping responsibilities for patentable subject matter); Vertinsky, *supra* note 231, at 526 (“[T]he role of industry groups in influencing and even drafting intellectual property legislation has been noted, and a lack of general familiarity with the specialized issues raised by patent law may intensify the room for capture by special interest groups.”); cf. Nard, *supra* note 4, at 103 (arguing that “the common law . . . is more responsive [than Congress] to the technological communities that form part of patent law's fabric”).

249. See Stuart Minor Benjamin & Arti K. Rai, *Fixing Innovation Policy: A Structural Perspective*, 77 GEO. WASH. L. REV. 1, 21 (2008) (“[W]ith the abolition of the Office of Technology Assessment . . . in the mid-1990s, the ability of Congress to secure unbiased advice on questions of innovation policy is limited.”).

finding that is often too blithely thought a peculiar legislative competence.²⁵⁰ Meanwhile, the mass of special-interest legislation, commonly favoring concentrated publisher interests, that seems to have overtaken much of U.S. copyright law provides a cautionary example of what a frequently engaged Congress might produce.²⁵¹

On the other hand, the sparseness of present legislative instruction on patentable subject matter is such that some high-level work by Congress could help.²⁵² For purposes of providing a framework for action by the USPTO and the courts, the U.S. Patent Act could feature more than a list of four broad categories of potentially patentable subject matter—"process, machine, manufacture, or composition of matter"²⁵³—supplemented by a substantially circular definition of the term "process" as "process, art or method."²⁵⁴ Congress could amend the Act to abrogate or to codify one or more of the courts' primary or secondary exclusions. Congress could also choose to list additional statutory exclusions. For example, Congress could codify something along the lines of an exclusion for patent claims "encompassing a human being," thereby ratifying an exclusion that the USPTO's guidelines have already recognized.²⁵⁵ Likewise, Congress could follow through on legislative proposals to prohibit patents for methods of tax planning or avoidance.²⁵⁶

In carving out such exclusions, Congress would not lack legal models. Both the international TRIPS agreement and the European Patent Convention already provide for exclusions along these lines. Despite requiring member

250. Cf. BURK & LEMLEY, *supra* note 1, at 105 ("[A]lthough legislatures are often characterized as having the better resources to investigate and develop factual evidence, this capability is often more theoretical than actual."); Brian Galle & Mark Seidenfeld, *Administrative Law's Federalism: Preemption, Delegation, and Agencies at the Edge of Federal Power*, 57 DUKE L.J. 1933, 1939 (2008) (contesting notions that administrative agencies are "not as transparent, deliberative, or accountable" as Congress).

251. See Clarisa Long, *Our Uniform Patent System*, FED. LAW., Feb. 2008, at 44, 48 (citing the U.S. Copyright Act as "an example of what can happen when we create special exemptions and other industry-specific legislation").

252. Cf. Thomas, *supra* note 117, at 1184 ("A legislative approach appears the best possibility for reminding the patent system that not everything we do is technological").

253. 35 U.S.C. § 101 (2006).

254. *Id.* § 100(b).

255. See, e.g., MPEP, *supra* note 28, § 2105, at 2100-5 ("If the broadest reasonable interpretation of the claimed invention as a whole encompasses a human being, then a rejection under 35 U.S.C. 101 must be made indicating that the claimed invention is directed to nonstatutory subject matter."). Congress has repeatedly supported such an exclusion through appropriations bills but has not amended the Patent Act to require such an exclusion. Golden, *supra* note 49, at 318.

256. See, e.g., Patent Reform Act of 2007, H.R. 1908, 110th Cong. § 10 (2007) (proposing to amend the Patent Act to prohibit patents for "tax planning method[s]"); S. 2369, 110th Cong. § 1 (2007) (proposing to amend the Patent Act to prohibit patents for "tax planning invention[s]"); Stop Tax Haven Abuse Act, S. 681, 110th Cong. § 303 (2007) (proposing to amend the Patent Act to prohibit patents for inventions "designed to minimize, avoid, defer, or otherwise affect the liability for Federal, State, local, or foreign tax").

states to make patents available “in all fields of technology” involving “industrial application,”²⁵⁷ the TRIPS agreement specifically permits member states to “exclude from patentability . . . diagnostic, therapeutic and surgical methods for the treatment of humans or animals” and also “plants and animals other than micro-organisms.”²⁵⁸ The European Patent Convention takes advantage of this allowance and explicitly forbids patents for “discoveries, scientific theories and mathematical methods”,²⁵⁹ “aesthetic creations”,²⁶⁰ “schemes, rules, and methods for performing mental acts, playing games or doing business, and programs for computers”,²⁶¹ “presentations of information”,²⁶² “plant or animal varieties”,²⁶³ and “methods for treatment of the human or animal body.”²⁶⁴ Without commenting on the wisdom of these particular exclusions here, one might concede that they would provide clearer marching orders for the USPTO and the courts than the U.S. Patent Act’s current comparative silence.²⁶⁵ Moreover, congressional action to adopt such exclusions might be sufficiently general that either a Madisonian clash of private interests or the relative diffuseness of interests involved could give the legislation reasonable odds of avoiding too much special-interest bias.²⁶⁶

But any plausible degree of congressional clarification would still leave much hard work to be done by the courts and the USPTO. These entities would need to construe the scope of whatever categories Congress excluded, often in situations where, whether through artful patent drafting or through the natural development and intermixing of technologies and their effects, patent claims defy categorical archetypes.

Such a situation arose in *Diamond v. Diehr*, the 1981 case in which the Supreme Court confronted patent claims for a process that centrally relied on

257. TRIPS, *supra* note 145, art. 27(1).

258. *Id.* art. 27(3).

259. European Patent Convention art. 52(2)(a), Dec. 13, 2007, available at <http://www.epo.org/patents/law/legal-texts/epc.html>.

260. *Id.* art. 52(2)(b).

261. *Id.* art. 52(2)(c).

262. *Id.* art. 52(2)(d).

263. *Id.* art. 53(b).

264. *Id.* art. 53(c).

265. *But cf.* Duffy, *supra* note 132, at 638 (“[T]he 1973 European Patent Convention attempted to codify a rule forbidding patents on ‘programs for computers,’ and yet that rule has been so eroded that most commentators and practitioners believe that computer programs have become as patentable in Europe as they are in the United States.” (footnote omitted)); Susan Marsnik & Robert E. Thomas, *Drawing a Line in the Patent Subject Matter Sands: Does Europe Provide a Solution to the Business Method and Software Patent Problem?*, B.C. INT’L & COMP. L. REV. (forthcoming 2011) (manuscript at 3), available at <http://ssrn.com/abstract=1676898> (“The clarity of the [European Patent Convention] exclusion is severely muddled by modifying language in Article 52(3) and by the complexity and fragmentation of the European patent system.” (footnote omitted)).

266. *Cf.* THE FEDERALIST No. 10, at 83 (James Madison) (Clinton Rossiter ed., 1961) (“Extend the sphere and you take in a greater variety of parties and interests; you make it less probable that a majority . . . will have a common motive to invade the rights of other citizens; or if such a common motive exists, it will be more difficult . . . to act in unison . . .”).

a computer program to implement a relatively simple mathematical test.²⁶⁷ The claim's combination of a computer program with a traditional manufacturing process complicated the problem of classification. The pro-patentability majority ultimately viewed the process as "nothing more than a process for molding rubber products."²⁶⁸ The dissenting Justices viewed the category-straddling claim differently, seeing the case as one implicating the "broad question" of the patentability of "computer programs."²⁶⁹

Statutory delimitation of excluded categories will not eliminate, and might even multiply, such line-drawing problems. As with questions of patent infringement, technological change and developments in claim drafting can be expected consistently to test legal boundaries. Consequently, legal doctrines for subject-matter eligibility will require continued, adaptive interpretation in order to maintain a semblance of clarity and substantive relevance. Moreover, such satisfactory adaptation will likely demand sustained, detailed attention to the nature of relevant technologies and developments in patent description and claim drafting. In this regard, the European example is instructive: the exclusions stated in the European Patent Convention have only set the table for continued struggle to determine the excluded categories' scope.²⁷⁰

Congress's shortcomings in time, expertise, and vulnerability to special interests appear likely to be at or near their peak in dealing with such subsidiary problems of line drawing. With respect to such problems, Congress would have to respond not just to major tectonic shifts in technology, such as the development of modern digital computers and associated software, but also to more minor developments, such as the incorporation of digital-computer technology in familiar "Industrial Age" devices or processes like refrigerators, cars, or methods of molding plastic. The time cycles for required tweaks would be tighter, and the typical scope and nature of individual tweaks would likely be narrower and more abstruse. Congress's attention span, understanding, and capacity for timely action would be much more severely strained than in providing basic policy direction through the identification of general categories of included or excluded matter. Further, the narrower nature of the issues involved would likely increase the risk that the spectrum of interested parties would be substantially unbalanced. In light

267. See *Diamond v. Diehr*, 450 U.S. 175, 177–79 (1981) (describing the respondents' claimed "contribution to the art").

268. *Id.* at 191.

269. *Id.* at 216 (Stevens, J., dissenting).

270. See GRAEME B. DINWOODIE ET AL., *INTERNATIONAL AND COMPARATIVE PATENT LAW* 86–88 (2002) (discussing European developments with respect to the patenting of claimed plant and animal inventions); PHILIP LEITH, *SOFTWARE AND PATENTS IN EUROPE* 27–34 (2007) (discussing interpretation of the European Patent Convention's exclusions from patentability in the context of computer-related innovation).

of such concerns, we might hope that an agency armed with relevant rule-making authority would do substantially better.

3. *The USPTO's Comparative Institutional Competence.*—With Congress apparently unable to do an adequate job of developing legal doctrine on subject-matter eligibility, we are back to the debate over whether responsibility for that task should lie primarily with the courts or an administrative agency. My sense is that the latter is the better locus for such responsibility. This results from the previous section's conclusion that, even if Congress provides some additional instruction in this area, it will almost necessarily remain very high level. The courts' basic difficulties in settling on effectual meanings for categories of patentable or unpatentable subject matter will therefore remain.

Indeed, the courts' difficulties recall the problem of interpretation and application that underlay perhaps the most prominent case in U.S. administrative law,²⁷¹ *Chevron, Inc. v. Natural Resources Defense Council*.²⁷² In this case, the U.S. Supreme Court confronted a statutory term, "stationary source," that the malleability of technology had turned into a linguistic muddle.²⁷³ The problem in *Chevron* was how to define a "source" of pollutant emissions when, for example, a single factory complex could include multiple buildings, each having multiple smokestacks.²⁷⁴ Should the factory complex be viewed as a single source subject to one overall level of emissions regulation? Should each distinct building or each smokestack instead be counted as a separate source? Alternative candidate constructions

271. See STEPHEN G. BREYER ET AL., *ADMINISTRATIVE LAW AND REGULATORY POLICY: PROBLEMS, TEXT, AND CASES* 247 (6th ed. 2006) ("In a remarkably short period, *Chevron* has become one of the most cited cases in all of American law.").

272. 467 U.S. 837 (1984).

273. The Supreme Court described the interpretive issue in *Chevron* as follows:

The EPA regulation promulgated to implement [a] permit requirement allows a State to adopt a plantwide definition of the term "stationary source." Under this definition, an existing plant that contains several pollution-emitting devices may install or modify one piece of equipment without meeting the permit conditions if the alteration will not increase the total emissions from the plant. The question presented . . . is whether the EPA's decision to allow States to treat all of the pollution-emitting devices within the same industrial grouping as though they were encased within a single "bubble" is based on a reasonable construction of the statutory term "stationary source."

Id. at 840.

274. See *id.*

appeared equally consistent with statutory provisions' plain language.²⁷⁵ Further, legislative history was uninformative.²⁷⁶

The unanimous Court resolved this nettle through an institutional shuffle. The Court appears to have concluded that under these circumstances, the proper resolution of this definitional question was fundamentally a technical policy question that, to the extent possible, courts should leave to reasonable resolution by the agency charged with implementing the statute.²⁷⁷ The courts would reserve the power to check that the agency's resolution was indeed reasonable, but the courts would not bind the agency to either the agency's past interpretations or even past judicial interpretations.²⁷⁸ Such reasoning gave rise to the *Chevron* doctrine, which provides that, in various circumstances, courts should understand agencies to have primary interpretive authority with respect to congressional acts.²⁷⁹

This Article's argument comports with fundamental intuitions underlying the Supreme Court's opinion in *Chevron*. As with *Chevron*'s smokestacks and plant designs, the malleability of technology and of approaches to describing it will often reveal statutory ambiguity with respect to subject-matter eligibility that traditional techniques of statutory interpretation cannot satisfactorily resolve. The categorical, policy-laden nature of the resulting interpretive questions suggests that these questions are better left to primary resolution by a policy organ specially concerned with such questions—namely, an administrative agency.²⁸⁰

275. See *Chevron*, 467 U.S. at 862 (“To the extent any congressional ‘intent’ can be discerned from this language, it would appear that the listing of overlapping, illustrative terms was intended to enlarge, rather than to confine, the scope of the agency’s power to regulate particular sources in order to effectuate the policies of the Act.”).

276. *Id.* (“Based on our examination of the legislative history, we agree with the Court of Appeals that it is unilluminating.”).

277. See *id.* at 865 (“[T]he Administrator’s interpretation represents a reasonable accommodation of manifestly competing interests and is entitled to deference: the regulatory scheme is technical and complex, the agency considered the matter in a detailed and reasoned fashion, and the decision involves reconciling conflicting policies.” (footnotes omitted)).

278. See *id.* at 863 (“An initial agency interpretation is not instantly carved in stone.”); see also *Nat’l Cable & Telecomms. Ass’n v. Brand X Internet Servs.*, 545 U.S. 967, 981 (2005) (“Agency inconsistency is not a basis for declining to analyze the agency’s interpretation under the *Chevron* framework.”); *id.* at 982 (“A court’s prior judicial construction of a statute trumps an agency construction otherwise entitled to *Chevron* deference only if the prior court decision holds that its construction follows from the unambiguous terms of the statute and thus leaves no room for agency discretion.”).

279. JOHN F. MANNING & MATTHEW C. STEPHENSON, *LEGISLATION AND REGULATION* 822–23 (2010) (stating that “the overall thrust of *Chevron* is fairly clear: If the responsible administrative agency has reasonably resolved a statutory ambiguity, the reviewing court should accept the agency’s resolution”).

280. Cf. Olson, *supra* note 245, at 185 (indicating that “the optimal solution [to regulating subject-matter eligibility] may be to assign an administrative agency the task of conducting explicit utilitarian analysis and rulemaking”); Allen K. Yu, *Within Subject Eligibility—A Disease and a Cure*, 84 S. CALIF. L. REV. (forthcoming 2011) (manuscript at 58), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1653454 (“A strong version of a policy-driven administrative

But a significant question remains. What agency should have this responsibility? The most obvious choice appears to be the USPTO. The USPTO's need to resolve questions of subject-matter eligibility as part of its daily business suggests that it will—despite concerns about rulemaking “ossification”²⁸¹—have ample incentive to update its understandings in this area as new technologies and claim-drafting techniques demand.²⁸²

A proposal to give the USPTO such authority will likely encounter an array of objections. At least in response to proposals to give the USPTO primary interpretive authority for *all* questions of substantive patent law, commentators have voiced objections that largely to reflect the USPTO's existing capacities or liability to capture and bias.

A first objection, recently voiced by Dan Burk and Mark Lemley, is that the USPTO currently lacks the resources and experience to handle substantive questions of statutory interpretation more competently than the courts.²⁸³ An initial response is that the USPTO's capacities can be expected to grow with its responsibilities. Further, the USPTO's current interpretive capacities are not as feeble as might commonly be thought. The USPTO already does a significant amount of rulemaking, whether in the form of substantive guidance to examiners and patent applicants or in the form of procedural rules that have substantial, practical effects. In performing such rulemaking, the USPTO has regularly undergone processes of notice-and-comment rulemaking that are characteristic of an exercise of binding rule-making authority.²⁸⁴ Although the USPTO recently abandoned a controversial rules package that sought to limit continuation applications and claims,²⁸⁵ the USPTO has had significant triumphs in connection with substantive guidelines on patent law's utility and disclosure requirements, as is discussed further below.²⁸⁶

[intellectual property] regime would not only call on [an] administrative agency to define subject matter eligibility, but perhaps also to invalidate issued . . . patents on the ground they impede rather than foster innovations . . .”).

281. See Jason Webb Yackee & Susan Webb Yackee, *Testing the Ossification Thesis: An Empirical Examination of Federal Regulatory Volume and Speed, 1950–1990*, at 108 (Nov. 30, 2010) (unpublished manuscript), available at <http://ssrn.com/abstract=1699878> (finding only “mixed and relatively weak evidence that ossification is either a serious or widespread problem” in a study of rulemaking by divisions of the U.S. Department of the Interior).

282. See *supra* text accompanying notes 231–35.

283. BURK & LEMLEY, *supra* note 1, at 168 (describing the USPTO as having “virtually no policy staff and—at least until recently—little experience or apparent inclination to take a leadership role in setting patent policy”).

284. See *infra* text accompanying notes 346–57.

285. Press Release, U.S. Patent & Trademark Office, *USPTO Rescinds Controversial Patent Regulations Package Proposed by Previous Administration* (Oct. 8, 2009), http://www.uspto.gov/news/09_21.jsp (describing the rules and their rescission).

286. See *infra* text accompanying notes 350–57.

Moreover, any USPTO deficiencies must be viewed in context. No agency currently has a mix of personnel, expertise, and experience that would be ideal for exercising primary interpretive authority on questions of patentable subject matter. Although the FTC has shown a capacity to influence patent-law developments by holding hearings or issuing reports,²⁸⁷ the FTC has not sought to demonstrate expertise in crafting specific patent-law doctrine. Because two of the central challenges for subject-matter eligibility involve formulating doctrine (1) that the USPTO's thousands of examiners can effectively administer and (2) that can provide effective guidance to patent attorneys, patent agents, and their clients, it would seem less than ideal to entrust substantive development of rules on subject-matter eligibility to an agency that is detached from processes of patent examination and prosecution.

A second and more serious set of objections to giving the USPTO substantive rulemaking authority derives from concerns that the USPTO will be overly subject to capture or bias.²⁸⁸ In this regard, the primary concern tends to be that, in wielding rulemaking authority, the USPTO will be overly solicitous of patent-applicant interests. USPTO documents and officials sharpen this concern when, presumably as part of an effort to improve the quality of USPTO operations, they describe applicants as "customers."²⁸⁹ The *ex parte* nature of the dominant forms of USPTO proceedings—examination proceedings and *ex parte* reexamination proceedings—means that USPTO examiners primarily interact with parties seeking to obtain patent rights. There is therefore natural cause for concern that USPTO personnel will become subject to a form of intellectual or informational capture and tend to view the world through patent applicants' lenses.²⁹⁰

There are additional public-choice problems. Major patent holders and trade groups have been visibly active in lobbying political actors who oversee USPTO performance, whether located on Capitol Hill or in the White House.²⁹¹ By pressuring public officeholders to put pressure on the USPTO, these sorts of substantial, concentrated interests could overwhelm the more diffuse, less easily coordinated interests of both the holders of smaller patent

287. See, e.g., *Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc.*, 548 U.S. 124, 134 (2006) (Breyer, J., dissenting) (citing a 2003 FTC report); *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 396 (2006) (Kennedy, J., concurring) (citing the same report).

288. See, e.g., Nard, *supra* note 4, at 56–57 (“[J]udicial primacy acts as a bulwark against the more politicized legislative process or capture-prone administrative rulemaking.”).

289. See BURK & LEMLEY, *supra* note 1, at 107 (“It is little wonder . . . that the [US]PTO in the 1990s stated its mission as ‘to help our customers get patents.’ That’s capture.”).

290. Cf. Wendy E. Wagner, *Administrative Law, Filter Failure, and Information Capture*, 59 DUKE L.J. 1321, 1325 (2010) (“In the regulatory context, information capture refers to the excessive use of information and related information costs as a means of gaining control over regulatory decisionmaking in informal rulemakings.”).

291. Cf. Golden, *supra* note 16, at 507 (discussing the lobbying efforts of industry groups).

portfolios and the non-patent-holding public.²⁹² Further, to the extent USPTO officials plan to seek employment in the private sector after their public service, major patent holders might offer greater and more lucrative opportunities, and might be more likely to hire persons who have already shown sympathy for their interests.²⁹³ Even if applicant interests do not capture the USPTO, the USPTO might suffer from mission bias.²⁹⁴ USPTO policy and practice could systematically deviate from the public interest in any of a number of directions. For example, USPTO concern with avoiding ridicule for granting apparently poor-quality patents could lead its examiners to focus too exclusively on avoiding wrongly issued patents (Type I errors) at the cost of underemphasizing avoidance of improperly rejecting meritorious claims (Type II errors).²⁹⁵ Alternatively, the USPTO could define its mission excessively in terms of processing patent applications, with the speedy issue of patent rights to satisfied “customers” becoming the easiest route to a sense of “mission accomplished.”

But concerns of capture and bias are normal in administrative law. A number of commentators suggest that such concerns have been systematically exaggerated.²⁹⁶ In any event, such concerns generally have not been understood to justify wholesale disempowerment of administrative agencies. Despite concerns that agencies such as the Securities and Exchange Commission,²⁹⁷ the Environmental Protection Agency,²⁹⁸ or the Food and

292. See Nicholas Bagley & Richard L. Revesz, *Centralized Oversight of the Administrative State*, 106 COLUM. L. REV. 1260, 1305 (2006) (“Like any elected official, the President will be particularly attentive to those groups that can provide him with the resources, support, or votes to win elections or promote his political agenda.”); Mark Seidenfeld, *Who Decides Who Decides: Federal Regulatory Preemption of State Tort Law*, 65 N.Y.U. ANN. SURV. AM. L. 611, 647 (2010) (“If capture [of agency rulemaking] happens, it often occurs when the industry convinces Congress and the White House—the political overseers of the agency—to demand a pro-industry outcome.”).

293. Rachel E. Barkow, *Insulating Agencies: Avoiding Capture Through Institutional Design*, 89 TEXAS L. REV. 15, 23 (2010) (“[C]apture operates because of the revolving-door phenomenon: the heads of agencies often anticipate entering or returning to employment with the regulated industry once their government service terminates.”).

294. See Seidenfeld, *supra* note 292, at 643 (“[Agency s]taff members . . . may harbor an idiosyncratic understanding of [the agency’s] mission.”).

295. See *supra* notes 139–45 and accompanying text.

296. See, e.g., David B. Spence & Frank Cross, *A Public Choice Case for the Administrative State*, 89 GEO. L.J. 97, 119 (2000) (concluding that “[t]he agency drift, or tunnel vision, argument . . . overstates the magnitude and significance of the drift problem”); *id.* at 121–22 (“[A]gency capture is no longer regarded as a valid descriptive theory of bureaucratic behavior.”); Matthew C. Stephenson, *Public Regulation of Private Enforcement: The Case for Expanding the Role of Administrative Agencies*, 91 VA. L. REV. 93, 131 (2005) (“[R]ecent research suggests that the ‘agency capture’ problem has been wildly overstated.”).

297. See Barkow, *supra* note 293, at 46 (“The effect of the revolving door is often cited as one of the reasons why the SEC failed to adequately protect consumers by addressing pressing problems in the trading industry.”).

298. See Joel A. Mintz, *Has Industry Captured the EPA?: Appraising Marver Bernstein’s Captive Agency Theory After Fifty Years*, 17 FORDHAM ENVTL. L. REV. 1, 36 (2005) (indicating that “EPA’s enforcement work has been nearly captured by industry several times and that it was

Drug Administration²⁹⁹ might be prone to capture by regulated interests, they have been entrusted with substantive rulemaking responsibilities that seem at least as important as any the USPTO would have under a grant of substantive rulemaking authority with respect to questions of patentable subject matter.³⁰⁰

Nonetheless, for at least three reasons, capture or bias concerns might be unusually sharp with respect to the USPTO. First, high degrees of uncertainty about the overall social value or harm generated by any particular patent grant and the typical lack of public salience of such grants mean that public-interest concerns are perhaps an even weaker check on the interests of prospective and existing patentees than in a context such as environmental regulation, in which disastrous events can lead to a focused public outcry.³⁰¹ Additionally, in the environmental context, suits for public nuisance or other forms of tort can check tendencies toward agency capture by helping to fill or highlight gaps in the agency's web of regulation.³⁰²

A second special concern derives from the nature of the U.S. patent bar. Although the mass of attorneys litigating patent cases has become substantially diverse,³⁰³ the USPTO still works with a special bar of patent prosecutors who have passed a separate "patent bar" examination and who, even to qualify to take this examination, need either to have completed sufficient coursework in an approved field or to have passed a further special test on engineering fundamentals.³⁰⁴ To the extent relatively clear limitations

partially captured on one occasion"); Wagner, *supra* note 290, at 1378 (describing reasons to think "that a significant share of EPA rulemakings might be susceptible to information capture").

299. See Merrill Goozner, *Conflicts of Interest in the Drug Industry's Relationship with the Government*, 35 HOFSTRA L. REV. 737, 738–39 (2006) (alleging that "industry capture" has turned "the nation's food and drug watchdog" into "an under funded lapdog"); Jason Iuliano, *Killing Us Sweetly: How to Take Industry Out of the FDA*, 6 J. FOOD L. & POL'Y 31, 33 (2010) (contending that the FDA "is nothing more than a corporate lapdog").

300. Masur, *supra* note 7, at 6–7 (arguing that, "[u]nless one believes that the administrative state should be dismantled wholesale, there is no compelling reason to resist granting substantive rule-making authority to the [US]PTO").

301. Cf. Lynn E. Blais & Wendy E. Wagner, *Emerging Science, Adaptive Regulation, and the Problem of Rulemaking Ruts*, 86 TEXAS L. REV. 1701, 1712 (2008) ("[W]hen new information is accessible, salient, and of the sort that can unify broad public concern about an issue, institutional responses can be swift and dramatic.").

302. See Mark Seidenfeld & Janna Satz Nugent, *"The Friendship of the People": Citizen Participation in Environmental Enforcement*, 73 GEO. WASH. L. REV. 269, 295 (2005) ("The scenario of cozy relationships between companies and [environmental] regulators is in part what drives the need for public interest group involvement in the enforcement process.").

303. See Golden, *supra* note 11, at 684 ("Nowadays, there is no monolithic patent bar to speak of.").

304. See U.S. Patent & Trademark Office, General Requirements Bulletin for Admission to the Examination for Registration to Practice in Patent Cases Before the United States Patent and Trademark Office 3–8 (Rev. 2008) (listing eligibility requirements for taking the patent bar, including a bachelor's degree or adequate coursework in specified subjects, or passage of "the Fundamentals of Engineering (FE) test").

on patentable subject matter are expected to reduce business for patent prosecutors, members of this special bar might consistently oppose such limitations. Such opposition might exert a strong gravitational pull on an administrative agency that works nearly exclusively with such lawyers and agents in adjudicative contexts, and many of whose personnel might contemplate a later career in patent prosecution.

A third possible cause for special concern derives from the USPTO's funding situation. To the extent funding for the USPTO is, in effect, capped by fees that it collects from patent applicants and owners,³⁰⁵ the USPTO might have an unusually strong administrative interest in preserving or increasing its "customer base" through a broad view of patentable subject matter.

An initial response to such concerns could be that, even with substantive rulemaking authority on issues of patentable subject matter, the USPTO's discretion is likely to be relatively tightly confined. As noted earlier, the United States' treaty obligations under TRIPS, which Congress might explicitly charge the USPTO with following, already require an expansive approach to subject-matter eligibility, one that generally does not discriminate against any technological field in awarding patent rights.³⁰⁶ Thus, to the extent bias or capture is a concern, it should be recalled, for starters, that there will only be sharply limited room for USPTO bias or capture to operate. The proposal here would give the USPTO only rulemaking authority on questions of patentable subject matter, and the range of potential ways in which such questions could be resolved is already relatively strictly confined.

In any event, recent history suggests that concerns of USPTO capture and bias are overstated with respect to issues of patentable subject matter. Although such concerns are generally presented as reasons to believe the USPTO will be overly permissive of subject matter, the USPTO has in fact commonly led fights for subject-matter limits. All of the Supreme Court's subject-matter eligibility decisions in the 1970s and early 1980s responded to proceedings in which the USPTO had rejected claimed subject matter as unpatentable. In *Gottschalk v. Benson* and *Parker v. Flook*, the Supreme Court upheld USPTO rejections.³⁰⁷ In *Diamond v. Chakrabarty* and

305. See Rai, *supra* note 4, at 2062 ("[T]he PTO is an entirely fee-based organization.").

306. See *supra* text accompanying note 145. The USPTO's 2010 guidelines on patentable subject matter themselves highlight the relatively limited field in which current debates over subject matter exclusions operate. See Interim Guidance for Determining Subject Matter Eligibility for Process Claims in View of *Bilski v. Kappos*, 75 Fed. Reg. 43,922, 43,923 (July 27, 2010) ("Although this guidance presents a change in existing examination practice, it is anticipated that subject matter eligibility determinations will not increase in complexity for the large majority of examiners, who do not routinely encounter claims that implicate the abstract idea exception.").

307. See *Parker v. Flook*, 437 U.S. 584, 585–87, 594 (1978) (agreeing with the USPTO's rejection of claims for a method that used a mathematical algorithm to calculate a threshold value

Diamond v. Diehr, the Supreme Court rejected USPTO efforts to enforce subject-matter limits.³⁰⁸ Likewise, the Federal Circuit's major en banc opinion on patentable subject matter in the 1990s, *In re Alappat*,³⁰⁹ reversed USPTO rejections.³¹⁰ Hence, in the last several decades, the U.S. courts, not the USPTO, seem primarily responsible for the capacious view of patentable subject matter that the courts have recently moved to rein in. In this area at least, there might be cause for belief that the courts, including a semispecialized Federal Circuit and a Supreme Court closely watched by major business interests, are more vulnerable to "capture" than the USPTO.³¹¹

Indeed, the USPTO faces a variety of pressures that do not align strictly with applicant interests in broader conceptions of patentable subject matter. First, there is the long-standing problem of the USPTO's backlog of pending applications and associated problems of delays in examination. These problems consistently provide cause for criticizing USPTO performance, and the USPTO has vigorously sought ways to address them.³¹² One obvious way to eliminate part of the USPTO backlog or, at least, to limit the flow of new applications adding to that backlog would be to take a stringent view of patentable subject matter.

A second, self-interested basis for USPTO reluctance to embrace doubtful forms of patentable subject matter arises from administrative difficulties that efforts to process such matter tend to create, including a likely greater risk of near-term embarrassment from the granting of poor-

above which an alarm would be triggered); *Gottschalk v. Benson*, 409 U.S. 63, 64–65 (1972) (agreeing with the USPTO's rejection of claims for "a method of programming a general-purpose digital computer to convert signals from binary-coded decimal form into pure binary form").

308. See *Diamond v. Diehr*, 450 U.S. 175, 180–81, 184 (1981) (disagreeing with the USPTO's rejection of claims for a method using a mathematical formula to determine when to halt a curing of rubber); *Diamond v. Chakrabarty*, 447 U.S. 303, 306, 309 (1980) (disagreeing with the USPTO's rejection of claims for genetically engineered bacteria, which had been based on the ground that "living things . . . are not patentable subject matter").

309. 33 F.3d 1526 (Fed. Cir. 1994) (en banc).

310. See *id.* at 1537, 1539–42 (disagreeing with the USPTO's rejection of claims to "a means for creating a smooth waveform display in a digital oscilloscope").

311. See Golden, *supra* note 11, at 689 ("[T]he existence of strong lobbying for 'pro-business' justices and the rise of an apparently influential Supreme Court bar suggest that concerns of bias or capture may be at least as great with respect to the Supreme Court as with respect to the Federal Circuit.").

312. See Arti Rai, Stuart Graham, & Mark Doms, Patent Reform 5 (Apr. 13, 2010) (unpublished white paper), available at http://www.commerce.gov/sites/default/files/documents/migrated/Patent_Reform-paper.pdf ("As regards timeliness, the problem is quite acute. The USPTO currently has a backlog of more than 750,000 patent applications, an accumulation that has doubled over the last decade."); see also Tony Dutra, *Attacking Patent Backlog, Pendency Problem Is Overriding Theme of 21st Annual PTO Day*, 81 PAT., TRADEMARK & COPYRIGHT J. 217, 217 (2010) ("The patent application backlog and the Patent and Trademark Office's plan to reduce it . . . dominated much of the discussion . . . at the 21st annual 'PTO Day' conference in Washington, D.C.").

quality patents. When the USPTO argued against the subject-matter eligibility of various software-related innovations in the 1970s and early 1980s, one plausible motivation, reflected in an opinion by Justice Stevens, was concern about the USPTO's ability to evaluate the quality of such subject matter competently.³¹³ Extension of patent rights to a new field of technology predictably places an immediate stress on an examination corps that will likely lack sufficient personnel having significant experience with the new technology. Without much in the way of preexisting patent material to consult as prior art, even qualified examiners are likely to have unusual difficulty discovering the most relevant art for purposes of evaluating questions of novelty or nonobviousness.

Third, as far as USPTO concerns with funding and "customer satisfaction" go, it is important to recall that extension of patent rights to new forms of subject matter does not necessarily favor all classes of innovators. Innovators who develop products or processes that are downstream of, or complementary to, innovations within a questionable subject-matter class might rationally believe that permitting patents for such subject matter will tend to force greater profit sharing with a new class of patent holders.³¹⁴ Similarly, supposed USPTO interest in more patents, which in turn generate more USPTO fees, does not necessarily favor subject-matter expansion. When the granting of patents for a particular form of subject matter generates a net drag on innovation, the USPTO might expect that the result of less innovation will be fewer patents.³¹⁵

Finally, the institutional framework in which the USPTO operates is not fixed and can be altered to reduce concerns of capture and bias. Even without new institutional reforms, non-patent-holding interest groups have recently become active in supervising USPTO processes—whether through the launching of *ex parte* or *inter partes* reexamination proceedings³¹⁶ or

313. See *Diehr*, 450 U.S. at 218 (Stevens, J., dissenting) ("No doubt [Commissioners of Patents and Trademarks who had opposed patents for specific computer-related innovations] may have been motivated by a concern about the ability of the Patent Office to process effectively the flood of applications that would inevitably flow from a decision that computer programs are patentable.").

314. Recent debates over patent reform have illustrated how business interests can split between those who advocate greater or lesser patent rights. See Golden, *supra* note 16, at 507 (discussing lobbying relating to recent patent-reform bills); see also Michael W. Carroll, *One Size Does Not Fit All: A Framework for Tailoring Intellectual Property Rights*, 70 OHIO ST. L.J. 1361, 1431 (2009) ("As long as a substantial subset of industry players is more concerned about access to inputs than control over outputs, the public choice risks associated with a more [intellectual-property-right-]tailoring-friendly legislative environment are lower than advertised.").

315. See John M. Golden, *Innovation Dynamics, Patents, and Dynamic-Elasticity Tests for the Promotion of Progress*, 24 HARV. J.L. & TECH. 47, 99 (2010) (discussing circumstances under which patent rights might be particularly likely to lead to a net slowing of innovation).

316. See Golden, *supra* note 49, at 320 (discussing requests for reexamination filed by Consumer Watchdog and the Public Patent Foundation).

through the challenge of patent claims' validity in federal courts.³¹⁷ Such active involvement of non-patentees would likely be further promoted through the institution of European-style opposition proceedings³¹⁸ or, at least, expanded opportunities for U.S.-style reexamination that encompass questions of subject-matter eligibility and enable non-patent-applicants to challenge adverse decisions both in the USPTO and in the courts.³¹⁹

A further institutional provision that could help protect against institutionally conditioned bias would be a congressional requirement that USPTO exercises of substantive rulemaking authority occur through notice-and-comment rulemaking processes rather than through adjudication.³²⁰ Under the Administrative Procedure Act (APA), notice-and-comment rulemaking is, by nature, open for involvement by any "interested person[.]"³²¹ Further, to the extent ensuring adequate representation of non-patentee interests is a concern, Congress could require that USPTO substantive rules be vetted, either during or prior to notice-and-comment rulemaking, by an advisory board specifically required to contain members who represent non-patentee interests.³²² For example, such a board could be required to include representatives from the FTC, the Department of Energy, the National Institutes of Health, academic science, and a consumer interest group. Rather than merely silently approving or disapproving of a proposed rule, the board could be required to provide comments, potentially majority

317. See *Ass'n for Molecular Pathology v. U.S. Patent & Trademark Office*, 702 F. Supp. 2d 181, 186–88 (S.D.N.Y. 2010) (listing as plaintiffs the Association for Molecular Pathology, the American College of Medical Genetics, the American Society for Clinical Pathology, Breast Cancer Action, and others).

318. See Mark D. Janis, *Rethinking Reexamination: Toward a Viable Administrative Revocation System for U.S. Patent Law*, 11 HARV. J.L. & TECH. 1, 99 (1997) ("Major patent jurisdictions in Europe have had long experience with administrative review procedures—oppositions, revocations, and 'nullity' proceedings—for adjudicating patent validity.").

319. Cf. Jonathan Masur, *Patent Inflation* 29 (Univ. of Chi. Pub. Law & Legal Theory, Working Paper No. 316, 2010), available at <http://ssrn.com/abstract=1623929> (arguing for the general importance, as a matter of institutional design, of permitting "challengers who lose before the [US]PTO to appeal to the Federal Circuit").

320. There are, in fact, likely a number of good reasons for favoring rulemaking over adjudication in this context. Richard Pierce, Sidney Shapiro, and Paul Verkuil have reported "near universal judicial and scholarly criticism of agency use of adjudication as a vehicle for formulating general rules . . . based on considerations that can be placed in four general categories: quality, efficiency, fairness and political accountability." PIERCE ET AL., *supra* note 78, at 280. Particularly given the USPTO's status as "a high caseload" agency with several thousand examiners separately adjudicating questions of patentability, rulemaking might be the better "vehicle for responsible officials to exercise control over policies." Daniel J. Gifford, *Adjudication in Independent Tribunals: The Role of an Alternative Agency Structure*, 66 NOTRE DAME L. REV. 965, 1023 (1991).

321. 5 U.S.C. § 553(c) (2006) ("After notice required by this section, the agency shall give interested persons an opportunity to participate in the rule making through submission of written data, views, or arguments with or without opportunity for oral presentation.").

322. Cf. 42 U.S.C. § 7409(d) (2006) (providing for "an independent scientific review committee" to review the EPA's air-pollutant criteria and air-quality standards).

and minority comments, that would become part of the rulemaking record, and thus a potential basis for judicial review if contrary to the rule ultimately adopted.³²³ Such review and comment would supplement Office of Management and Budget review, pursuant to executive order, of USPTO rulemaking.³²⁴

Congress could also require that the personnel responsible for USPTO substantive rulemaking be substantially insulated from those having responsibility for USPTO adjudication. The APA and other acts already provide models for how adjudicators, such as administrative law judges, can be insulated from undue influence by other agency personnel.³²⁵ Congress could similarly provide for substantial autonomy of USPTO rulemaking personnel, including budgetary independence.³²⁶

Of course, Congress could accomplish the same end by housing substantive rulemaking authority in a separate agency entirely, perhaps a newly formed “Office of Innovation Policy”³²⁷ having at least this dash of real lawmaking muscle. A rough model for such an approach exists under the Occupational Safety and Health Act. This Act establishes two separate entities, the Occupational Safety and Health Review Commission and the Occupational Safety and Health Administration, that perform (1) adjudicatory functions and (2) rulemaking and rule-enforcement functions, respectively.³²⁸

323. PIERCE ET AL., *supra* note 78, at 326 (“An agency’s failure to respond in its statement of basis and purpose to well-supported arguments contained in public comments critical of the agency’s proposed rule can form the basis for reversal of the agency’s final rule.” (citation omitted)).

324. See generally MANNING & STEPHENSON, *supra* note 279, at 550–53 (discussing the development of “centralized regulatory review” through successive executive orders).

325. See 5 U.S.C. § 554(d) (generally requiring that an “employee who presides at the reception of evidence” in formal adjudication strictly limit *ex parte* contacts with respect to “fact[s] in issue” and “[not] be responsible to or subject to the supervision or direction of an employee or agent engaged in . . . investigative or prosecuting functions”); *id.* § 557(d)(1) (further limiting *ex parte* communications with agency decision makers); see also PIERCE ET AL., *supra* note 78, at 322 (describing administrative law judges as “almost entirely independent of the agencies at which they preside”).

326. See Burstein, *supra* note 7 (manuscript at 59) (calling for USPTO rulemaking and examination functions to have “separate budgets and personnel policies”).

327. Benjamin & Rai, *supra* note 249, at 61 (“We ultimately propose . . . an Office of Innovation Policy (‘OIP’) that would draw upon, and feed into, existing regulatory review processes . . .”).

328. See STRAUSS ET AL., *supra* note 241, at 403 (“As with OSHA and OSHRC, Congress has occasionally created a bureaucratic structure explicitly separating quasi-legislative and quasi-executive from quasi-judicial responsibilities.”); see also *About OSHA*, U.S. DEP’T OF LABOR, <http://osha.gov/about.html> (describing OSHA’s congressionally mandated mission as “ensur[ing] safe and healthful working conditions for working men and women by setting and enforcing standards and by providing training, outreach, education and assistance”); U.S. OCCUPATIONAL SAFETY & HEALTH REV. COMM’N, <http://www.oshrc.gov> (describing OSHRC as “an independent Federal agency created to decide contests of citations or penalties resulting from OSHA inspections of American work places”).

But such a split-agency approach seems likely to represent overkill. Both the narrowness of the proposed rulemaking authority and doubts about the magnitude of capture or bias concerns suggest that housing such authority in a separate agency will likely cost more than it is worth. Prior split-agency approaches have led Richard Fallon to note the “variety of perplexities” in interagency relations that result.³²⁹

Moreover, there will likely be situation-specific costs to multiplying institutional walls between USPTO rulemakers and examiners. USPTO examinations will likely provide important information about (1) pressure points in subject-matter law, (2) patent-applicant claiming strategies, and (3) the meaningfulness of subject-matter limitations in actual practice.³³⁰ Given applicants’ freedom to craft claims to conform, at least formally, with doctrinal demands, and given limitations on the time and resources the USPTO has available to provide anything approaching timely patent examination, careful monitoring of examination processes will almost surely be necessary to ensure that the law in action bears significant relation to issued regulations’ intended effects. Further, as occurred dramatically with the flood of applications for gene fragments filed in the 1990s,³³¹ information about incoming applications and ongoing but nonpublic USPTO examinations can provide timely notice of new pressure points. Such early warning can enable rulemakers to study an issue and to propose a response before the “lock-in” effects from too many ill-considered grants become overwhelming. Placing rulemaking authority in an administrative entity that is entirely separate from the USPTO can be expected to complicate, if not fully frustrate, efforts at cooperation.

Recent USPTO initiatives on subject-matter eligibility demonstrate both the agency’s already-existing rulemaking capacity, and its ability to act quickly (a) to clarify understandings of existing law, (b) to solicit broad-based input, and (c) to address detailed concerns that frequently arise in claim drafting and examination. For example, in late January 2010, USPTO Director David Kappos used a one-page memorandum to patent examiners to provide instruction on how claims for an invention using a computer-readable medium might be narrowed to avoid encompassing transitory electromagnetic signals that the Federal Circuit has held to be

329. Richard H. Fallon, Jr., *Enforcing Aviation Safety Regulations: The Case for a Split-Enforcement Model of Agency Adjudication*, 4 ADMIN. L.J. 389, 431 (1991); see also George Robert Johnson, Jr., *The Split-Enforcement Model: Some Conclusions from the OSHA and MSHA Experiences*, 39 ADMIN. L. REV. 315, 348 (1987) (“Divided regulatory programs must be expected inherently to encounter more administrative problems and difficulties than might a similar program housed entirely under one administrative roof.”).

330. Cf. Burstein, *supra* note 7 (manuscript at 59) (arguing that “there is likely to be significant value in keeping [rulemaking and examination] functions housed in the same agency” because, for example, examination can bring to light “issues deserving of policy attention”).

331. See MERGES & DUFFY, *supra* note 50, at 252 (discussing a “race to patent . . . gene fragments” in the 1990s).

unpatentable.³³² Citing an earlier memorandum from 1987 “suggesting that applicants add the limitation ‘non-human’ to a claim covering a multicellular organism,” the Director indicated that “adding the limitation ‘non-transitory’” might prevent a subject-matter rejection while also preserving patentability under written-description and enablement requirements.³³³ This suggestion illustrates the kind of attention to malleable claim-drafting detail that will likely be crucial to socially desirable regulation of patentable subject matter.

More dramatically, on the very day that the Supreme Court issued its long-awaited decision in *Bilski v. Kappos*, the USPTO issued a two-page memorandum, made publicly available via the USPTO’s website, that instructed examiners how to recalibrate evaluation of subject-matter eligibility in light of the decision.³³⁴ In the wake of the Court’s holding that the machine-or-transformation test was “‘a useful and important clue’” but not a definitive requirement,³³⁵ the USPTO instructed examiners to use a branched set of rebuttable presumptions:

If a claimed method meets the machine-or-transformation test, the method is likely patent-eligible under section 101 unless there is a clear indication that the method is directed to an abstract idea. If a claimed method does not meet the machine-or-transformation test, the examiner should reject the claim under section 101 unless there is a clear indication that the method is not directed to an abstract idea.³³⁶

Significantly, however, the USPTO’s same-day memorandum also emphasized that it provided only “interim guidance”³³⁷ and that the agency would “be developing further guidance.”³³⁸ Indeed, on July 27, 2010, less than one month after the Supreme Court’s decision in *Bilski*, the USPTO published significantly more detailed guidelines in the *Federal Register* and solicited public comment.³³⁹ These guidelines apparently abandoned the

332. Memorandum from David J. Kappos, Director of the U.S. Patent & Trademark Office, Subject Matter Eligibility of Computer Readable Media (Jan. 26, 2010), available at http://www.uspto.gov/patents/law/notices/101_crm_20100127.pdf.

333. *Id.*

334. Memorandum to Patent Examining Corps, *supra* note 232; see also Press Release, U.S. Patent & Trademark Office, USPTO Issues Statement in Response to Supreme Court Ruling in *Bilski v. Kappos* (June 28, 2010), available at http://www.uspto.gov/news/pr/2010/10_30.jsp (announcing that the USPTO was “distributing interim guidance for the examining corps” that same day (internal quotation marks omitted)).

335. Memorandum to Patent Examining Corps, *supra* note 232, at 1.

336. *Id.* at 2.

337. *Id.* at 1.

338. *Id.* at 2.

339. Interim Guidance for Determining Subject Matter Eligibility for Process Claims in View of *Bilski v. Kappos*, 75 Fed. Reg. 43,922 (July 27, 2010).

presumption-based approach of the same-day instructions.³⁴⁰ Instead of basing presumptions of eligibility or ineligibility on satisfaction or failure, respectively, of the machine-or-transformation test—i.e., instead of “presuming” to make presumptions—the USPTO more humbly invited responses to the following questions:

1. What are examples of claims that do not meet the machine-or-transformation test but nevertheless remain patent-eligible because they do not recite an abstract idea?
2. What are examples of claims that meet the machine-or-transformation test but nevertheless are not patent-eligible because they recite an abstract idea?³⁴¹

The guidelines then gave an overview of the Supreme Court’s decisions in *Bilski* and on related petitions for certiorari,³⁴² reproducing the claims rejected in *Bilski* “as examples of claims that run afoul of the abstract idea exception.”³⁴³

In a further effort to clarify how, as a matter of day-to-day practice, examiners and patent prosecutors should approach questions of subject-matter eligibility, the guidelines also provided a multicolumn description of factors and subfactors that could affect a determination of subject-matter eligibility.³⁴⁴ The list included (1) various factors relating to the machine-or-transformation test; (2) additional factors relating to whether a claim would effectively “monopolize a natural force,” “scientific fact,” or ways of thinking about them; and (3) additional factors relating to whether a claim would effectively monopolize “a general concept.”³⁴⁵ The guidelines also provided a nonexclusive list of “[e]xamples of general concepts”; such examples included “[b]asic economic practices or theories (e.g., hedging, insurance, financial transactions, marketing)”; “[b]asic legal theories (e.g., contracts, dispute resolution, rules of law)”; “[m]athematical concepts (e.g., algorithms, spatial relationships, geometry)”; and “[m]ental activity (e.g., forming a judgment, observation, evaluation, or opinion).”³⁴⁶

The USPTO’s guidelines in the wake of *Bilski* are a work in progress, with many aspects that cry for critique and possible amendment. But whatever their flaws, they demonstrate (1) the progress that the USPTO can make, in a relatively short time, in articulating to examiners and the public an

340. *But see id.* at 43,923 (describing the new interim guidance “as a supplement” to “the memorandum to the Patent Examining Corps . . . dated June 28, 2010”).

341. *Id.*

342. *Id.* at 43,923–25.

343. *Id.* at 43,924.

344. *Id.* at 43,925–26.

345. *Id.*

346. *Id.* at 43,926.

which the USPTO will likely seek to make such progress. In combination with the USPTO's June 28 memorandum, the interim guidelines illustrate the USPTO's ability to act quickly and flexibly, and in a way that actively solicits input and can generate improvement over time.

Likewise, recent USPTO successes with purely interpretive rulemaking on matters of substance fortify this Article's contention that the USPTO can add value by engaging in binding, rather than merely interpretive, rulemaking on subject-matter eligibility. The USPTO's 2010 guidelines on subject-matter eligibility are merely one illustration of the USPTO's regular engagement in notice-and-comment processes involving revised instructions to examiners on questions of substantive patent law. Indeed, the 2010 guidelines marked the third time in five years that the USPTO had engaged in notice-and-comment processes with respect to questions of subject-matter eligibility: in December 2005³⁴⁷ and September 2009,³⁴⁸ the USPTO likewise published requests for comments on interim guidelines on subject-matter eligibility.³⁴⁹ Within the last thirteen years, the USPTO has similarly invited comments on new guidelines with respect to other requirements for patentability, including guidelines on utility³⁵⁰ and an adequate written description.³⁵¹ Notably, the USPTO's requests for comments on the utility and written-description guidelines attracted a broad range of commentators, including the American Intellectual Property Law Association, the American College of Medical Genetics, the Association of American Medical Colleges, the Biotechnology Industry Organization, the National Breast Cancer Coalition, the National Institutes of Health, and various individuals.³⁵² Likewise, the USPTO's request for comments on its 2005 subject-matter

347. Request for Comments on Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility, 70 Fed. Reg. 75,451 (Dec. 20, 2005).

348. Request for Comments on Interim Examination Instructions for Evaluating Patent Subject Matter Eligibility, 74 Fed. Reg. 47,780 (Sept. 17, 2009).

349. In both instances, the USPTO later published notices that extended the deadline for submitting comments. See Additional Period for Comments on Interim Examination Instructions for Evaluating Patent Subject Matter Eligibility, 74 Fed. Reg. 52,184 (Oct. 9, 2009); Request for Comments on Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility, 71 Fed. Reg. 34,307 (June 14, 2006).

350. Revised Interim Utility Examination Guidelines; Request for Comments; Correction, 65 Fed. Reg. 3,425 (Jan. 21, 2000); Revised Utility Examination Guidelines; Request for Comments, 64 Fed. Reg. 71,440 (Dec. 21, 1999).

351. Revised Interim Guidelines for Examination of Patent Applications Under the 35 U.S.C. § 112, ¶ 1 "Written Description" Requirement; Request for Comments, 64 Fed. Reg. 71,427 (Dec. 21, 1999); Request for Comments on Interim Guidelines for Examination of Patent Applications Under the 35 U.S.C. § 112 ¶ 1 "Written Description" Requirement, 63 Fed. Reg. 32,639 (June 15, 1998).

352. U.S. Patent & Trademark Office, Public Comments on the United States Patent and Trademark Office "Revised Interim Utility Examination Guidelines," <http://www.uspto.gov/web/offices/com/sol/comments/utilguide/index.html>; Comments on Interim Guidelines for Examination of Patent Applications under the 35 U.S.C. § 112 "Written Description" Requirement, <http://www.uspto.gov/ip/rules/comments/com112.jsp>.

guidelines drew responses from the American Bar Association, the American Intellectual Property Law Association, the Association of American Medical Colleges, the Computer and Communications Industry Association, the Motion Picture Association of America, the National Institutes of Health, Patients Not Patents, and individuals.³⁵³

Moreover, the final versions of the USPTO's utility and written-description guidelines have had substantial legal effect.³⁵⁴ In 2005, the Federal Circuit essentially gave its blessing to the new utility guidelines,³⁵⁵ which had effectively made unpatentable a substantial category of claims for patents on genetic fragments called "expressed sequence tags."³⁵⁶ Even before this, in 2002, the Federal Circuit had found persuasive and "adopt[ed]" the written-description guidelines' discussion of circumstances under which "functional descriptions of genetic material" or other chemical compounds would satisfy patent law's written description requirement.³⁵⁷

In short, even without substantive rulemaking authority, the USPTO has demonstrated substantial competence—at least as adjudged by its primary institutional competitor, the Federal Circuit—in handling thorny questions of substantive patent law. But interpretive authority bound to preexisting judicial precedent is unlikely to enable the USPTO to help clean up the mess that is subject-matter-eligibility law. Given the tangled state of existing judge-

353. U.S. Patent & Trademark Office, Comments Regarding Revisions and Technical Corrections Affecting Requirements for Ex Parte and Inter Partes Reexamination, <http://www.uspto.gov/ip/rules/comments/ab98.jsp>. The list of commentators on the USPTO's 2009 guidelines appears to have been less balanced, see U.S. Patent & Trademark Office, Comments on Request for Comments on Interim Examination Instructions for Evaluating Patent Subject Matter Eligibility (Sept. 17, 2009), <http://www.uspto.gov/patents/law/comments/ab98comments.jsp>, perhaps because more resource-strapped parties were directing their attention more exclusively to the U.S. Supreme Court case of *Bilski v. Kappos*, which already awaited oral argument and threatened to render much of the USPTO's 2009 guidelines moot.

354. See Dutra, *supra* note 177, at 325 (reporting former Federal Circuit Chief Judge Michel's statement that the Federal Circuit had, "[i]n some . . . decisions, . . . explicitly approved the [USPTO's 2001] written description guidelines as being valid and worth deferring to"); Rai, *supra* note 4, at 2053 ("In the 2005 case *In re Fisher*, the PTO succeeded in convincing the Federal Circuit of the validity of its heightened standard for evaluating the utility of patent applications." (footnote omitted)).

355. *In re Fisher*, 421 F.3d 1365, 1372 (Fed. Cir. 2005) ("The PTO's standards for assessing whether a claimed invention has a specific and substantial utility comport with this court's interpretation of the utility requirement of § 101.").

356. See *id.* at 1379 ("We conclude that substantial evidence supports the [BPIA's] findings that each of the five claimed [expressed sequence tags] lacks [] a specific and substantial utility and that they are not enabled."); see also John M. Golden, *Biotechnology, Technology Policy, and Patentability: Natural Products and Invention in the American System*, 50 EMORY L.J. 101, 188 (2001) (describing comments associated with the guidelines as "mak[ing] clear that the [USPTO] considers bare genetic sequences unpatentable").

357. *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956, 964 (Fed. Cir. 2002); cf. *Ariad Pharms., Inc. v. Eli Lilly & Co.*, 598 F.3d 1336, 1352–53 (Fed. Cir. 2010) (en banc) (citing the USPTO's 2001 guidelines in support of a statement regarding the nature of the written-description requirement "when a patent claims a genus by its function or result").

made doctrine, and given the ease with which new technologies and claim-drafting techniques generate further tangles, the USPTO will likely need more freedom of operation than merely subordinate interpretive authority allows.

IV. Conclusion

Limitations on subject-matter eligibility provide a categorical filter that can improve patent-system performance. Two mathematical models illustrate (1) how more individualized tests for patentability can fail to ensure that patents improve social welfare and (2) how a particular form of subject matter—fundamental principles having a high number of potential kinds of uses—can generate particularly high social costs and thus qualify as a form of subject matter that the patent system would best filter out.

The question then becomes what form of process and what institution the law should charge with developing and maintaining such a categorical filter. This Article argues that, in the United States at least, administrative rulemaking appears to be the best process for this purpose. Additionally, the best agency to carry out such rulemaking is the USPTO. The USPTO has a demonstrated capacity and known incentive to respond promptly and meaningfully to questions of subject-matter eligibility. Moreover, giving the USPTO rulemaking authority with respect to subject-matter eligibility does not require giving it plenary rulemaking authority that encompasses all matters of patent-law substance. Just as other regimes of U.S. law have divided tasks of adjudication and enforcement between different institutions, the U.S. patent system can divide areas of primary interpretive authority between the USPTO and Article III courts. Such an institutional innovation appears the best way to leverage the relative competences of these institutions.

* * *

Share Repurchases, Equity Issuances, and the Optimal Design of Executive Pay

Jesse M. Fried*

This Article identifies a potential cost to public investors of tying executive pay to the stock's future value—even its long-term value. In particular, such an arrangement can incentivize executives to engage in share repurchases (when the current stock price is low) and equity issuances (when the current stock price is high) that reduce “aggregate shareholder value”: the amount of value flowing to all the firm's shareholders over time. The Article also puts forward a mechanism that ties executive pay to aggregate shareholder value and thereby eliminates the identified distortions.

I. Introduction

Public-company executives in the United States receive most of their pay in the form of equity compensation—restricted stock, stock options, and other incentives whose payoff is tied to the future value of their firms' shares.¹ Among S&P 500 CEOs in 2009, on average more than 60% of compensation came in the form of restricted stock and stock-option grants.² Equity-based compensation is increasingly common in other countries as well.³

The purpose of equity compensation is to better align executives' interests with those of the firm's shareholders.⁴ Tying executives' payoffs more closely to the stock's future value should give executives stronger in-

* Professor of Law, Harvard Law School. For financial support, I am grateful to the John M. Olin Center for Law, Economics, and Business and the Harvard Law School Program on Corporate Governance. Elaine Choi, Daniel Doktor, Shawn Grover, Matt Hutchins, Audrey Lee, Da Lin, and Katherine Petti provided valuable research assistance. For helpful conversations and comments, I would like to thank Ken Ayotte, Joe Bachelder, Lucian Bebchuk, Bernie Black, Tom Brennan, Allen Ferrell, Victor Fleischer, Rob Jackson, Ira Kay, James Kim, Reinier Kraakman, Claudia Landeo, Kate Litvak, Mark Ramseyer, Guhan Subramanian, Fred Tung, participants at the 2010 ALEA Meeting, the Harvard Law and Economics Seminar, and the Northwestern Law and Economics Colloquium, and especially Louis Kaplow and Steve Shavell.

1. John E. Core et al., *Executive Equity Compensation and Incentives: A Survey*, FRBNY ECON. POL'Y REV., Apr. 2003, at 29.

2. See EQUILAR, 2010 CEO PAY STRATEGIES: COMPENSATION AT S&P 500 COMPANIES 11 (2010).

3. See Brian R. Cheffins & Randall S. Thomas, *The Globalization (Americanization?) of Executive Pay*, 1 BERKELEY BUS. L.J. 233, 246–47 (2004) (describing the increasing use of equity-based pay in Europe and Asia).

4. Michael C. Jensen & Kevin J. Murphy, *CEO Incentives—It's Not How Much You Pay, But How*, HARV. BUS. REV., May–June 1990, at 138, 139.

centives to generate value for shareholders.⁵ Not surprisingly, the use of equity-based pay has long been encouraged by investors, regulators, and academics.⁶

Most compensation arrangements tie executive pay to the *short-term* stock price.⁷ Unfortunately, a stock's short-term price does not necessarily reflect the stock's long-term value. Thus, over the last decade, there has been growing recognition that tying pay to the short-term stock price encourages executives to focus on short-term results at the expense of long-term value.⁸ But recognition of the problems associated with tying pay to the short-term stock price has not diminished enthusiasm for the use of equity compensation itself; rather, it has led commentators to emphasize the importance of tying equity payoffs to the *long-term* stock price (which better reflects the stock's long-term value).⁹

This Article identifies a different and more subtle economic problem with equity pay—a problem that arises whenever an executive's payoff is tied to the stock's future value, even its long-term value. In particular, the Article shows that tying payoffs to the stock's future value fails to reward executives for maximizing what I have called “aggregate shareholder value”: the amount of value flowing to all of the firm's shareholders over time.¹⁰ Indeed, I show that tying an executive's payoff to the stock's future value, even its long-term value, can encourage the executive to take steps that

5. *Id.*

6. See, e.g., Jeffrey N. Gordon, “Say on Pay”: *Cautionary Notes on the U.K. Experience and the Case for Shareholder Opt-In*, 46 HARV. J. ON LEGIS. 323, 352 (2009) (reporting that the widespread adoption of stock options in the 1990s resulted, in part, from institutional investor pressure on firms); Jensen & Murphy, *supra* note 4, at 141 (urging boards to use more stock options to better tie equity pay to performance).

7. See LUCIAN BEBCHUK & JESSE FRIED, PAY WITHOUT PERFORMANCE: THE UNFULFILLED PROMISE OF EXECUTIVE COMPENSATION 176–77 (2004) (describing executives' ability to unwind stock options and restricted shares as soon as they vest).

8. See, e.g., *id.* at 184 (analyzing problems resulting from the broad freedom of executives to unload equity incentives in the short term).

9. See, e.g., *id.* at 175 (suggesting that executives be required to hold stock for the long term); Lucian A. Bebchuk & Jesse M. Fried, *Paying for Long-Term Performance*, 158 U. PA. L. REV. 1915, 1928–36 (2010) (putting forward more detailed recommendations for long-term holding requirements and explaining that executives should be allowed to unwind only a small fraction of their equity each year); Sanjai Bhagat & Roberta Romano, *Reforming Executive Compensation: Focusing and Committing to the Long-Term*, 26 YALE J. ON REG. 359, 361 (2009) (suggesting that executives be paid only with restricted stock and stock options that cannot be unwound until after retirement).

10. See Jesse M. Fried, *Informed Trading and False Signaling with Open Market Repurchases*, 93 CALIF. L. REV. 1323, 1331 (2005). Aggregate shareholder value is the net cash flow from the firm to all of the firm's shareholders over time, where net cash flow equals the cash received by shareholders from the firm via dividends and share repurchases less any cash paid by shareholders to the firm for their shares. I use the term *aggregate shareholder value* rather than *shareholder value* because *shareholder value* may be taken to mean the value flowing to the firm's *current* shareholders.

destroy aggregate shareholder value (and therefore economic or “social” value¹¹).

The Article describes and analyzes two distortions caused by tying an executive’s payoff to a stock’s future value. First, when the stock’s current price is below its actual value, an executive whose pay is tied to the stock’s future value can be rewarded for diverting cash from productive investments in the firm to fund bargain-price share repurchases. I call this distortion—which involves socially excessive cash distributions by the firm—“costly contraction.”

Second, when the stock’s current price is higher than its actual value, the executive can be rewarded for issuing new shares even if the cash or other assets received for the newly issued shares cannot be used productively by the firm. I term this distortion—which involves socially excessive investment by the firm—“costly expansion.” Both costly contraction and costly expansion can boost the stock’s future value and executives’ payouts even as they destroy aggregate shareholder value.

The reason why tying executive pay to the stock’s future value leads to these two distortions is straightforward. Tying pay to the stock’s future value aligns an executive’s interests with the interests of only one group of shareholders, whom I call “nontrading shareholders”: investors who neither sell any of their shares nor buy any additional shares until the executive receives the value of her shares. It fails to align the executive’s interests with those of two other groups of shareholders: (1) “redeeming shareholders”—investors who sell shares to the firm before the executive cashes out her equity, and (2) “investing shareholders”—investors who buy (additional) shares from the firm before the executive’s cash-out date. Thus, tying an executive’s pay to the stock’s future value rewards the executive for transferring value to nontrading shareholders from redeeming shareholders (by buying the latter’s shares at a low price) and investing shareholders (by selling them shares at an inflated price), even if aggregate shareholder value—the amount of value flowing to all three groups of shareholders—is thereby diminished.¹²

11. For purposes of this Article, however, I assume that the firm’s current and future shareholders are the only residual claimants to the firm’s cash flow, and thus that aggregate shareholder value is equivalent to social value. (I will thus use the terms “aggregate shareholder value” and “social value” interchangeably.) This assumption, made purely for expositional convenience, does not affect the Article’s analysis about the distortions caused by tying executives’ payoffs to the stock price, or about the desirability of the constant-share proposal this Article puts forward.

12. Tying pay to the stock’s future value also fails to tie executives’ payoffs to the value flowing to two other groups of shareholders: (1) shareholders who sell their shares in the market before the executive’s cash-out date and (2) the investors who buy these selling shareholders’ stock. However, the cash that changes hands when investors buy and sell a firm’s shares to each other in the secondary market does not affect aggregate shareholder value—the total amount of value flowing from the firm to shareholders over time. Rather, trading in the secondary market merely

I also put forward a mechanism that would perfectly tie executive pay to aggregate shareholder value: the “constant-share” approach to equity pay. Under this approach, an executive would be required to adjust her equity holdings in the firm whenever the firm purchases or sells its own shares to keep her percentage ownership constant through the transaction. Thus, the executive would be required to sell some of her shares whenever the firm repurchases its own stock and to buy additional shares when the firm issues new equity.

In essence, the constant-share approach requires the executive to participate equally as both a redeeming shareholder and as a nontrading shareholder when the firm repurchases shares, and as both an investing shareholder and a nontrading shareholder when the firm issues shares. The constant-share approach thus ties the executive’s payoff to the value flowing to all of the firm’s shareholders and rewards the executive for engaging in repurchases and equity issuances if and only if those transactions increase aggregate shareholder value.¹³

Before proceeding, I would like to be clear about the normative assumptions underlying this Article’s analysis. Consistent with standard and widely used notions of efficiency that underlie most of the economically oriented scholarship on corporate governance, I assume that executive-compensation arrangements should reward executives for generating—not destroying—social value.¹⁴ Thus, as Michael Jensen has argued, executives should not be incentivized to redistribute value from future shareholders to current shareholders (or from one group of current shareholders to another) in ways that reduce social value, even if some of the firm’s current shareholders are made better off.¹⁵

I would also like to indicate my objectives in this Article. Actual implementation of the constant-share approach at any given firm would

redistributes value among different shareholders. Thus, these shareholders’ returns can be ignored when analyzing the extent to which equity compensation ties executives’ payoffs to aggregate shareholder value.

13. In this Article, I abstract from the question of how much—and what elements of—aggregate shareholder value should be paid to executives. That is, I do not consider here how much equity executives should receive, whether equity pay should take the form of stock or options, or the extent to which the payoffs from these instruments should be designed to filter out changes in the stock price that are due to market-wide or industry-wide fluctuations. My objectives are (1) to demonstrate that any arrangement that ties pay to the stock’s future value can reward an executive for destroying aggregate shareholder value, and (2) to show how pay can be tied to aggregate shareholder value to eliminate these distortions.

14. Cf. William T. Allen et al., COMMENTARIES AND CASES ON THE LAW OF BUSINESS ORGANIZATION 7 (3d ed. 2009) (urging the use of Kaldor-Hicks efficiency as a criterion for evaluating corporate law and corporate governance arrangements).

15. See Michael C. Jensen, *Agency Costs of Overvalued Equity*, 34 FIN. MGMT. 5, 16 (2005) (arguing that managers and the board should treat all shareholders (including future shareholders) equally to maximize the firm’s long-run economic value).

require certain technical adjustments¹⁶ as well as the adoption of anti-circumvention arrangements—issues that are beyond the scope of this project. I am thus not advocating adoption of any particular arrangement here. Rather, my goals in this Article are as follows: (1) to demonstrate that, just as there is a potential economic cost associated with tying executive pay to the short-term stock price, there is a potential economic cost associated with tying executive pay to the stock's future value, even its long-term value, at least for the typical company that repurchases and issues shares; and (2) to show conceptually how compensation arrangements should be structured to tie executive pay to the value generated by a firm for its shareholders over time.

The remainder of the Article is organized as follows. Part II explains why equity pay can give executives an incentive to engage in excessive repurchases. It begins by showing that tying executive pay to the stock's future value rewards executives for engaging in repurchases when the stock's current price is below its actual value. It next explains how repurchases can reduce aggregate shareholder value by distributing cash that should be invested in the firm's own projects. It concludes by demonstrating that executives holding stock can be rewarded for conducting bargain-price repurchases even when those repurchases destroy social value.

Part III explains how equity pay gives executives an incentive to engage in excessive equity issuances. It begins by showing that tying executive pay to the stock's future value rewards executives for issuing equity when it is overpriced. It then demonstrates that the investments financed by such equity issuances can reduce social value. It concludes by showing that executives can be rewarded for inflated-price equity issuances even if the investments financed thereby destroy social value.

Part IV introduces the constant-share approach to equity pay, under which executives must maintain their proportional ownership as the firm transacts in its own stock. Such an approach, it shows, eliminates the incentive to engage in costly contraction and costly expansion created by tying executives' payoffs to the stock's future value. A conclusion follows.

II. Equity Pay and Costly Contraction

In this Part, I explain why tying executive payoffs to the stock's future value, even its long-term value, can reward executives for engaging in repurchases that reduce social value. Subpart A briefly discusses the growing use of repurchases as a means to distribute cash to shareholders. Subpart B shows that executives holding stock in the firm have a strong incentive to undertake repurchases when the firm's current stock price is below its actual

16. For example, much of an executive's equity is likely to consist of vested stock options and unvested shares and stock options. Measuring the executive's proportional equity ownership would require assigning a share-equivalent to these instruments.

value, and it summarizes the considerable evidence that executives frequently conduct such bargain-price repurchases. Subpart C explains how repurchases can reduce social value by diverting cash from valuable firm projects. Subpart D demonstrates that executives holding stock in the firm can be rewarded for engaging in such value-wasting repurchases when the current stock price is sufficiently low.

A. Widespread Use of Repurchases

Publicly traded U.S. firms annually generate hundreds of billions of dollars in earnings.¹⁷ Each year, firms must decide how much of their retained earnings should be distributed to shareholders rather than left in the firm. Executives must also decide the form that such distribution should take: dividends, repurchases, or a combination of both.¹⁸

Share repurchases have become increasingly common and are now considered the dominant form of cash payout.¹⁹ Over 90% of U.S. public firms that distribute cash engage in repurchases.²⁰ In 2007, S&P 500 firms distributed almost \$600 billion through repurchases.²¹

A repurchase will typically take one of two forms: (1) an “open market repurchase” (OMR), in which the firm buys its own stock on the market through a broker,²² or (2) a “repurchase tender offer” (RTO), in which the firm offers to buy back its own stock directly from shareholders, usually at a premium over the market price.²³ Because over 90% of repurchases take the form of OMRs,²⁴ my analysis focuses primarily on OMRs.²⁵

17. See Gustavo Grullon & Roni Michaely, *Dividends, Share Repurchases, and the Substitution Hypothesis*, 57 J. FIN. 1649, 1655 tbl.1 (2002) (reporting annual aggregate earnings of U.S. firms from 1972 through 2000).

18. See generally Douglas J. Skinner, *The Evolving Relation Between Earnings, Dividends, and Stock Repurchases*, 87 J. FIN. ECON. 582 (2008) (comparing the percentages of firms that pay dividends, firms that repurchase shares, and firms that do both).

19. See *id.* at 584.

20. See Gustavo Grullon & David L. Ikenberry, *What Do We Know About Stock Repurchases?*, J. APPLIED CORP. FIN., Spring 2000, at 31, 33–34 (reporting that in the 1990s additional cash flows were channeled into share repurchases instead of dividends); Skinner, *supra* note 20, at 583 (explaining that in 2005 only 7% of firms paid dividends and did not distribute any cash through repurchases).

21. Press Release, Standard & Poor's, S&P 500 Buybacks Set Record of \$589 Billion in 2007 (Apr. 7, 2008), available at http://www2.standardandpoors.com/spf/pdf/index/040708_SP500_BUYBACK_PR.pdf; see also Paul A. Griffin & Ning Zhu, *Accounting Rules? Stock Buybacks and Stock Options: Additional Evidence*, 6 J. CONTEMP. ACCT. & ECON. 1, 1 (2010) (reporting \$1 trillion of repurchases market wide in 2007).

22. Jesse M. Fried, *Informed Trading and False Signaling with Open Market Repurchases*, 93 CALIF. L. REV. 1323, 1335 (2005).

23. Jesse M. Fried, *Insider Signaling and Insider Trading with Repurchase Tender Offers*, 67 U. CHI. L. REV. 421, 421 (2000).

24. See Monica L. Banyl et al., *Errors in Estimating Share Repurchases*, 14 J. CORP. FIN. 460, 460 (2008) (reporting that since 1996 open market repurchase programs have accounted for 88% of all announced repurchase programs, and the announced value of these open market repurchases has been over 93% of the total reported value of repurchase programs); Grullon & Ikenberry, *supra*

Economists believe that the growing use of repurchases rather than dividends is likely the result of the widespread use of stock options to compensate executives.²⁶ Stock options provide a payoff equal to the difference between the (usually fixed) exercise price and the exercise-date stock price.²⁷ A dividend reduces firm assets without affecting the number of shares.²⁸ As a result, a dividend reduces per-share value and the share price, thereby diminishing the value of an executive's stock options.²⁹

A repurchase, in contrast, does not reduce per-share value (as much or at all) because it decreases firm assets and the number of shares outstanding by approximately the same proportion.³⁰ Executives compensated with stock options are thus biased in favor of repurchases.³¹ Not surprisingly, executives with larger option packages tend to pay lower dividends and distribute more cash through share repurchases.³²

However, as I explain below, even absent this stock-option bias, executives will often have an incentive to prefer repurchases over dividends.

note 22, at 33–34 & fig.1 (reporting that between 1980 and 1999, open-market programs comprised about 92% of the total repurchase announcements and 91% of the total value of all repurchase announcements).

25. However, for purposes of this Article, the mechanism by which firms repurchase stock is irrelevant.

26. See, e.g., George W. Fenn & Nellie Liang, *Corporate Payout Policy and Managerial Stock Incentives*, 60 J. FIN. ECON. 45, 48 (2001) (finding a statistically significant positive relationship between repurchases and management stock options, and concluding that management stock options help explain the rise in repurchases at the expense of dividends). Repurchases may also offer shareholders a number of possible advantages over dividends. In many cases, they are a more tax-efficient mechanism than dividends for distributing cash. Repurchases (unlike dividends) also may enable firms to acquire shares for increasingly popular stock-option plans or provide liquidity to a firm's selling shareholders. See Fried, *supra* note 24, at 1336–40 (describing the possible advantages of repurchases over dividends and explaining why many of the advantages are likely to be quite modest).

27. See, e.g., Christine Jolls, *Stock Repurchases and Incentive Compensation* 1–2 (Nat'l Bureau of Econ. Research, Working Paper No. 6467, 1998), available at <http://www.nber.org/papers/w6467.pdf> (describing the structure of stock options and exploring how this structure affects management incentives).

28. *Id.*

29. *Id.* One could preserve the value of an executive's stock option following a dividend by reducing the exercise price by the amount of the dividend. However, for various reasons such dividend adjustments are uncommon. See Kathleen M. Kahle, *When a Buyback Isn't a Buyback: Open Market Repurchases and Employee Options*, 63 J. FIN. ECON. 235, 242 n.2 (2002) (“[O]nly 1% of CEOs with options have dividend protection.” (citation omitted)). Thus, dividends typically reduce the value of an executive's stock options. Jolls, *supra* note 29, at 2.

30. Jolls, *supra* note 31, at 1.

31. Fenn & Liang, *supra* note 30, at 65. Executives paid with restricted stock will also have an incentive to repurchase shares rather than issue dividends if the executives are not entitled to receive the value of any dividends paid while the restricted stock is vesting. However, most executives compensated with restricted stock appear to be entitled to receive dividends while the stock is still vesting. See Phyllis Plitch, *Executives Find Restricted Stock Pays Dividends from the Get-Go*, WALL ST. J., Feb. 28, 2005, at C3 (reporting an estimate that 90% of U.S. publicly traded companies award dividends on unvested restricted stock).

32. Fenn & Liang, *supra* note 30, at 47–48.

In particular, any executive whose payoff is tied to the stock's future value will get a larger payoff by engaging in a bargain-price repurchase than by issuing a dividend.

B. Using Repurchases to Buy Low

Executives whose payoff is tied to the future value of their firm's stock have a strong incentive to repurchase stock when the stock's current price is below its actual value. Indeed, there is substantial evidence that executives frequently engage in bargain-price repurchases. Bargain pricing persists even after firms announce that they plan to repurchase their own shares.

1. *Executives' Payoffs.*—As I have shown elsewhere,³³ a repurchase is economically equivalent to the following two-step transaction: First, redeeming (selling) shareholders sell their shares to nontrading (continuing) shareholders directly at the repurchase price. Second, the firm issues a dividend to nontrading shareholders equal to the dollar amount of the repurchase. Thus, a bargain-price repurchase transfers value from redeeming shareholders to continuing shareholders, including executives holding equity.³⁴

A simple example can be used to illustrate how a bargain-price repurchase transfers value to executives holding equity in the firm. Consider ABC Corporation (ABC) that has two shares outstanding and is liquidated at Liquidation Date.³⁵ One share is held by its CEO. The other share is held by public shareholders. Consider two scenarios:

No-Transaction Scenario: Suppose that if ABC does not repurchase any of its equity prior to Liquidation Date, it will distribute \$20 to the holders of its two shares at Liquidation Date. The no-transaction value of each of ABC's two shares at Liquidation Date is thus \$10.

Repurchase Scenario: Now suppose that ABC can conduct a repurchase before Liquidation Date when the stock trades at \$8 (\$2 less than its actual value of \$10), buying back public shareholders' single share at that price. Assume that the \$8 spent on the repurchase reduces ABC's Liquidation Date value from \$20 to \$12. At Liquidation Date, the value of ABC's remaining share (held by CEO) is thus \$12.

It should be easy to see that the bargain-price repurchase boosts CEO's payout without increasing social value—the value flowing to ABC's

33. Fried, *supra* note 10, at 1344–46.

34. When a firm buys stock at a price below its actual value, the precise distributional effects depend on whether the redeeming shareholders would have otherwise sold their shares to new investors for the same price. If so, the redeeming shareholder cannot be said to "lose" any value as a result of the bargain-price repurchase. Instead, the repurchase deprives would-be new investors of a gain. For simplicity, however, I will assume that it is the redeeming shareholders that lose money as the result of the bargain-price repurchase.

35. I assume that ABC does not issue any dividends (or sell any equity) before Liquidation Date.

shareholders over time. In both the No-Transaction and Repurchase Scenarios, ABC's social value is \$20.³⁶ But CEO's payout in the No-Transaction Scenario is only \$10, while in the Repurchase Scenario it is \$12. The results are summarized in Table 1 below.

Table 1. Rewards for Bargain-Price Repurchase

	Social Value	CEO Payoff
No Repurchase	\$20	\$10
Repurchase	\$20	\$12

2. *Evidence of Bargain Repurchases.*—Having seen that executives holding equity have an incentive to conduct bargain repurchases, I now turn to the considerable empirical evidence that they do so. This evidence includes (a) executives' own statements and behavior, and (b) stock-price movements following repurchase announcements.

a. *Executives' Own Statements and Behavior.*—Executives admit that they frequently use repurchases to buy stock when it is cheap. According to the authors of a major 2005 survey of financial executives regarding their firms' payout policies, "[t]he most popular response for all repurchase questions on the entire survey is that firms repurchase when their stock is a good value, relative to its true value: 86.4% of all firms agree or strongly agree with this supposition."³⁷ According to the survey's organizers, "executives tell us that they accelerate (or initiate) share repurchases when their company's stock price is low."³⁸

Empirical studies confirm that executives' desire to buy stock at a low price is linked to their equity ownership. One study found that abnormal returns following repurchase announcements, which are associated with pre-repurchase underpricing, are positively correlated with pre-buyback executive stock ownership.³⁹ Another found that relatively infrequent

36. In the No-Transaction Scenario, all \$20 flows to shareholders at Liquidation Date. In the Repurchase Scenario, \$8 flows to shareholders during the repurchase and \$12 flows to shareholders at Liquidation Date (for a total of \$20). Throughout the examples in this Article, I ignore the time value of money (or alternatively, assume it is zero). This assumption, made purely for convenience, does not affect the analysis.

37. Alon Brav et al., *Payout Policy in the 21st Century*, 77 J. FIN. ECON. 483, 514 (2005).

38. *Id.* Earlier studies yielded similar responses. When asked in a 1988 survey what the most important circumstance precipitating a repurchase was, 66% of the surveyed executives responded "low stock price," which was six times as many as the next most popular answer, "need for treasury stock." George P. Tsetsekos et al., *A Survey of Stock Repurchase Motivations and Practices of Major U.S. Corporations*, 7 J. APPLIED BUS. RES. 15, 17–18 tbl.2 (1991).

39. See Elias Raad & H.K. Wu, *Insider Trading Effects on Stock Returns Around Open-Market Stock Repurchase Announcements: An Empirical Study*, 18 J. FIN. RES. 45, 57 (1995).

repurchase announcers—those firms that are more likely to be engaged in bargain repurchasing than repurchasing shares to acquire stock for employee-option programs—also tend to have higher levels of executive ownership.⁴⁰ Both of these studies indicate that executives are more likely to engage in bargain-price repurchases when executives hold more equity.

Moreover, executives actively manipulate earnings to drive the stock price down around repurchases and thereby increase the amount of value transferred to themselves and other non-selling shareholders.⁴¹ Such earnings manipulation is more aggressive when the CEO's equity ownership is higher, providing additional evidence that executives conduct repurchases to boost the value of their equity.⁴²

One might wonder why insider trading laws do not prevent executives from engaging in such indirect insider trading. But as I have explained elsewhere,⁴³ there are substantial limits on the law's ability to deter corporations from using inside information to trade in their own stock.⁴⁴ Executives are thus left with considerable ability to use the corporation for indirect insider trading when the stock price is below its actual value.

b. Post-announcement Stock Returns.—Stock-price movements following repurchase announcements also suggest that inside information drives many repurchases. If executives use repurchases to buy stock at a low price, the stock prices of firms announcing repurchases should, on average, subsequently outperform those of firms not announcing repurchases. Indeed, stock prices of firms announcing repurchases increase faster than stock prices of similar firms not announcing repurchases. One study found that shares of firms announcing repurchases earn abnormal returns of 6.7% in the first year following the announcement and 23.6% over the subsequent four years.⁴⁵

40. See Murali Jagannathan & Clifford Stephens, *Motives for Multiple Open-Market Repurchase Programs*, 32 FIN. MGMT. 71, 71–72 (2003).

41. See Guojin Gong, Henock Louis & Amy X. Sun, *Earnings Management and Firm Performance Following Open-Market Repurchases*, 63 J. FIN. 947, 983 (2008) (reporting that firms adjust accruals to decrease their reported earnings before stock repurchases).

42. *Id.*

43. Fried, *supra* note 10, at 1343.

44. Indeed, it is not completely clear under current law whether it is ever illegal for a corporation to buy its own stock in the public markets on inside information. See Mark J. Loewenstein & William K.S. Wang, *The Corporation as Insider Trader*, 30 DEL. J. CORP. L. 45, 70–72 (2005) (noting that the SEC takes the position that an issuer trading on material inside information would violate Rule 10b-5 but that this position has not been endorsed by any court).

45. Konan Chan et al., *Economic Sources of Gain in Stock Repurchases*, 39 J. FIN. & QUANT. ANALYSIS 461, 463 (2004); see also Urs Peyer & Theo Vermaelen, *The Nature and Persistence of Buyback Anomalies*, 22 REV. FIN. STUD. 1693, 1701 (2009) (finding, in a large sample of firms announcing OMRs, a 24.25% cumulative market-adjusted return over 48 months following OMR announcements).

Post-announcement returns are even higher in those firms that actually repurchase shares after making a repurchase announcement.⁴⁶ Focusing on “value firms” (firms with a high book-to-market ratio) that had announced repurchases, one study found that among the firms in which managers subsequently repurchased more than 4% of the firm’s shares in the year following the repurchase announcement, four-year post-announcement abnormal returns were 57%. For value firms that announced a repurchase but did not subsequently repurchase any shares, the authors could not find any evidence of post-announcement abnormal returns.⁴⁷ These post-announcement returns provide further strong evidence (along with executives’ own statements and their behavior) that executives often use repurchases to indirectly buy underpriced stock.

3. *Why Bargain Pricing Persists After Repurchase Announcements.*— Because executives often use repurchases to buy stock at a low price, a repurchase announcement will tend to signal that the value of the stock is higher than the current market price. This signal, in turn, can be expected to boost the stock price, reducing the amount of underpricing and executives’ ability to profit from a bargain-price repurchase. Indeed, if a repurchase announcement clearly signaled a certain amount of underpricing, the announcement should—in an efficient market—immediately eliminate the underpricing.

However, investors do not appear to immediately impart the information contained in repurchase announcements into the stock price, just as they do not immediately impart other types of information into the stock price.⁴⁸ The failure of shareholders to immediately react to certain types of public information has been labeled by economists “investor underreaction.”⁴⁹ In short, markets are not as efficient as some commentators might believe.

46. Chan et al., *supra* note 51, at 476. After making a repurchase announcement, which is not binding, executives can choose how much (if any) equity to actually repurchase. See Fried, *supra* note 10, at 1335 (explaining that firms announcing authorization of open-market repurchases, the most common form of repurchases, are not required to indicate the number of shares they intend to repurchase or to commit to repurchasing any shares).

47. See Konan Chan et al., *Do Managers Time the Market? Evidence from Open-Market Share Repurchases*, 31 J. BANKING FIN. 2673, 2676, 2686–88 (2007). For other studies indicating that executives in the United States and elsewhere tend to repurchase stock when it is underpriced, see Paul Brockman & Dennis Y. Chung, *Managerial Timing and Corporate Liquidity: Evidence from Actual Share Repurchases*, 61 J. FIN. ECON. 417, 418 (2001), and Clifford P. Stephens & Michael S. Weisbach, *Actual Share Reacquisitions in Open-Market Repurchase Programs*, 53 J. FIN. 313, 313 (1998).

48. See David Ikenberry et al., *Market Underreaction to Open Market Share Repurchases*, 39 J. FIN. ECON. 181, 183 (1995) (discussing delayed market reactions to announcements of important corporate events such as repurchases, mergers, proxy contests, and spinoffs).

49. See, e.g., Harrison Hong & Jeremy C. Stein, *A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets*, 54 J. FIN. 2143, 2143 (1999); Ikenberry, *supra* note 53, at 183.

Moreover, even if markets were efficient at processing publicly available information, bargain pricing would still persist after repurchase announcements. A buyback announcement does not unambiguously signal that the stock is underpriced because boards may announce a buyback even when the stock is not underpriced. For example, as I have explained elsewhere, executives might announce a repurchase program that they have no plan to actually conduct simply to boost the stock price so they can unload their own shares at a higher price.⁵⁰ Indeed, a recent paper finds evidence of what I have called such “false signaling.”⁵¹ To the extent repurchase announcements are made for reasons other than conducting bargain repurchases, the resulting adjustment to the stock price will not completely eliminate any underpricing—even if the market is efficient.⁵²

Because of the underreaction problem and the fact that some repurchase announcements are not driven by the desire to buy stock at a low price, the market response to repurchase announcements is, on average, rather muted. Repurchase announcements are associated with short-term abnormal price increases averaging 3% to 4% in the 1980s⁵³ and approximately 2% in the 1990s.⁵⁴ The more muted the market’s response to a repurchase announcement, the more profits executives can reap repurchasing underpriced stock.

C. *The Possibility of Costly Repurchases*

We have just seen that executives whose payoff is tied to the stock’s future value have incentives to engage in bargain repurchases even if those repurchases do not increase social value. In addition, there is considerable evidence that executives do engage in bargain repurchases.⁵⁵ I will now

50. See Fried, *supra* note 10, at 1351–56 (developing the argument that executives can use repurchase announcements for false signaling and providing anecdotal accounts of such false signaling); Jesse M. Fried, *Open Market Repurchases: Signaling or Managerial Opportunism?*, 2 THEORETICAL INQUIRIES IN L. 865, 879–81 (2001) (explaining that executives can use repurchase announcements to boost the stock price before selling their shares). Alternatively, firms may announce a repurchase because they need to repurchase shares for employee stock-option programs.

51. See Konan Chan et al., *Share Repurchases as a Potential Tool to Mislead Investors*, 16 J. CORP. FIN. 137, 139 (2010) (finding evidence consistent with the notion of executives of poorly performing firms making share repurchase announcements without an intention to repurchase shares).

52. Because of the problem of false signaling, I have suggested that firms be required to disclose not only their intention to repurchase shares but also the exact details of any buy orders given to brokers shortly before the orders are made. Fried, *supra* note 10, at 1330. Such a pre-disclosure rule would increase the accuracy of price adjustments to repurchase announcements and reduce the amount of underpricing when the stock’s price is below its actual value.

53. See Ikenberry et al., *supra* note 55, at 190 (reporting that the average market reaction to OMR announcements for all of the OMRs announced between January 1980 and December 1990 by firms listed on the ASE, NYSE, and NASDAQ was 3.54%).

54. See Peyer & Vermaelen, *supra* note 51, at 1697 (finding that, in a sample of OMR announcements from 1991–2001, there were average abnormal stock price reactions of 2.39% in the three days around the announcement).

55. See *supra* note 44 and accompanying text.

explain how repurchases (whether or not they are bargain priced) can reduce social value.

1. How Repurchases Can Reduce Social Value.—A repurchase can reduce social value by distributing cash that, from a social perspective, could generate higher returns if invested in the firm's own projects. Suppose, for example, that \$100 left in the firm would generate a return of 15% (\$15). Suppose that if instead the \$100 were distributed to shareholders, the shareholders receiving the cash could generate returns of only 10% (\$10) outside the firm. Distributing the \$100 through a repurchase would thus destroy \$5 of social value.

In fact, there is evidence consistent with repurchases diverting cash that would otherwise be used in the firm. A recent study found that repurchases, especially those that appear to be driven by executive stock ownership, have a significantly negative effect on a firm's short-term investments and research and development.⁵⁶ The study found that, everything else equal, doubling repurchases led to an 8% reduction in research-and-development expenditures. An earlier study came to similar conclusions—that repurchases led to firms diverting cash from potentially productive investments.⁵⁷

To be sure, these two studies do not establish that most or even many repurchases destroy social value. But the studies do provide evidence that repurchases can divert cash from productive activities inside the firm, increasing the likelihood that some repurchases distribute cash that would generate more social value inside the firm.

2. Constraints on Firm Borrowing.—One might wonder why a firm that has a valuable project and whose stock trades at a low price cannot have its cake and eat it too. Indeed, in a world of perfect capital markets, there would be no need to sacrifice desirable firm projects to fund a bargain-price repurchase: firms could easily find the cash both to buy their stock at a low price and to invest in their high-value projects.

Consider the example above where shareholders can earn 10% on the cash they receive from the corporation and a 15% project is sacrificed to fund a \$100 repurchase. In a world of perfect capital markets, the corporation should be able to obtain financing for any project with a positive net present

56. See Alok Bhargava, *Executive Compensation, Share Repurchases, and Investment Expenditures: Econometric Evidence from U.S. Firms* (2010) (unpublished manuscript) (on file with author).

57. See Daniel A. Bens et al., *Real Investment Implications of Employee Stock Option Exercises*, 40 J. ACCT. RES. 359, 359 (2002) (finding some evidence that firms that repurchase shares to satisfy option exercises exhibit subsequent poor performance because the repurchases divert cash from productive investments).

value.⁵⁸ Thus, a firm's ability to invest in desirable projects would not depend on having cash on hand. A firm could both repurchase \$100 worth of shares and invest \$100 in the desirable project simply by borrowing another \$100 for the project.

However, there are a number of reasons why a firm may not borrow enough money to fully fund the desirable project. First, information asymmetry may prevent a firm from borrowing money on cost-effective terms.⁵⁹ While the firm's executives may know that the firm's prospects are good, outside lenders asked to provide capital may lack sufficient information to reach the same conclusion. Outside lenders may thus demand terms that make the financing of the desirable project too costly, leading executives to forgo the project.

Second, even if a firm could borrow on reasonable terms from a lender, the borrowing may not be permitted by the firm's existing arrangements. For example, loan covenants with existing lenders might bar the firm from borrowing additional funds. Covenants are inevitably both underinclusive and overinclusive: they fail to prevent some value-decreasing activities and unfortunately prevent some value-increasing activities.⁶⁰ In this case, a loan covenant would be overinclusive: it would prevent the firm from financing a desirable project with additional debt.⁶¹

Third, executives who are risk averse may wish to avoid the additional discipline imposed by more debt. Even if credit could be obtained on reasonable terms and the firm's existing arrangements would permit such a borrowing, the executives might personally be better off forgoing the valuable project rather than having the firm take on more debt. For any one or all three of these reasons, there may be a trade-off between a firm's ability to repurchase its shares and its ability to fund productive activities inside the firm.

58. Stewart C. Myers & Nicholas S. Majluf, *Corporate Financing and Investment Decisions when Firms Have Information that Investors Do Not Have*, 13 J. FIN. ECON. 187, 187 (1984).

59. See generally *id.* at 187-220.

60. See Lucian Arye Bebchuk & Jesse M. Fried, *The Uneasy Case for the Priority of Secured Claims in Bankruptcy*, 105 YALE L.J. 857, 879 (1996) (noting that the difficulty of specifying all possible contingencies is likely to cause covenants to be overinclusive in some respects); Marcel Kahan & David Yermack, *Investment Opportunities and the Design of Debt Securities*, 14 J.L. ECON. & ORG. 136, 138-40 (1998) (explaining that the potential overinclusiveness of particular covenants may cause the parties to avoid such covenants when their ability to renegotiate these covenants is diminished).

61. In principle, these covenants preventing a value-increasing investment could be renegotiated, with the resulting surplus shared between the lender and the borrower. But such renegotiation is often difficult or costly, particularly when the borrower must simultaneously renegotiate with multiple creditors to obtain the modifications needed to facilitate the new investment.

D. Rewards for Costly Contraction

We have seen that executives holding stock can profit from bargain-price repurchases and that repurchases can reduce social value. I will now show that executives whose payoff is tied to the stock's future value can be rewarded for engaging in value-wasting repurchases. I will then explain that other components of executives' pay packages fail to mitigate this distortion.

1. *Equity Pay and Costly Repurchases.*—To see why equity rewards executives for costly contraction, let us return to the example of ABC Corporation introduced in subpart B. As before, ABC has two shares outstanding and is liquidated at Liquidation Date.⁶² One share is held by its CEO. The other share is held by public shareholders. Consider two scenarios:

No-Transaction Scenario: Suppose that if ABC does not repurchase any of its equity prior to Liquidation Date, it will distribute \$20 to the holders of its two shares at Liquidation Date. The no-transaction value of each of ABC's two shares at Liquidation Date is thus \$10.

Costly Repurchase Scenario: Now suppose that ABC can conduct a repurchase before Liquidation Date when the stock trades at \$8 (\$2 less than its actual value of \$10), buying back public shareholders' single share at that price. Assume that the \$8 spent on the repurchase reduces ABC's Liquidation Date value by \$9, from \$20 to \$11, because ABC must give up a valuable project. At Liquidation Date, the value of ABC's remaining share (held by CEO) is thus \$11.

It should be easy to see that the costly bargain-price repurchase boosts CEO's payout even though it reduces social value. In the No-Transaction Scenario, ABC's social value is \$20; in the Costly Repurchase Scenario, ABC's social value is \$19.⁶³ But CEO's payout in the Costly Repurchase Scenario is \$11, \$1 more than in the No-Transaction Scenario. The effect of the repurchase on shareholder value and CEO's payoff can be summarized in the following table:

Table 2. Rewards for Costly Repurchase

	Social Value	CEO Payoff
No Repurchase	\$20	\$10
Repurchase	\$19	\$11

62. I assume that ABC does not issue any dividends (or sell any equity) before Liquidation Date.

63. In the No-Transaction Scenario, all \$20 flows to shareholders at Liquidation Date. In the Repurchase Scenario, \$8 flows to shareholders during the repurchase and \$11 flows to shareholders at Liquidation Date (for a total of \$19).

As one can see, an executive whose payoff is tied to the stock's future value can be rewarded for repurchasing shares even when the transaction destroys social value. The problem is that equity aligns an executive's interests with those of nontrading shareholders but not with those of redeeming shareholders. Thus, the executive has an incentive to transfer value from redeeming shareholders to nontrading shareholders even when such transfer destroys social value. Importantly, this distortion arises even if the executive's payoff is tied to the stock's long-term value.

2. *Do Other Pay Components Mitigate?*— We just saw that an executive whose payoff is tied to the stock's future value can be rewarded for engaging in a repurchase that reduces social value. But executives are typically free to sell much of their equity in the short term.⁶⁴ One might think that having stock that could be sold in the short term would tie the executive's payout to that of short-term shareholders, including redeeming shareholders who sell stock back to the corporation when the firm conducts a repurchase.

However, the fact that executives *can* sell stock in the short term does not mean that they *will* sell stock in the short term. When executives know that the stock is underpriced and conduct a repurchase in order to indirectly buy stock at a low price, they can be expected to hold onto their personal shares until the stock price rises. In fact, there is evidence that executives buy additional shares for their personal accounts before and during bargain repurchases.⁶⁵ Thus, other components of executives' compensation arrangements will not mitigate executives' financial incentives to engage in costly repurchases.⁶⁶

III. Equity Pay and Costly Expansion

Part II demonstrated that executives whose payoff is tied to the stock's future value can be rewarded for inefficiently contracting the firm's operations when the firm's current stock price is below its actual value. This Part shows that such executives are also rewarded for inefficiently expanding the firm's operations when the firm's current stock price is above its actual value. In particular, such executives benefit from having the firm sell additional equity at inflated prices, even when the cash raised by such an issuance is invested in ways that reduce social value.

Subpart A discusses the widespread use of equity issuances by firms. Subpart B explains that executives whose payoff is tied to the stock's future value benefit when the firms sell overpriced stock. It also provides evidence

64. See BEBCHUK & FRIED, *supra* note 7, at 174–79.

65. See Raad & Wu, *supra* note 45, at 57 (reporting higher levels of insider stock purchases in the month immediately preceding a share repurchase announcement).

66. Of course, executives whose total pay is tied to firm size may have somewhat less incentive to engage in a costly repurchase, or indeed any repurchase.

that such inflated-price issuances are common. Subpart C explains why equity issuances can reduce social value. Subpart D then shows that executives can benefit even if inflated-price issuances lead to a reduction in social value.

A. *The Use of Share Issuances*

After undergoing an IPO, a publicly traded firm usually continues to issue shares throughout its life.⁶⁷ Such issuances typically serve one of three purposes. First, stock is issued to employees as part of their compensation packages.⁶⁸ Second, stock issuances are used to raise cash for operations and strategic investments or to pay down debt.⁶⁹ Third, in the context of corporate acquisitions, stock is often issued to target shareholders as consideration for their shares in the target company.⁷⁰

Although equity issuances have different purposes, all have the same economic consequence: they directly or indirectly move cash or other assets into the firm.⁷¹ Equity issuances thus have the opposite effect of repurchases. While repurchases remove value from the firm and put it into the hands of shareholders, equity issuances take value from shareholders and put it into the firm's hands.

B. *Using Share Issuances to Sell High*

This subpart explains that executives whose payoff is tied to the stock's future value have a strong incentive to issue shares when the stock's current price is higher than its actual value. It then describes the substantial evidence that executives frequently engage in inflated-price issuances.

1. *Executives' Payoffs.*—An equity issuance has analogous distributional effects to a share repurchase. As we saw in Part II, a share

67. See Eugene F. Fama & Kenneth R. French, *Financing Decisions: Who Issues Stock?*, 76 J. FIN. ECON. 549, 550 (2005) (reporting that 86% of publicly traded firms in their sample issued stock between 1993 and 2002 after their IPOs).

68. Among the largest 200 firms in 2007, the range of shares allocated to equity compensation plans ranged from 0.92% of outstanding shares to 62.6% of outstanding shares, with the median around 10.5%. PEARL MEYER & PARTNERS, 2008 EQUITY STAKE STUDY: STUDY OF THE TOP 200 CORPORATIONS 2 (2009).

69. See Fama & French, *supra* note 74, at 573–74 (describing various purposes for stock issuances). These cash-raising issuances may take the form of seasoned equity offerings, private placements, convertible debt, warrants, or rights issues. *Id.* at 550.

70. See *id.* at 554 (explaining the tax advantage of using acquirer-firm stock to purchase shares of targets).

71. The issuance of equity for compensation indirectly moves cash into the firm. The firm gives equity to executives and other employees, who eventually sell the equity for cash on the open market to investors. This practice has the same economic effect as a transaction in which the investors buy stock from the firm for cash and the firm then uses the cash to compensate executives and other employees. The issuance of stock to raise cash to pay down debt also indirectly shifts cash into the firm by reducing future interest payments.

repurchase transfers value from redeeming shareholders to nontrading shareholders when the stock price is below the stock's actual value. Similarly, a stock issuance transfers value from investing shareholders to nontrading shareholders when the stock's price is above its actual value.

Like a share repurchase, an equity issuance is economically equivalent to a two-part transaction that involves investing shareholders trading directly with nontrading shareholders. In particular, a stock issuance is economically equivalent to (1) the firm issuing shares pro rata to nontrading and investing shareholders for the issuance price, and (2) nontrading shareholders selling their portion of the issued shares to the investing shareholders for the issuance price. Thus, an equity issuance transfers value from investing shareholders to nontrading shareholders (including executives holding stock) when the sale price exceeds the value of the issued stock.

To illustrate the incentive of executives holding stock to conduct inflated-price offerings, consider again ABC Corporation. As before, it has two shares outstanding, one held by CEO and one held by public shareholders, and it is liquidated at Liquidation Date.⁷² Consider two scenarios:

No-Transaction Scenario: Suppose that if ABC does not sell any of its equity prior to Liquidation Date, it will distribute \$20 to the holders of its two shares at Liquidation Date. The no-transaction value of each of ABC's two shares at Liquidation Date is thus \$10.

Equity Issuance Scenario: Now suppose that ABC can conduct an equity issuance before Liquidation Date when the stock trades at \$13 (\$3 more than its actual value of \$10), selling a third share at that price. Assume that the \$13 received by ABC increases ABC's Liquidation Date value from \$20 to \$33. At Liquidation Date, the value of each of ABC's three shares (including that held by CEO) is thus \$11.

It should be easy to see that the inflated-price issuance boosts CEO's payout without increasing social value—the value flowing to ABC's shareholders over time. In both the No-Transaction Scenario and Equity Issuance Scenario, ABC's social value is \$20.⁷³ But CEO's payout in the No-Transaction Scenario is only \$10, while in the Equity Issuance Scenario it is \$11. The results are summarized in Table 3 below.

72. I assume that ABC does not issue any dividends (or repurchase any equity) before Liquidation Date.

73. In the No-Transaction Scenario, all \$20 flows to shareholders at Liquidation Date. In the Equity Issuance Scenario, \$13 flows from shareholders during the equity issuance, and \$33 flows back to shareholders at Liquidation Date (for a net amount of \$20).

Table 3. Rewards for Inflated-Price Issuance

	Social Value	CEO Payoff
No Equity Issuance	\$20	\$10
Equity Issuance	\$20	\$11

2. *Evidence of Inflated-Price Issuances.*—There is considerable evidence that firms tend to conduct seasoned equity offerings (SEOs)—transactions in which cash is raised from new and existing shareholders—when the stock is overpriced.⁷⁴ For example, one well-known study found that firms undertaking SEOs systematically underperform benchmark stocks over the five-year post-offering period.⁷⁵ This pattern of underperformance indicates that the shares sold were, on average, overpriced at the time of the SEO.⁷⁶

When a firm's shares are overpriced, the firm is also more likely to acquire other companies and use its shares as consideration in the merger.⁷⁷ There is evidence that such acquisitions boost the long-term stock value of the acquiring firms' shares by enabling the acquiring firms to purchase assets cheaply.⁷⁸ To the extent executives of these firms hold equity that they

74. See, e.g., Tim Loughran & Jay R. Ritter, *The New Issues Puzzle*, 50 J. FIN. 23, 25, 47 (1995) (examining 3702 seasoned equity offerings between 1970 and 1980 and finding evidence consistent with firms announcing stock issues when the stock is grossly overvalued, the market failing to revalue the stock appropriately, and the stock remaining overvalued when the issue occurs).

75. See Loughran & Ritter, *supra* note 74, at 23–25 (examining SEO underperformance between 1970 and 1990); see also Jeffrey Pontiff & Artemiza Woodgate, *Share Issuance and Cross-sectional Returns*, 63 J. FIN. 921, 943–44 (2008) (finding evidence of post-SEO stock underperformance in a more recent sample of U.S. SEOs). For evidence that SEOs are used in other countries to sell stock at an inflated price, see Brian J. Henderson et al., *World Markets for Raising New Capital*, 82 J. FIN. ECON. 63, 66 (2006) (concluding that “firms are more likely to issue equity when the stock market appears to be overvalued”).

76. The failure of investors to immediately impart all of the information signaled by these transactions into the stock price is another example of investor underreaction discussed earlier. See, e.g., Loughran & Ritter, *supra* note 74, at 47–48 (discussing the market's “misvaluation” of SEOs).

77. See, e.g., Ming Dong et al., *Does Investor Misvaluation Drive the Takeover Market?*, 61 J. FIN. 725, 757 (2006) (finding that overpriced firms are more likely to try to acquire other firms that are less overpriced); Matthew Rhodes-Kropf et al., *Valuation Waves and Merger Activity: The Empirical Evidence*, 77 J. FIN. ECON. 561, 600–01 (2005) (concluding that the “vast majority” of mergers involve “highly overvalued bidders”); Itzhak Ben-David et al., *Are Stock Acquirers Overvalued? Evidence from Short Selling Activity* 23–24 (Fisher Coll. of Bus. Working Paper Series, Working Paper No. 2010-03-011, 2010), available at <http://www.ssrn.com/abstract=1572686> (determining that short-selling activity is consistent with acquirers using overvalued stock to buy other companies); cf. Matthew Rhodes-Kropf & S. Viswanathan, *Market Valuation and Merger Waves*, 59 J. FIN. 2685, 2710 (2004) (presenting a model in which acquirers are more likely to use stock when they are overvalued); Andrei Shleifer & Robert W. Vishny, *Stock Market Driven Acquisitions*, 70 J. FIN. ECON. 295, 300 (2003) (proposing that overvalued firms engage in stock-financed acquisitions so that the overvalued firms' shareholders can benefit from obtaining hard assets at a discount).

78. See Pavel G. Savor & Qi Lu, *Do Stock Mergers Create Value for Acquirers?*, 64 J. FIN. 1061, 1063 (2009) (finding that the shares of a sample of stock-financed bidders that completed

cannot currently cash out, they will benefit from using overpriced stock to acquire target assets at a discount.⁷⁹

Recall from Part II that there is evidence that executives manipulate the stock price down around stock buybacks to increase their profits from bargain-price repurchases. Similarly, executives manipulate the stock price up around equity offerings in order to increase the amount transferred from investors buying stock from the firm. One study found that seasoned equity issuers are more likely to manipulate earnings than nonissuers and that such manipulations boost the price around the equity offering.⁸⁰ Such earnings manipulations also occur when the stock is being used to acquire another company.⁸¹

C. *The Possibility of Costly Expansion*

As we have seen, executives whose payoff is tied to the stock's future value are rewarded for conducting inflated-price issuances, and there is considerable evidence that such issuances are common. This subpart shows that equity issuances can reduce social value by diverting money into firm investments that yield lower returns than investments outside of the firm.

1. *Why Expansion Can Reduce Social Value.*—Just as a share repurchase can reduce social value by distributing cash that should be invested into the firm's own projects, an equity issuance can reduce social value by enabling the firm to engage in projects that yield a lower return than

their acquisitions outperformed a control sample of stock-financed bidders that failed to complete their acquisitions by 25–30% over a three-year horizon, and demonstrating that the outperformance was due to the successful bidders ability to acquire cheap assets); cf. Tim Loughran & Anand M. Vijh, *Do Long-Term Shareholders Benefit from Corporate Acquisitions?*, 52 J. FIN. 1765, 1775 (1997) (finding that managers of acquiring firms use stock to pay for the acquisitions when their stock is likely to be overvalued and cash when their stock is likely to be undervalued).

79. To the extent the stock is overpriced and insiders can cash out some of their equity immediately, they will have an incentive to unwind this equity even as they seek to boost the value of their remaining shares by having the firm issue overpriced stock. See Daniel Bradley et al., *Do Insiders Practice What They Preach? Informed Option Exercises Around Acquisitions* 4–5 (Feb. 2009) (unpublished manuscript), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1364787 (finding that, around the announcement of stock-financed acquisitions, insiders of the acquiring firm exercise stock options and sell the underlying shares).

80. See Siew Hong Teoh et al., *Earnings Management and the Underperformance of Seasoned Equity Offerings*, 50 J. FIN. ECON. 63, 64–65 (1998) (reporting that seasoned equity issuers raise reported earnings by altering discretionary accruals and that this manipulation lowers post-offering returns); cf. Daniel A. Cohen & Paul Zarowin, *Accrual-Based and Real Earnings Management Activities Around Seasoned Equity Offerings* 4, 10 (June 2008) (unpublished manuscript), available at <http://ssrn.com/abstract=108193> (finding use of both accrual-based and real earnings management in a sample of 1,511 SEOs between 1987 and 2006).

81. See, e.g., Merle Erickson & Shiing-wu Wang, *Earnings Management by Acquiring Firms in Stock for Stock Mergers*, 27 J. ACCT. & ECON. 149, 151 (1999) (finding, in a sample of stock-financed mergers between 1985 and 1990, that acquirers managed earnings upward before announcing the merger); Henock Louis, *Earnings Management and the Market Performance of Acquiring Firms*, 74 J. FIN. ECON. 121, 134, 136 tbl.4 (2004) (finding that acquiring firms overstate earnings prior to stock-for-stock acquisitions).

projects outside of the firm. Suppose, for example, that the \$100 raised by an equity issuance would generate a return of 10% outside of the firm. Suppose further that, if invested in the firm, the \$100 would generate returns of 5%. The \$100 equity issuance would thus destroy \$5 of social value.

Substantial evidence suggests that expansions facilitated by equity issuances can reduce social value. For example, acquisitions—many of which are financed by the acquirer issuing stock—frequently reduce the aggregate wealth of both acquirer and target shareholders.⁸² One study found that during the period from 1998 to 2001, the combined value of acquirer and target stock fell over \$100 billion following acquisition announcements.⁸³

It is well-known that much of this value destruction occurs when firms have overpriced stock to pay for their acquisitions.⁸⁴ One oft-cited example of a value-destroying acquisition financed by overpriced equity is America Online's (AOL) acquisition of Time Warner in 2000.⁸⁵ AOL, with a market capitalization of over \$200 billion, used \$162 billion of its own stock to acquire Time Warner.⁸⁶ Whatever or not AOL's executives expected the merger to generate synergy benefits, it failed to do so. Because the companies were worth more separated than together, AOL and Time Warner parted ways nine years later.⁸⁷ When AOL was spun off, it was worth \$3.5 billion while Time Warner was valued at about \$36 billion.⁸⁸

Although the merger was a bust, AOL's original shareholders would have benefitted substantially from the transaction if they were planning to hold their shares for the long term. Had AOL not acquired Time Warner, AOL shareholders would have seen the value of their AOL shares decline from over \$200 billion to several billion dollars over the next decade.

82. 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, 759 (2005) (finding that decreases in acquirer shareholder value were not due to wealth transfers from acquiring shareholders to target shareholders because the aggregate value of both the acquirer and the target decreased significantly).

83. *Id.* To the extent that some of the decrease in the acquirer stock price following acquisition announcements is due to the offer signaling that the acquirer is overpriced, not all of this loss in shareholder value necessarily represents a destruction of social value.

84. See Michael C. Jensen, *Agency Costs of Overvalued Equity*, 34 FIN. MGMT. 5, 10 (2005) (arguing that managers of firms with overpriced stock make poor acquisitions in part to buy assets cheaply).

85. See Tim Arango, *How the AOL-Time Warner Merger Went So Wrong*, N.Y. TIMES, Jan. 11, 2010, http://www.nytimes.com/2010/01/11/business/media/11merger.html?_r=1 (reporting that the 2000 deal valued the combined firm at \$350 billion and that ten years later the combined values of the companies, which have since been separated, was about one-seventh of their combined values on the day of the merger).

86. See Daniel Okrent, *Happily Ever After?*, TIME, Jan. 24, 2000, at 39, 39 (reporting that the transaction was an all-stock acquisition for about \$162 billion of AOL stock).

87. See W. David Garnder, *AOL Completes Spin-Off From Time Warner*, INFORMATION WEEK (Dec. 10, 2009), <http://www.informationweek.com/news/internet/ebusiness/showArticle.jhtml?articleID=222001597>.

88. *Id.*

Instead, AOL shareholders ended up owning a large fraction of Time Warner, which ten years later had a market capitalization more than ten times that of AOL. Thus, acquisitions using high-priced stock can boost the long-term value of the acquiring firm's shares even if the acquisitions end up destroying social value.

2. *Why Not Sell Overpriced Shares and Distribute the Cash to Shareholders?*—A firm that sells inflated-price equity could, in principle, simply distribute the sale proceeds to its shareholders rather than invest the proceeds in unproductive activities. Indeed, the firm's current shareholders and its executives would be much better off if the cash raised by an inflated-price offering could be used to generate high returns outside of the firm rather than low returns inside the firm.

Why then would executives ever use the cash raised from an inflated-price offering to make unproductive investments in the firm? The reason is simple: a firm selling equity must inform its old and new investors of the purpose of the financing.⁸⁹ If the firm announces that it will take all of the funds raised and hold them in cash or distribute them to shareholders, investors are likely to infer that the firm is issuing stock merely to exploit the overpricing of its stock.⁹⁰ Investors may thus refuse to purchase shares, preventing the firm from selling overpriced equity.

Because firms selling equity cannot simply distribute the cash to investors, firms generally must use the funds for some other purpose, such as increasing investment.⁹¹ Thus, if the firm has poor investment opportunities, the only way to sell overpriced stock may be to use the proceeds for value-wasting investments. Unfortunately, as we will see shortly, executives holding stock in their firms can be rewarded for such value-wasting investments.

D. Rewards for Costly Expansion

Having seen that equity issuances can reduce social value, we will now see that executives whose payoff is tied to the stock's future value can be rewarded by engaging in inflated-price equity issuances that reduce social

89. See, e.g., SEC, Registration Statement Under the Securities Act of 1933, at 10 (Form S-3), available at www.sec.gov/about/forms/forms-3.pdf (requiring the issuer to furnish the information required by Item 504 of Regulation S-K, namely the "principal purposes for the which the proceeds are to be used").

90. See, e.g., Merritt B. Fox, *Civil Liability and Mandatory Disclosure*, 109 COLUM. L. REV. 237, 262 & n.65 (2009) (observing that the distribution of proceeds from a new equity sale would signal that the issuer sold the shares simply because it believed the stock was overpriced).

91. See Woojin Kim & Michael S. Weisbach, *Motivations for Public Equity Offers: An International Perspective*, 87 J. FIN. ECON. 281, 283 (2008) (reporting that equity offerings are done both to raise investment capital and to exploit favorable market conditions); Ming Dong et al., *Stock Market Misvaluation and Corporate Investment* 4 (Munich Personal RePEc Archive, Paper No. 3109, 2007), available at <http://mpra.ub.uni-muenchen.de/3109> (finding that cash raised by overpriced firms issuing equity is used to increase investment).

value. I will then explain that other components of executives' pay packages do not mitigate this distortion.

1. Equity Pay and Costly Expansion.— Executives whose payoffs are tied to the stock's future value can be rewarded for engaging in inflated-price equity issuances, even if those issuances reduce social value.

Return again to ABC Corporation. As before, it has two shares outstanding, one held by CEO and one held by public shareholders, and is liquidated at Liquidation Date.⁹² Consider two scenarios:

No-Transaction Scenario: Suppose that if ABC does not sell any of its equity prior to Liquidation Date, it will distribute \$20 to the holders of its two shares at Liquidation Date. The no-transaction value of each of ABC's two shares at Liquidation Date is thus \$10.

Equity Issuance Scenario: Now suppose that ABC can conduct an equity issuance before Liquidation Date when the stock trades at \$14 (\$4 more than its actual value of \$10), selling a third share at that price. Assume that the \$14 received by ABC increases ABC's Liquidation Date value from \$20 to \$33 because \$1 of social value is destroyed by the transaction. At Liquidation Date, the value of each of ABC's three shares (including that held by CEO) is thus \$11.

It should be easy to see that the inflated-price issuance boosts CEO's payout while reducing social value. In the No-Transaction Scenario, ABC's social value is \$20. In the Equity Issuance Scenario, it is \$19 (\$33 distributed to shareholders at Liquidation Date less the \$14 raised from investors).⁹³ But CEO's payout in the Equity Issuance Scenario is \$1 higher (\$11 rather than \$10). The net effect of the inflated-price equity sale on social value and CEO's payoff can be summarized in Table 4 below:

Table 4. Rewards for Costly Expansion

	Social Value	CEO Payoff
No Equity Issuance	\$20	\$10
Equity Issuance	\$19	\$11

As one can see, the inflated-price equity issuance rewards CEO even though social value is reduced. Thus, the equity given to CEO can incentivize her to direct the firm to sell shares even when the transaction destroys social value.

92. I assume that ABC does not issue any dividends (or repurchase any equity) before Liquidation Date.

93. In the No-Transaction Scenario, all \$20 flows to shareholders at Liquidation Date. In the Equity Issuance Scenario, \$14 flows from shareholders during the equity issuance and \$33 flows back to shareholders at Liquidation Date (for a net amount of \$19).

The problem is that tying an executive's payoff to the stock's future value aligns the executive's interests with those of nontrading shareholders but not with those of investing shareholders who buy additional stock from the firm. Thus, the executive has an incentive to take steps that transfer value from investing shareholders to nontrading shareholders even when such steps would destroy social value. And, as in the case of costly repurchases, this distortion arises even if the executive's payoff is tied to the stock's long-term value.

2. *Do Other Pay Components Mitigate?*—As we have seen, an executive whose payoff is tied to the stock's future value can be rewarded for engaging in value-wasting equity issuances when the stock is overpriced. However, equity comprises only part of an executive's compensation. The executive likely also receives a cash salary and bonus. In addition, the executive may hold stock that she can unload currently. Do these other forms of compensation mitigate the adverse incentives created by the equity held by the executive?

Unfortunately, none of these other components of the executive's pay package undermines executives' incentive to engage in costly expansion when the stock is overpriced. Indeed, given the well-known correlation between market capitalization and executive pay, salary and other forms of compensation may well rise if the executive expands the firm.⁹⁴

What about any equity that the executive is free to unwind currently? The executive will have an incentive to sell the unwindable stock and still conduct an overpriced equity issuance to boost the value of her remaining shares. In fact, there is evidence that executives whose firms are selling overpriced stock simultaneously unload some of their own shares.⁹⁵ In short, the incentive to engage in costly expansion is not weakened by other components of executives' pay packages.

IV. The Constant-Share Approach

Parts II and III demonstrated that executives whose payoffs are tied to the future value of their firms' shares can be rewarded for engaging in both stock repurchases and equity issuances that reduce social value. This problem arises because such an arrangement fails to tie executives' payoffs to aggregate shareholder value—the amount of value flowing to all of the firm's shareholders over time. Rather, it ties executives' payoffs to the value flowing to one subset of shareholders—nontrading shareholders—who neither sell any of their stock to the firm nor buy any additional stock from the firm before the executives receive the value of their stock.

94. Lucian Bebchuk & Yaniv Grinstein, *Firm Expansion and CEO Pay* 2–3 (Harvard John M. Olin Discussion Paper Series, Discussion Paper No. 533, 2007), available at http://papers.ssrn.com/abstract_id=838245.

95. See *supra* note 88.

In this Part, I explain how equity pay could be structured to reward executives for repurchases and equity issuances if and only if these transactions increase social value. Subpart A puts forward and provides an overview of a “constant-share” approach to equity-based pay that would tie executives’ payoffs to aggregate shareholder value—the value flowing to nontrading, redeeming, and investing shareholders—and thereby eliminate executives’ incentives to engage in value-reducing repurchases and equity issuances merely to boost the stock’s future value. Subpart B describes in more detail how the constant-share approach operates in the context of a repurchase. Subpart C describes how the approach operates in connection with an equity issuance.

My main goal in this Part is to explain the concept behind the constant-share approach rather than put forward a fully fleshed-out proposal for implementing it. Thus, in explaining the conceptual underpinnings of the constant-share approach, I make some simplifying assumptions. In particular, I assume the executive owns vested stock subject to a holding requirement (and no other equity in the firm), is risk-neutral, has sufficient liquidity to purchase additional shares, and cannot use hedging or other techniques to circumvent the constant-share arrangement. I also ignore tax considerations and the time value of money. A mathematical model of the constant-share approach can be found in the Appendix.

A. Description of the Approach

Under the constant-share approach, executives would be required to adjust their equity positions whenever the firm repurchases or issues shares such that executives’ fractional ownership in the firm remains constant throughout the transaction. Thus, when the firm repurchases shares, the executive would be required to sell some of her shares to the firm. And when the firm issues shares, the executive would be required to buy additional shares.

Whether the firm buys or sells its own stock, the executive would transact with the firm on the same terms as the firm transacts with other investors. Thus, when the firm repurchases shares, the executive must sell shares to the firm at the repurchase price. Similarly, the executive must buy shares at the issue price when the firm sells shares.

For example, suppose that an executive (CEO) holds a certain fraction of the firm’s equity at a particular point in time (say 2%). If the firm repurchases 10% of its shares, CEO would be required to sell to the firm, at the same price the firm was paying for the repurchased shares, 10% of her 2% block, or 0.2% of the firm’s shares. Similarly, if the firm increases its outstanding shares by 10% in a seasoned offering, CEO would be required to buy, at the same price the firm was receiving for the newly issued stock, an amount equal to 10% of her 2% block, or 0.2% of the firm’s outstanding shares. The effect of CEO’s transactions would be to leave CEO with the

same fraction of the firm's outstanding shares after the repurchase or share issuance as before—in this case, 2%.⁹⁶

As I explain in more detail below, requiring the executive to participate in repurchases as a redeeming shareholder and in equity issuances as an investing shareholder in the same proportion as she owns stock in the company ensures that the executive's equity payoff is tied to the value flowing to all of the firm's shareholders, not just the value flowing to nontrading shareholders. Essentially, the constant-share approach would make CEO both a nontrading and redeeming shareholder in the case of repurchases and both a nontrading and investing shareholder in the case of equity issuances. Thus, the executive no longer financially benefits from repurchases and stock issuances that merely transfer wealth from one set of shareholders to another without generating any social value. Instead, the executive has an incentive to engage in repurchases and equity issuances only if they increase aggregate shareholder value.

The constant-share approach provides another benefit in addition to eliminating the incentive to engage in costly contractions and costly expansions. As Parts II and III explained, executives frequently manipulate the stock price around repurchases and equity issuances. They engage in such manipulation because it boosts the value of their equity. Once executives have decided to conduct a repurchase, driving down the stock price before the firm starts buying back stock increases the value of executives' equity because the executives are indirect buyers of the repurchased stock. Similarly, assuming an equity issuance will occur, driving up the stock price makes executives holding stock better off because they are indirect sellers of the issued equity. The constant-share approach would eliminate executives' equity-driven incentive to engage in such manipulation by ensuring that executives are neither indirect buyers nor indirect sellers in these transactions.

96. Although I assume in the text that CEO sells and purchases shares whenever the firm transacts in its own stock, CEO's adjustments could be effected without the purchase or sale of actual shares. For example, if CEO owns 2% of the firm's shares and the firm repurchases 1,000,000 shares, CEO could enter into an agreement with the firm to swap, at some point in the future, the then-current value of 20,000 of her shares to the firm for their value at the time of the repurchase price plus an amount accounting for the time value of the money. Similarly, if the firm issues 1,000,000 new shares, CEO could be required to swap the value of 20,000 shares at the issuance price plus interest (or something analogous) for the then-current value of the shares.

Moreover, the adjustments required by the constant-share approach need not be made every time the firm transacts in its own stock. Rather, the firm could track its repurchases and equity offerings each year (and CEO's positions on the eve of each of these transactions) and, at year end, require CEO to engage in a single transaction with the firm that leaves CEO in the same position as contemporaneous adjustments. The Appendix explains how an *ex post* adjustment would be effected.

B. Constant-Share Approach in a Repurchasing Firm

When a firm repurchases shares, the constant-share approach ties an executive's equity payoff to the total amount of value flowing to both nontrading and redeeming shareholders. This mechanism eliminates the executive's equity-driven incentive to engage in a repurchase merely because it transfers value from redeeming shareholders. Instead, the executive is rewarded for engaging in a repurchase if and only if the repurchase increases aggregate shareholder value.

To see why this is the case, suppose that an executive (CEO), at the time of a possible repurchase, owns 10% of the firm's stock. Suppose that the firm is considering purchasing 20% of the outstanding shares at the current trading price.

Absent the constant-share mechanism, CEO would own 12.5% of the firm's shares after the repurchase.⁹⁷ Thus, CEO would receive 12.5% of the value flowing to nontrading shareholders—the shareholders who do not redeem their shares in the repurchase—and 0% of the value flowing to redeeming shareholders. CEO's equity payoff would thus reflect only the value flowing to nontrading shareholders rather than to all of the firm's shareholders affected by the transaction. As we saw in Part II, CEO would be rewarded for engaging in bargain-price repurchases even if they reduce social value.

Under the constant-share approach, CEO would be required to participate in the 20% repurchase in an amount proportionate to her pre-transaction ownership interest of the firm (10%). Thus, shares sold by CEO to the firm would constitute 10% of the 20% block acquired by the firm (or 2% of the firm's outstanding stock). Put another way, because the firm is repurchasing 20% of all of its shares, CEO is required to sell to the firm 20% of her 10% interest.

After the repurchase, CEO would continue to own 10% of the firm's equity. Thus, CEO would receive 10% of the value flowing to nontrading shareholders. But because CEO also held 10% of the equity repurchased by the firm, CEO would receive 10% of the value flowing to redeeming shareholders. As a result, CEO's payoff would equal 10% of the value flowing to all of the firm's shareholders affected by the transaction. Thus, CEO has an incentive to conduct repurchases if and only if the repurchases increase aggregate shareholder value.

C. Constant-Share Approach in a Share-Issuing Firm

I now turn to consider how the constant-share approach operates when a firm issues equity. As I will show, the constant-share approach ties an

97. Because the firm is repurchasing 20% of its stock, the proportional interest of each remaining shareholder, including CEO, will increase by 25%.

executive's payoff to the total amount of value flowing to both nontrading and investing shareholders. This mechanism eliminates the executive's incentive to engage in an equity issuance merely because the issuance transfers value from investing shareholders. Instead, the executive will be rewarded for engaging in any equity issuance that actually increases aggregate shareholder value.

To see why this is the case, suppose that an executive (CEO), at the time of a possible equity issuance, owns 10% of the firm's stock. Suppose that the firm is considering issuing an amount of stock equal to 25% of the outstanding shares at the current trading price.

Absent a constant-share mechanism, CEO would own 8% of the firm's stock after the transaction.⁹⁸ Thus, CEO would receive 8% of the value flowing to nontrading shareholders--the shareholders who do not invest in the equity issuance--and 0% of the value flowing to investing shareholders. CEO's equity payoff would thus reflect the value flowing to nontrading shareholders rather than to all of the firm's shareholders affected by the transaction. As we saw in Part III, CEO would be rewarded for engaging in inflated-price issuances even if they reduce aggregate shareholder value.

Under the constant-share approach, CEO would be required to buy 10% of the shares sold by the firm. Because the firm is selling an amount of shares equal to 25% of its pre-sale outstanding equity, CEO would be required to buy 2.5% of the pre-sale equity. That is, CEO would be required to increase her 10% stock ownership by 25%, the proportion by which the equity offering increases the firm's outstanding shares.

After the equity issuance, CEO would continue to own 10% of the firm's equity. Thus, CEO would receive 10% of the value flowing to nontrading shareholders. But because CEO also buys 10% of the equity issued by the firm, CEO receives 10% of the value flowing to investing shareholders. As a result, CEO's payoff equals 10% of the value flowing to all of the firm's shareholders affected by the transaction. Thus, CEO has an incentive to conduct the equity issuance if and only if the issuance increases aggregate shareholder value.⁹⁹

98. Because the firm is issuing an amount of equity equal to 25% of its pre-transaction outstanding shares, the proportional interest of each remaining shareholder, including the CEO, will drop by 20%.

99. I am assuming that CEO is risk-neutral. If CEO is risk-averse, she may be deterred from engaging in value-increasing equity-financed expansions if she must buy enough equity to maintain her proportional ownership of the firm. For such an executive, a requirement that she purchase somewhat less equity than is needed to maintain her percentage ownership when the firm issues stock may both deter costly expansions and facilitate value-creating expansions.

V. Conclusion

Tying executive pay to the future value of the firm's stock is widely viewed as a useful means of incentivizing executives to generate value for public shareholders. This Article has identified a potential flaw with such equity-based pay arrangements, even if they tie the executive's payoff to the stock's long-term value. In particular, tying executive pay to the stock's future value can reward executives for engaging in share repurchases and equity issuances that reduce what I term "aggregate shareholder value"—the amount of cash flowing to all of a firm's shareholders over time.

The Article has also put forward a new "constant-share" approach to equity-based compensation that ties executives' equity payoffs to aggregate shareholder value. Under this approach, executives would be required to sell some of their shares (or buy additional shares) whenever the firm repurchases its own stock (or issues new equity) so that the executives' proportional ownership in the firm remains constant as the firm transacts in its own stock. The Article showed that the constant-share approach would eliminate executives' incentives to engage in share repurchases and equity issuances that reduce aggregate shareholder value.

I hope that the analysis I have offered will help sharpen understanding of the potentially negative effects of tying executive pay to the future value of the firm's stock—even its long-term value—and assist regulators, directors, and shareholders in improving executive compensation and corporate governance in public companies.

APPENDIX

The Appendix introduces an analytical framework for examining the effect of equity compensation on executives' incentives to engage in repurchases and equity issuances and also introduces the operation of the constant-share approach.

Analytical Framework

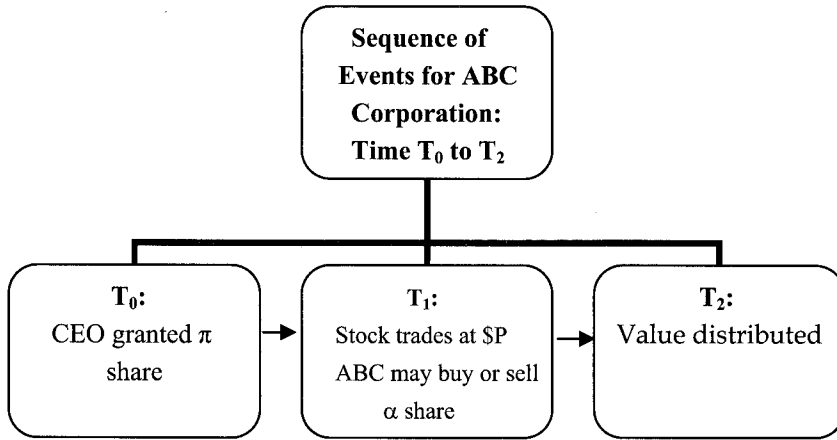
Consider a Corporation (ABC) that initially has a single share outstanding and exists in three sequential periods: Time $T=0$, Time $T=1$, and Time $T=2$.

- At $T=0$, ABC has a single risk-neutral manager (CEO) who is granted a fraction, π , of ABC's equity, which she must hold until $T=2$. In analyzing CEO's incentives, I assume CEO seeks solely to maximize the dollar payoff from her compensation arrangement.
- At $T=1$, ABC's share (or a fraction of it) can be traded for a price, P_1 , per share. At $T=1$, ABC may or may not repurchase or issue an additional amount of equity equal to a fraction, α , of its single share.
- At $T=2$, ABC is liquidated and its value is distributed pro rata to its shareholders.

ABC's $T=2$ value will depend on whether there has been a transaction in ABC's stock at $T=1$. In the absence of any transactions in the firm's stock (such as a repurchase or sale of equity), ABC's $T=2$ value is V .

If there is a repurchase (or sale) of equity, ABC's $T=2$ value will be reduced (increased) by the amount paid (received) for any stock repurchased (sold) at $T=1$ plus an amount, X , representing the efficiency effects of the transaction on ABC's value. From an economic perspective, ABC should repurchase equity or issue equity if and only if (iff) $X > 0$.

Figure 1. Sequence of Events for ABC Corporation



Social Value and Final-Period Stock Value. Social value (SV) is the net amount of value flowing from ABC to its shareholders between $T=0$ and $T=2$. The final-period stock value is ABC's $T=2$ value, divided by the number of shares outstanding at $T=2$.

No-Transaction Scenario. Denote SV_n as ABC's social value when ABC neither repurchases nor issues equity at $T=1$. It should be easy to see that

$$(1a) \quad SV_n = V.$$

Denote P_{2n} as the final-period stock value if there is no repurchase or equity issuance. It should also be easy to see that

$$(2a) \quad P_{2n} = V.$$

Repurchase Scenario. Denote SV_r as ABC's social value when ABC repurchases α share at $T=1$ for a price P_1 , and the repurchase changes ABC's value by X . It follows that

$$(1b) \quad SV_r = \alpha P_1 + (V - \alpha P_1 + X) = (V + X).$$

Denote P_{2r} as the final-period stock value if there is a repurchase.

$$(2b) \quad P_{2r} = (V - \alpha P_1 + X)/(1 - \alpha).$$

Equity-Issuance Scenario. Denote SV_i as ABC's social value when ABC issues α share at $T=1$ for a price P_1 , which changes ABC's value by X .

$$(1c) SV_1 = (V + \alpha P_1 + X) - \alpha P_1 = (V + X).$$

Denote P_{2i} as the final-period stock value if there is an equity issuance.

$$(2c) P_{2i} = (V + \alpha P_1 + X)/(1 + \alpha).$$

SV and the final-period stock value for each scenario are summarized in the table below.

Table 5. Social Value and Stock Value

	Social Value	Stock Value
No transaction	V	V
Repurchase	V+X	$(V - \alpha P_1 + X)/(1 - \alpha)$
Equity issuance	V+X	$(V + \alpha P_1 + X)/(1 + \alpha)$

CEO's Incentive to Engage in Costly Repurchases

Consider CEO's incentive to repurchase at $T=1$ when, as assumed previously, her payoff is tied to the $T=2$ stock value. Given CEO's incentive to maximize the $T=2$ stock value, it follows from (2a) and (2b) that CEO will repurchase at $T=1$ iff

$$(3) (V - \alpha P_1 + X)/(1 - \alpha) > V.$$

Simplifying (3) yields

$$(4) V - P_1 > -X/\alpha.$$

It follows from (4) that CEO has an incentive to engage in a costly (value-reducing) repurchase when

$$(5) 0 > X > \alpha(P_1 - V).$$

Remark: It should be easy to see from (5) that if $P_1 \geq V$ (the stock is either properly priced or overpriced at $T=1$), CEO does not have an incentive to conduct a costly repurchase. However, if $P_1 < V$ (the stock is underpriced at $T=1$), CEO may have an interest in conducting a costly repurchase.

CEO's Incentive to Engage in Costly Expansion

Now consider CEO's incentive to issue equity at $T=1$ when, as assumed previously, her payoff is tied to the $T=2$ stock value. Given CEO's incentive to maximize the $T=2$ stock value, it follows from (2a) and (2c) that CEO will issue equity at $T=1$ iff

$$(6) (V + \alpha P_1 + X)/(1 + \alpha) > V.$$

Simplifying (6) yields

$$(7) P_1 - V > -X/\alpha.$$

It follows from (7) that CEO has an incentive to engage in a costly (value-wasting) equity issuance when

$$(8) 0 > X > \alpha(V - P_1).$$

Remark: It should be easy to see from (8) that if $P_1 \leq V$ (the stock is either properly priced or underpriced in the short term), CEO does not have an incentive to engage in costly expansion. However, if $P_1 > V$ (the stock is overpriced in the short term), CEO may benefit from engaging in costly expansion.

Constant-Share Approach

Under the constant-share approach, CEO must participate in a repurchase (issuance) by selling (buying) a fraction of the shares purchased (sold) by the company equal to her pre-transaction percentage interest in ABC, π . This approach incentivizes CEO to conduct a repurchase or equity issuance if and only if it increases social value.

Denote the CEO's payoff if there is no repurchase as W_n , if there is a repurchase as W_r , and if there is a stock issuance as W_i .

If there is no repurchase or equity issuance, it should be easy to see from (2a) that

$$(9a) W_n = \pi V.$$

If there is a repurchase of α share at $T=1$ for price P_1 , CEO will be required to sell $\pi\alpha$ share at price P_1 to ABC. CEO will receive $\pi\alpha P_1$ at $T=1$ and be left with $\pi(1 - \alpha)$ share at $T=2$. It follows from this and (2a) that

$$(9b) W_r = \pi\alpha P_1 + \pi(1 - \alpha)(V - \alpha P_1 + X)/(1 - \alpha) = \pi(V + X).$$

If there is an issuance of equity at $T=1$, CEO will pay $\pi\alpha P_1$ and be left with $\pi(1 + \alpha)$ share at $T=2$. It follows from this and (2c) that

$$(9c) W_i = -\pi\alpha P_1 + \pi(1 + \alpha)(V + \alpha P_1 + X)/(1 + \alpha) = \pi(V + X).$$

The CEO's payoff under each scenario can be summarized in the following table.

Table 6: Social Value and CEO Payoff Under the Constant-Share Approach

	Shareholder Value	CEO Payoff
No transaction	V	πV
Repurchase	V+X	$\pi(V+X)$
Equity issuance	V+X	$\pi(V+X)$

Thus, CEO has an incentive to undertake a repurchase or equity issuance iff $X > 0$. That is, CEO would undertake the transaction if and only if it increases social value.

Ex Post Implementation

Above, it was assumed that the CEO would participate pro rata in any equity transaction at $T=1$. Thus, because CEO owns π of ABC's equity, she would sell or buy $\pi\alpha$ share when ABC buys or sells α of its equity. Under this constant-share approach, CEO's payoff is tied to social value.

However, CEO's payoff could be tied to shareholder value through the use of an ex post adjustment made to CEO's position *after* the equity transaction took place. Denote as $\beta\pi$ the amount of shares CEO must sell (or buy) after ABC's repurchase (or equity offering) in order to tie CEO's payoff to social value. Such an adjustment would change the amount of equity outstanding and ABC's value.

Suppose ABC repurchases α share at $T=1$ for price P_1 , and then CEO sells $\beta\pi$ share for price P_1 .

ABC's $T=2$ value will be $V - [\alpha + \beta\pi]P_1 + X$, and it will have $1 - \alpha - \beta\pi$ outstanding share.

Denote CEO's payoff in a repurchase when there is an ex post adjustment as W_r' . It follows that

$$(10) W_r' = \beta\pi P_1 + \pi(V - [\alpha + \beta\pi]P_1 + X)/(1 - \alpha - \beta\pi).$$

Aligning CEO's payoff with social value requires that $W_r' = W_r$, which in turn implies that

$$(11) W_r' = \pi(V + X).$$

From (10) and (11), it follows that

$$(12) \beta = \alpha/(1 - \pi).$$

It can easily be shown that the post-transaction adjustment in the case of an equity issuance is the same. CEO must buy $\beta\pi$ share at P_1 , where $\beta = \alpha/(1 - \pi)$.

* * *

Book Review

But I Thought the Earth Belonged to the Living

DEAD HANDS: A SOCIAL HISTORY OF WILLS, TRUSTS, AND INHERITANCE LAW. By Lawrence M. Friedman. Stanford, California: Stanford University Press, 2009. 240 pages. \$60.00 (cloth), \$22.95 (paper).

Mark L. Ascher*

Thomas Jefferson famously wrote that “*the earth belongs . . . to the living.*”¹ Sadly, the statement has never been an accurate restatement of the law, either here or in England, as Anglo-American law has long deferred to the Dead Hand. The law of wills allows the dead virtually unfettered discretion in divvying up that which used to belong to them, and probate law generally requires the living to effectuate those desires. The law of trusts allows both the living and the dead to create trusts for the disposition of property over extended periods of time, including long after their death, and the law of fiduciary administration requires the living to manage and enforce these trusts. But Jefferson certainly was not trying to restate the law of the dead. Instead, his statement seems primarily aspirational, encouraging us never to let the control we accord the Dead Hand get completely out of hand. There is also a note of admonition, imploring us never to allow our laws to be used in the establishment of an aristocracy.

For the last several decades, certain trends in the law of the dead have threatened to put us sharply at odds with Jefferson’s vision. Though each of these trends has been legislative, the changes they embody are not the product of, nor can they withstand, serious policy analysis. Instead, they are the product of intense and well-placed lobbying by a few particularly motivated special interest groups. As time goes by, and as more and more people reflect on what has happened, these trends may, therefore, eventually fizzle out. Indeed, there is some evidence that this has already begun to occur. But if these trends continue, and if the changes they embody prove enduring, my own view is that we will have broken faith with one of our most sacred

* Joseph D. Jamail Centennial Chair in Law, The University of Texas School of Law. B.A., Marquette University; M.A., Kansas State University; J.D., Harvard University; L.L.M. (in Taxation), New York University. Thanks to Professors Jack Getman and John Langbein for their thoughtful comments on prior versions of this Review.

1. Letter from Thomas Jefferson to James Madison (Sept. 6, 1789) (emphasis added), in 3 MEMOIRS, CORRESPONDENCE, AND PRIVATE PAPERS OF THOMAS JEFFERSON 28, 31 (Thomas Jefferson Randolph ed., 1829); see also 1 THOMAS PAINE, *Rights of Man*, in THE COMPLETE WRITINGS OF THOMAS PAINE 241, 251 (Philip S. Foner ed., 1945).

principles, egalitarianism, and that we may even have unleashed developments that could someday threaten the viability of republican government as we know it.

Seemingly written in response to these trends, a trio of intriguing new books takes up the reach and longevity of the Dead Hand. Each in its own way concludes that our laws now cater, as never before, to the wishes of the dead. In a field that is supposedly as moribund as law ever gets, the publication of three new books on essentially the same topic at essentially the same time strongly suggests that something significant is, indeed, afoot. Interestingly, the books disagree, at least to some extent, as to why the trends have developed. Somewhat surprisingly, they also disagree as to the changes' significance.

This Review takes a close look at one of these books, *Dead Hands: A Social History of Wills, Trusts, and Inheritance Law*,² by Professor Lawrence M. Friedman. Its account of the law of the dead is as close to a toe-tapper as I've ever read. It's a deceptively easy read, however, because Professor Friedman's considerable skills as an author often mask the complexity of the underlying concepts. It does, however, stake out several positions as to which there is ample room for debate. In considering these positions, this Review borrows heavily from the other two books: *From Here to Eternity? Property and the Dead Hand*,³ by Professor Ronald Chester, and *Immortality and the Law: The Rising Power of the American Dead*,⁴ by Professor Ray D. Madoff.

Professor Friedman styles *Dead Hands* "A Social History of Wills, Trusts, and Inheritance Law." Yet it often reads less like history and more like an inviting study guide.⁵ In the Introduction, Professor Friedman identifies four "leitmotifs." The first is changes in family structure: "It has shifted emphasis from what we might call the *bloodline family* to the *family of affection and dependence*."⁶ These changes, he argues, help to account for rather substantial changes in the law of intestate succession, for example.⁷ The second is that we live in an "age of elaborate record keeping, computers, data bases, and vibrant bureaucracies."⁸ So nonprobate transfers are not so risky, and wills no longer rule the roost. The third is "demographic and

2. LAWRENCE M. FRIEDMAN, *DEAD HANDS: A SOCIAL HISTORY OF WILLS, TRUSTS, AND INHERITANCE LAW* (2009).

3. RONALD CHESTER, *FROM HERE TO ETERNITY? PROPERTY AND THE DEAD HAND* (2007).

4. RAY D. MADOFF, *IMMORTALITY AND THE LAW: THE RISING POWER OF THE AMERICAN DEAD* (2010).

5. It nicely tracks the usual Estates and Trusts course from intestate succession to the transfer taxes, and it does so with such insight and infectious enthusiasm that I am seriously considering recommending it as optional reading.

6. FRIEDMAN, *supra* note 2, at 11.

7. *Id.* at 12.

8. *Id.* at 13.

cultural changes,”⁹ not least of which is that we live a lot longer than our ancestors. These changes affect not only when our heirs will receive their inheritance but also what will be left for them. The fourth is “attitudes in society toward wealth and the wealthy.”¹⁰ Though “America was born in revolt against a system of inherited, dynastic wealth,” our laws, according to Professor Friedman, now “*favor[]* dynastic wealth.”¹¹

Let us start with the longest and most elaborate chapter, “Distribution After Death”—basically intestate succession. Professor Friedman promptly identifies two changes of “particular importance.”¹² The first is “the merger of rules about real estate and rules about personal property.”¹³ The second is “the growing share of the estate that goes to a surviving spouse.”¹⁴ Both, he asserts, are connected to “the shift from emphasis on the *bloodline* family to the *family of affection and dependence*.”¹⁵

Professor Friedman’s primary illustration of the merger of the rules relating to real and personal property is our rejection of primogeniture. Primogeniture ensured that the decedent’s lands would pass intact to his oldest male heir. Other rules provided for a somewhat more equitable and far-flung transmission of personal property. The United States rejected primogeniture early on, in favor of partible inheritance, which typically results, after satisfaction of the (possibly greater) claims of the surviving spouse, in an equitable division (variously defined) among *all* of the decedent’s children and possibly also *their* issue. So it is true in this and other ways that the general trend has been a homogenization of the way we treat real and personal property at death.¹⁶ Yet merger of the rules relating to real and personal property strikes me as a change only a law professor could love. To me, the vastly more important change is our rejection of primogeniture in favor of a decidedly less dynastic and dramatically more egalitarian method of divvying up property at death. Nor is the switch from primogeniture to partible inheritance related in any significant way to changes in the family, unless one packs that phrase with a whole lot of baggage. In my view, the switch is instead a clear rejection of our English aristocratic roots, in favor of a more egalitarian and republican society.

9. *Id.*

10. *Id.* at 14.

11. *Id.*

12. *Id.* at 19.

13. *Id.*

14. *Id.*

15. *Id.*

16. See, e.g., UNIF. PROBATE CODE §§ 2-102 to 2-103 (amended 2008), 8 U.L.A. 38, 42 (Supp. 2010) (exemplifying the trend in rules concerning intestate succession). But see, e.g., TEX. PROB. CODE ANN. § 38(b) (West 2010) (prescribing different intestate succession rules for real and personal property); *id.* § 322B (requiring that bequests of personal property abate before devises of real property).

The second of Professor Friedman's particularly important changes, the transformation of the surviving spouse's entitlement, is undeniably true. It is also, in my view, the single most important change in the law of inheritance in centuries. In this respect, Professor Friedman's account seems right on the money. He notes that dower was once a widow's sole protection and that it involved only the decedent's land.¹⁷ Virtually every state, however, has replaced dower with either community property or an elective share, both of which apply to both real and personal property.¹⁸ Simultaneously, Professor Friedman traces the transition of the surviving spouse from being entitled under the law of intestate succession to essentially nothing, as recently as at the turn of the last century,¹⁹ to being the primary—and often exclusive—intestate taker.²⁰ Again, however, I fail to see how the movement of the surviving spouse to the head of the inheritance pack reflects a change in family structure.²¹ The dramatic increase in the surviving spouse's entitlement strikes me as almost entirely attributable to and emblematic of the ever-increasing stature of women in American society.²²

Professor Friedman is not alone in seeking to establish a link between various changes in inheritance law and changes in the family.²³ No doubt some such changes are best explainable on just such a basis. Adoption, for example, is mostly an American phenomenon; in England, it was essentially unknown.²⁴ Likewise, although the English no doubt knew a little something

17. FRIEDMAN, *supra* note 2, at 24.

18. RALPH C. BRASHIER, *INHERITANCE LAW AND THE EVOLVING FAMILY* 12–14 (2004) (noting that only a few states have retained dower, including Arkansas, Kentucky, Michigan, and Ohio).

19. *See, e.g.*, *Butler v. Sherwood*, 188 N.Y.S. 242, 243–44 (N.Y. App. Div. 1921) (holding that a wife's attempt to transfer both her farm and her personal property to her husband by deed effective upon her death was an invalid testamentary transfer), *aff'd per curiam*, 135 N.E. 957 (N.Y. 1922). Under the New York intestacy statute, because she was survived by a brother, her husband was entitled to at most two thousand dollars and one-half of the personal property. N.Y. DECEDENT EST. LAW §§ 98(3), 100 (Consol. 1909).

20. *See, e.g.*, UNIF. PROBATE CODE § 2-102 (specifying that the surviving spouse takes no less than the first \$150,000, as adjusted for inflation through 2008); FRIEDMAN, *supra* note 2, at 180–81 (“The surviving spouse (mostly a she) began our period in a very subordinate position. Now she bestrides the world of intestacy like a colossus.”).

21. Professor Friedman himself seems to sense this: “Marriage may be a weaker reed than it used to be, but you could not prove it through the intestacy laws.” FRIEDMAN, *supra* note 2, at 25.

22. *See* CHESTER, *supra* note 3, at 75 (attributing increases in spousal protection to the women's rights movement); Ariela R. Dubler, *In the Shadow of Marriage: Single Women and the Legal Construction of the Family and the State*, 112 *YALE L.J.* 1641, 1673–75 (2003) (noting a linkage between increases in spousal protection and the suffrage movement); Anne-Marie Rhodes, *Blood and Behavior*, 36 *ACTEC L.J.* 143, 168 (2010) (discussing increases in spousal protection in the context of the women's rights movement).

23. *See, e.g.*, Susan N. Gary, *Adapting Intestacy Laws to Changing Families*, 18 *LAW & INEQ.* 1, 6 (2000) (proposing changes in intestacy statutes to parallel changes in American family structures).

24. *See* FRIEDMAN, *supra* note 2, at 22 (noting that English law did not recognize adoptions).

about illegitimacy, we Americans seem to have perfected the practice.²⁵ Reshaping the law of inheritance to treat our adopted and nonmarital children essentially the same as we treat our other children seems intimately related to changes in how we conceive the family. These changes are also, however, matters of relative detail. In contrast, the biggest changes in the law of inheritance, our rejection of primogeniture and the dramatic increase in the surviving spouse's entitlement, appear to have very different roots.

Dead Hands also contains a lovely chapter on the ordinarily dreary law of wills, including their execution, revocation, and interpretation. Throughout, Professor Friedman carefully documents a decline in formality. He takes pains, however, to distance himself from those who would attribute these changes to "a broad revolt against formalism throughout the legal system":

[A]t least in *this* field—succession—I hate to pin the source of change on legal theory, or on some vague and general shift in ideology. . . . [A] change in legal culture calls for a specific explanation; and a vague reference to some sort of ideological change is simply not enough. And we must remember that not every branch of law is less formalistic than it used to be. . . . In the law of succession, what drives the motor of anti-formalism is perhaps the gradual but very marked change in the importance, and the role, of the classical will. It now has formidable rivals: the so-called will substitutes.²⁶

After a short chapter on will contests, Professor Friedman turns to will substitutes and continues his thought:

The will once had a virtual monopoly over gifts at death. This is no longer the case. The rise of will substitutes has, in turn, affected the law of wills itself. This is probably a key reason why the law of wills has become less formal and formalistic. After all, now one can draw up a document that looks like a will, sounds like a will, and acts like a will, but isn't a will. This document will not need witnesses, a ceremony, and the whole hocus-pocus surrounding the ordinary will. If so, then it is natural to ask if there is any point in insisting so rigidly on all the formalities and the hocus-pocus.²⁷

This is not only a breath of fresh air; it also has the ring of truth. I, too, have long been skeptical of enlisting the law of wills in a frontal assault on "formalism." There are good reasons for demanding at least some level of formalism when *the* most important party to the transaction is *always* unavailable for comment when it comes time to interpret the document. It

25. See MADOFF, *supra* note 4, at 61 ("The number of nonmarital children has reached epidemic proportions in the United States; almost 40 percent of all children are born out of wedlock.").

26. FRIEDMAN, *supra* note 2, at 80–81.

27. *Id.* at 100–01.

may also matter that the document in question has almost never been the product of negotiation. Ordinarily, the testator consults with only an attorney, whose primary motivation is to please the client. On other occasions, the testator seems to have consulted no one. Thus, comparisons with contract law, for example, have always struck me as perilous.²⁸ But Professor Friedman is certainly correct: “[I]n the twentieth century, will substitutes [did grow] like mushrooms.”²⁹ That this was also when hyperformalism in the law of wills began to subside may well be more than a coincidence.

It is in the chapter on trusts that issues relating to the reach and longevity of the Dead Hand begin to surface. According to Professor Friedman, “Legally speaking, over the past century and a half, the dynastic trust has won victory after victory.”³⁰ He points to several developments from the nineteenth century that he says paved the way for dynastic trusts. It is perhaps no accident that each of these developments originated in Massachusetts, where professional trustees first came to prominence. One development was the prudent-person rule governing trust investments.³¹ It assumes that most trusts are for the long haul; in any event, it accommodates investment practices appropriate for such trusts. Another was legitimization of so-called spendthrift trusts.³² These are trusts in which one or more of the beneficiaries’ interests are inalienable. In other words, the settlor of a trust, by adding a few more words to the trust document, can effectively render the beneficiaries’ interests off-limits to their creditors. Yet another was the so-called *Clafin* doctrine, under which the beneficiaries of a trust cannot by mutual consent terminate the trust ahead of time, even if they are all adult and competent, as long as any material purpose of the settlor remains unfulfilled.³³ The influence of each of these developments quickly spread,

28. Contracts of adhesion, outliers even in the world of contract law, are a particularly poor analogy. Such contracts may not serve all of the parties’ interests equally well, but ordinarily they serve at least one of the parties’ interests exceedingly well. In most instances, someone has thought them through *very* carefully, proofread them to distraction, and revised them over a series of transactions, if not over a period of years. Whatever the difficulties that may or may not result from the inequality of bargaining power and the fact that only one of the parties was responsible in any meaningful way for their content, contracts of adhesion generally work. Wills, by contrast, are *supposed to be* custom documents. Often they contain errors, both large and small. In all too many instances, they are simply train wrecks that do not work for *anyone*, except those who have nothing to lose in contesting them and those who litigate for a living.

29. FRIEDMAN, *supra* note 2, at 110.

30. *Id.* at 116.

31. The seminal case was *Harvard College v. Amory*, 26 Mass. (9 Pick.) 446 (1830).

32. The leading Massachusetts case was *Broadway National Bank v. Adams*, 133 Mass. 170 (1882), but the dictum in *Nichols v. Eaton*, 91 U.S. 716, 727 (1875), recognizing a settlor’s ability to protect transferred property “from the ills of life” surely had more impact.

33. See *Clafin v. Clafin*, 20 N.E. 454, 455–56 (Mass. 1889) (“The decision . . . rests upon the doctrine that a testator has a right to dispose of his own property with such restrictions and limitations, not repugnant to law, as he sees fit, and that his intentions ought to be carried out, unless they contravene some positive rule of law, or are against public policy.”).

and by the middle of the twentieth century, it permeated the country. The table was thus set, Professor Friedman argues, for the creation of dynastic trusts.

Why these developments occurred is much less clear. Certainly the prudent-person rule gave trustees additional investment flexibility. This may have increased settlors' confidence that the trust assets would be invested appropriately over time. In addition, creating a long-term trust no longer meant taking property off the market indefinitely. Likewise, the ability to insulate the beneficiaries' interests from their creditors and the knowledge that the beneficiaries could not by themselves easily terminate the trust prior to the stated termination date may have steered settlors in their dynastic intentions. None of these developments, however, is inherently dynastic; each can and far more frequently does apply to trusts of much shorter duration.³⁴ Nonetheless, based in part on these developments, Professor Friedman asserts that there had already occurred in this country a change in attitude toward wealth generally and toward dynastic wealth in particular. Though we had begun our national existence deeply skeptical of such wealth, we had by the end of the nineteenth century become much more accepting of it.³⁵ I am not so sure. We had not yet witnessed the flowering of the Progressive Movement. The Sixteenth Amendment and the progressive income tax were still to come. So were both the estate tax and the gift tax. Our national interest in and willingness to enact each of these taxes strongly suggests that we had not yet lost our suspicion of dynastic wealth. It makes sense that these developments would originate in Massachusetts, where trusts, and administering them, first became big business. So it also makes sense that the Massachusetts courts would stake out positions that would accommodate such arrangements. But it does not ring true that there had already been a shift in attitude toward dynastic wealth, in either Massachusetts or the country at large. The rapid spread of these developments may say much less about our collective attitude toward dynastic wealth than about the crucial role that *stare decisis* played as we pursued our Manifest Destiny across a vast and mostly vacant continent, leaving in our wake dozens of brand-new and semi-autonomous legal systems.³⁶ When

34. See, e.g., FRIEDMAN, *supra* note 2, at 114–18 (criticizing the application of the prudent-person rule to short-term trusts).

35. *Id.* at 120–21.

36. Professor Friedman himself, in his classic article, provided a much richer, full-bodied explanation:

The decisive cases . . . fall into a relatively narrow time-span, beginning about 1880, following in rapid succession for about 25 years, then tapering off . . . [W]hy did the courts so generally accept the spendthrift trust after 1880? Of course, the very fact that such trusts may have been spreading rapidly was a reason; judges do not normally like to reject patterns of behavior popular among members of their own social class. But deeper causes must have initiated and sustained the trend. The age (1880–1900) was an age of crisis; crisis engendered by the full impact of industrialism, the stresses and

the courts in Oklahoma, for example, took up their first trust cases, a vibrant law of trusts already existed back East. Massachusetts had by then been administering and litigating the affairs of trusts for well over a century. The Massachusetts courts and judges were well respected. Why shouldn't the Oklahoma courts have looked to Massachusetts in fashioning their own law of trusts? In any event, when the first *Restatement of Trusts* appeared in 1935, it incorporated all three developments.³⁷

Next comes a chapter entitled, "Control by the Dead and Its Limits: The Rise and Fall of the Rule against Perpetuities." In it, Professor Friedman nicely explains the operation of the rule against perpetuities. Mostly, though, he traces its fall from grace. Until the late 1990s, the rule against perpetuities was a thoroughly accepted, if much maligned, part of the American legal landscape against which those who created trusts had to operate.³⁸ Brushing

strains of a maturing, technical, organized mass economy. In the psychology of the period's social movements, some historians have detected a feeling of uneasiness . . . Consolidation of economic and social position was a dominant political motive; caste and class were hardening. The industrial "trust," the big union, the farmer's movements, the national grief over the passing of a symbolic frontier: these may suggest a social urge for rule and safety in an economic order increasingly beyond the individual's grasp. The dynastic trust and the spendthrift trust were "conservative" in the sense that they protected estate-entities against social change. That courts could, by their say-so, make secure a long-term "estate" against changing fortunes, may have been particularly pleasing at a time when so many values were called into question.

Lawrence M. Friedman, *The Dynastic Trust*, 73 YALE L.J. 547, 582-83 (1964) (footnote omitted).

37. See 1 RESTATEMENT OF TRUSTS §§ 152-53 (1935) (recognizing spendthrift trusts); *id.* § 227 (setting forth as default law the proposition that trust funds must be invested as a prudent person would); 2 *id.* § 337 (providing rules for terminating trusts by consent, subject to the *Clafflin* doctrine). Professor Scott, the Restatement's Reporter, published the first edition of his monumental treatise just four years later. See AUSTIN WAKEMAN SCOTT, *THE LAW OF TRUSTS* (1st ed. 1939). He also served as Reporter for the Second Restatement, which appeared in 1959, and published the second and third editions of his treatise in 1956 and 1967. In short, he presided over the American law of trusts during the middle half of the twentieth century. He spent virtually his entire life in Massachusetts at the Harvard Law School. Naturally, he was a keen observer of the Massachusetts cases, on which he lavished attention in his treatise. He thus played a crucial role in transmitting Massachusetts trust law, not only via the Restatements and the various editions of the treatise, but also via the several generations of well-placed law students whom he taught from 1910 to 1961. See Mark L. Ascher, *Scott, Austin Wakeman* (detailing Scott's influence), in *THE YALE BIOGRAPHICAL DICTIONARY OF AMERICAN LAW* 484, 484-85 (Roger K. Newman ed., 2009).

38. In a couple of states, there had long been no rule against perpetuities, at least as to personal property held in trust. See *Locklear v. Tucker*, 203 P.2d 380, 386 (Idaho 1949) (holding that the common law rule against perpetuities was no longer in force in Idaho); *Becker v. Chester*, 91 N.W. 87, 100-01 (Wis. 1902) (noting that the rule against perpetuities was abrogated in Wisconsin with respect to personalty). These states had remained, however, well off the beaten path of the trust business. Certainly, there had been no stampedes of settlors seeking to settle perpetual trusts subject to their laws. See RESTATEMENT (THIRD) OF PROP.: WILLS AND OTHER DONATIVE TRANSFERS ch. 27, intro. note at 119 (Tentative Draft No. 6, 2010) ("Before 1986, . . . transferors had little desire to take advantage of the absence of a Rule in those states in order to establish perpetual trusts for their descendants from time to time living forever."); Friedman, *supra* note 36, at 550 ("In Wisconsin there are apparently no real barriers to the creation of trusts lasting far longer than the common-law period of perpetuities. Yet settlors in Wisconsin do not seem to want to create private trusts in perpetuity."). I myself recall an acerbic remark uttered, probably in 1977, by

aside thick and pungent layers of arcane detail, one might best describe the function of the rule against perpetuities as establishing an upper limit of approximately ninety years on how long the living must put up with the Dead Hand. Not that that cramped many settlors' style. In my time as an estate planner, not one of the clients for whom I worked complained about the duration of the trusts we were helping them to create. Pretty clearly, the wealthy were *not*, at that time, clamoring to create perpetual private trusts.³⁹ In the late 1990s, however, various state legislatures began to repeal the rule against perpetuities, either in fact or in practical effect. It is now possible in about half of the states to create private trusts of essentially unlimited duration.⁴⁰

Professor Friedman does a superb job of telling the story of the decline of the rule against perpetuities. He does not attribute the decline to a change in our collective attitude toward wealth generally or toward dynastic wealth in particular. Instead, he says, banks and trust companies declared war on the rule against perpetuities:

Basically, banks and trust companies lobbied for the change. A tax on generation-skipping transfers (GST) entered the law in 1986 So, if I leave my estate to my daughter for life, and then to her children after she dies, the GST will tax the transfer when she dies. . . . But [\$5 million] of each transfer is exempt. Hence, if I can set up a trust that lasts many, many generations, the [\$5 million] gifts can escape tax for as long as the trust itself might last.

But how long is that? Only as long as the rule against perpetuities allows. However, if the state abolishes the rule, then a rich man can set up a very long-term trust and take full advantage of this tax loophole. Bankers and trust companies saw this opening and brought their considerable influence to bear on state legislatures, which were happy to oblige.⁴¹

my late, great estate planning professor, A. James Casner, also a long-time resident of Massachusetts, to the effect that no one in his right mind would find the prospect of creating a perpetual trust sufficiently attractive to subject it to the laws of either of these states.

39. See CHESTER, *supra* note 3, at 22 (“[A] strong case can be made that the relatively new push to extend dead hand control beyond one’s immediate circle is a creature not of basic changes in decedents’ attitudes, but of the aggressive marketing of estate planning ‘packages’ by professionals.”); Max M. Schanzenbach & Robert H. Sitkoff, *Perpetuities or Taxes? Explaining the Rise of the Perpetual Trust*, 27 CARDOZO L. REV. 2465, 2470 (2006) (“Although the rise of the perpetual trust might be viewed as evidence of a dynastic impulse, our findings suggest instead that the modern perpetual trust is primarily a creature of the federal transfer taxes.”); *supra* note 38.

40. See 2 AUSTIN WAKEMAN SCOTT, WILLIAM FRANKLIN FRATCHER & MARK L. ASCHER, SCOTT AND ASCHER ON TRUSTS § 9.3.9 n.17 (5th ed. 2006 & Cum. Supp. 2010) (listing twenty-six state statutes that have “actually or effectively repealed” the rule against perpetuities).

41. FRIEDMAN, *supra* note 2, at 131–32. Given the central role of the GST exemption in the war on the rule against perpetuities, Professor Waggoner recently urged Congress to limit the duration of the GST exemption to two generations. Lawrence W. Waggoner, *Congress Should Impose a Two-Generation Limit on the GST Exemption: Here’s Why* (Univ. of Mich. Law Sch. Pub.

This, then, is a story *not* about a change in popular attitudes but about the financial-services industry's ability to manipulate state legislatures.⁴²

It gets worse. In the ensuing race to the bottom, a second form of competition promptly emerged. Remember spendthrift trusts? In the late nineteenth century, the American courts decided that the settlor, by adding a few words to the governing instrument, could render the trust's beneficial interests inalienable.⁴³ The settlor could not, however, make his or her own interest(s) inalienable.⁴⁴ In other words, settlors could not create trusts for themselves (self-settled trusts) that would allow them to avoid their own creditors. Since the late 1990s, however, the legislatures of a few states have proven themselves willing to go even this extra mile, to allow their financial-services industries to market so-called asset-protection trusts.⁴⁵

This, then, is my biggest criticism of Professor Friedman's book: Where's the outrage? It is certainly true that the rule against perpetuities was, and where it remains in force still is, a crusty old curmudgeon. It is complicated beyond any easy description,⁴⁶ and its detractors have never shied away from ridiculing each and every one of its embarrassing little foibles.⁴⁷ Virtually everywhere, however, it remained in place until the financial-services industry declared war on it. Why? Because it served a crucial role.⁴⁸ There is no good reason why we should permit private trusts

Law & Legal Theory Working Paper Series, Working Paper No. 205, 2010), available at <http://ssrn.com/abstract=1640742>. Several years ago, the Staff of the Joint Committee on Taxation made basically the same recommendation. JOINT COMM. ON TAXATION, OPTIONS TO IMPROVE TAX COMPLIANCE AND REFORM TAX EXPENDITURES 393 (2005).

42. For a somewhat more nuanced but still abbreviated version of the story, see *infra* text accompanying notes 135–39. For fuller versions, see CHESTER, *supra* note 3, at 22–27; MADOFF, *supra* note 4, at 8, 76–82; Ira Mark Bloom, *The GST Tax Tail Is Killing the Rule Against Perpetuities*, 87 TAX NOTES 569 (2000); Jesse Dukeminier & James E. Krier, *The Rise of the Perpetual Trust*, 50 UCLA L. REV. 1303 (2003); Schanzenbach & Sitkoff, *supra* note 39; Robert H. Sitkoff & Max M. Schanzenbach, *Jurisdictional Competition for Trust Funds: An Empirical Analysis of Perpetuities and Taxes*, 115 YALE L.J. 356 (2005); Stewart E. Sterk, *Jurisdictional Competition to Abolish the Rule Against Perpetuities: R.I.P. for the R.A.P.*, 24 CARDOZO L. REV. 2097 (2003).

43. See *supra* text accompanying note 32.

44. 3 SCOTT, FRATCHER & ASCHER, *supra* note 40, § 15.4.

45. For more on asset-protection trusts, see *infra* notes 143–44 and accompanying text.

46. Yet Professor Gray's classic restatement takes only one sentence: "No interest is good unless it must vest, if at all, not later than twenty-one years after some life in being at the creation of the interest." JOHN CHIPMAN GRAY, *THE RULE AGAINST PERPETUITIES* § 201 (Roland Gray ed., 4th ed. 1942).

47. See, e.g., W. Barton Leach, *Perpetuities in a Nutshell*, 51 HARV. L. REV. 638, 638 (1938) (noting that teachers of the legal profession love highlighting the difficulty, complexity, and obscurity of the rule against perpetuities); W. Barton Leach, *Perpetuities in Perspective: Ending the Rule's Reign of Terror*, 65 HARV. L. REV. 721, 722 (1952) (stating that the rule against perpetuities is "made badly" and a "dangerous instrumentality in the hands of most members of [the legal] profession").

48. See, e.g., CHESTER, *supra* note 3, at 40 ("While the Rule against Perpetuities dealt with the problem of ending longterm trusts only imperfectly, it was certainly better than nothing.").

to live forever. Can you imagine being asked, as a lawyer or a judge, to interpret a trust deed from the year 1011?⁴⁹ The changes in custom, style, usage, and even language that inevitably occur over the course of a thousand years would necessarily reduce any interpretation to little more than speculation or, more likely, outright transposition of *our* mores onto the venerable document. Constitutional theory might seem, by comparison, obsessively faithful to the governing instrument. Times have changed so much since 1011 that trying our best to apply the document according to its own terms would be nothing short of quixotic. Such a document would certainly *not* have been executed in the United States; it probably would not even have been executed on what is currently U.S. soil. Its language would necessarily differ starkly from twenty-first century American English. Society itself back then was obviously *very* different, as was the “economy.” The legal landscape, including the “tax laws,” were too. And what about the beneficiaries? Is there *anyone* reading this Review who can name even *one* ancestor who was alive in 1011?

For centuries, we have allowed charities and charitable trusts to be perpetual, *because they benefit the public*. We sense that a charitable trust is intrinsically more valuable than a private trust. If nothing else, a charitable trust casts its benefits much more widely. We thus tolerate the prospect that we might, on behalf of a charitable trust, someday need to (mis)interpret an ancient document. We also understand that notions of public benefit are necessarily fluid, so we sense little conceptual difficulty when we also realize that a subsequent determination of what benefits the public unavoidably and often consciously reflects, almost to the exclusion of everything else, the current social context. In addition, *cy pres* has long been available to ameliorate and, if need be, to correct for the obsolescence that inevitably ensues when human beings attempt to breathe eternal life into their own pitiful little playthings.

By contrast, private trusts are supposed to be for the exclusive benefit of certain readily identifiable individuals, who are supposed to be identifiable solely by reference to the settlor’s intent.⁵⁰ Moreover, the trustee is supposed to provide for the beneficiaries and allocate the trust’s income and property among them, in certain narrowly defined ways, which again are supposed to

49. Lest the reader accuse me of being unfair or hyperbolic in discussing a trust deed from 1,000 years ago, I would note that 1,000 years is considerably shorter than *forever* and that a number of jurisdictions now permit perpetual private trusts. See RESTATEMENT (THIRD) OF PROP.: WILLS AND OTHER DONATIVE TRANSFERS ch. 27, intro. note at 124–25 (Tentative Draft No. 6, 2010) (listing eighteen statutory schemes that allow perpetual trusts). One thousand years is merely the most preposterously piggish period of time that has yet pushed its way into an American perpetuities statute. See ALASKA STAT. §§ 34.27.051–.070 (2010); COLO. REV. STAT. ANN. § 15-11-1102.5 (West 2009); UTAH CODE ANN. § 75-2-1203 (2007); WYO. STAT. ANN. § 34-1-139 (2009).

50. 2 SCOTT, FRATCHER & ASCHER, *supra* note 40, § 12.1.

be determined exclusively by reference to the settlor's intent.⁵¹ Yet if we allow private trusts to endure indefinitely, we must also admit that when the time comes to interpret the relevant governing instrument, possibly 1,000 years later, the interpretation will inevitably not be by reference to the settlor's intent but almost entirely by reference to the social context of a brave new world. Doing so seems entirely appropriate in the case of a charitable trust but just plain wrong in the case of a private trust. Professor Friedman and others seem to understand the problem and have suggested importing into the enforcement of private trusts something like the wide-ranging judicial discretion that *cy pres* has long brought to the enforcement of charitable trusts.⁵² Yet the law of private trusts does not currently come close,⁵³ and I would argue that it ought not try. Even the lawyers who help clients create perpetual private trusts seem to understand the problem because they often recommend inclusion at each generation of a special power to appoint the property outright among the settlor's issue.⁵⁴ Private trusts are simply different. Precisely because they are so different, they should not be perpetual.

Allowing private trusts to be perpetual is loony for yet another reason. I have never understood why we allow healthy, adult children an almost unlimited right to inherit their parents' property.⁵⁵ In most cases, however, children at least *knew* their parents. Maybe, even, in a miniscule number of instances, they contributed to a parent's acquisition of property. Grandchildren, too, generally knew their grandparents to at least *some* extent. But what about great-grandchildren? And great-great grandchildren? And great-great-great grandchildren? And great-great-great-great grandchildren? If at this point you think I'm beating a dead horse, think again. Forever is a long time. Yet at some point, not far beyond grandchildren, the beneficiaries of a perpetual private trust will have no claim that sounds meaningfully in affinity, contribution, or even acquaintance, on which to anchor a claim to an interest in what used to be the settlor's property.⁵⁶ They will be nothing

51. See generally 4 SCOTT, FRATCHER & ASCHER, *supra* note 40, chs. 19, 20 (explaining the investment of trust funds and the concepts of principal and income, respectively).

52. FRIEDMAN, *supra* note 2, at 121–22, 136; see also CHESTER, *supra* note 3, at 47–53.

53. See, e.g., UNIF. TRUST CODE § 412 (2006) (giving courts the power to modify the administrative or dispositive provisions of a trust due to unanticipated circumstances or inability to administer the trust effectively); see also CHESTER, *supra* note 3, at 47–53 (recounting various reform efforts regarding the modification of long-term trusts).

54. See Susan F. French, *Perpetual Trusts, Conservation Servitudes, and the Problem of the Future*, 27 CARDOZO L. REV. 2523, 2525 & n.8 (2006) (explaining a tax advantage under the Internal Revenue Code for conservation servitudes “granted in perpetuity”).

55. See Mark L. Ascher, *Curtailing Inherited Wealth*, 89 MICH. L. REV. 69, 74 (1990) (proposing strict limitations on disposition of wealth to healthy, adult children).

56. See CHESTER, *supra* note 3, at 4 (“Most decedents really have no chance to know generations more remote than that of their grandchildren. Thus the decedent could not really care in any human way about these hypothetical future descendants.”).

more than the winners of a genetic lottery run amuck. That someone might be entitled to receive periodic checks from a trustee merely because a wealthy great-great-great-great grandparent lived in a state that permitted perpetual private trusts strikes me as deeply offensive to our shared egalitarian ideals.⁵⁷ Permitting perpetual private trusts also strikes me as utterly devoid of any significant offsetting justification, such as providing the settlor with meaningful additional incentives to earn and save or increasing fidelity to the settlor's wishes, for reasons that should by now be apparent.

Another reason not to permit perpetual private trusts is that, in the final analysis, the only real beneficiaries will be the trustees and the lawyers.⁵⁸ Over time, the number of trust beneficiaries will increase geometrically. Soon there will be so many beneficiaries that trust administration in anything like the usual way will become prohibitively expensive and eventually impossible. One estimate suggests that the average settlor might have 450 descendants 150 years after creating the trust; 7,000 descendants after 250 years; 114,500 descendants after 350 years; and 1.8 million descendants after 450 years.⁵⁹ Thus, the "demographics of descent put the perpetual or multiple-centuries trust on a collision course with core principles of trust administration."⁶⁰ To avoid liability for distributing trust assets to one who is not a trust beneficiary, the trustee would at some point have to undertake expensive and repetitive inquiries just to keep track of the beneficiaries.⁶¹ For centuries, the rule against perpetuities has relieved trustees of most such burdens. In the case of a perpetual private trust, however, once the trustee has identified all of the beneficiaries, the fun has only just begun. Precisely how does one administer a trust that has thousands—or millions—of beneficiaries? It is fundamental that the trustee must act in accordance with the duty of impartiality. Not only must the trustee's actions be impartial—the trustee must, as appropriate, consult with, and take into account the

57. See MADOFF, *supra* note 4, at 8 ("The result [of dynasty trusts] will be the creation of new societal divisions between those who are beneficiaries of these tax-free, judgment-proof, long-term trusts, and the rest of society.").

58. Is it any wonder that elements of the financial-services industry lobbied so hard for the necessary legislation?

59. RESTATEMENT (THIRD) OF PROP.: WILLS AND OTHER DONATIVE TRANSFERS ch. 27, intro. note at 120 (Tentative Draft No. 6, 2010).

60. *Id.*

61. To avoid similar difficulties, the trend in the law of intestate succession has been to cut out "laughing heirs," usually defined as those beyond grandparents and their issue. See, e.g., UNIF. PROBATE CODE §§ 2-102 to 2-103 (amended 2008), 8 U.L.A. 38, 42 (Supp. 2010) (providing for intestate succession only to descendants of the decedent's grandparents or closer kin). If none of a decedent's heirs qualifies, the property that would have passed by intestate succession escheats to the state. *Id.* § 2-105. The "penalty," thus, is forfeiture. Refusing to allow private trusts to endure for more than a couple of generations imposes no such penalty. The family merely gains unrestricted access to the property sooner than the settlor may have wished.

significantly differing interests of, each of the beneficiaries.⁶² At some point, all too often after the trustees and the lawyers have pretty much picked the trust's bones clean, the court will have no choice but to do that which the rule against perpetuities has long done—put the trust out of its (and its beneficiaries') misery.⁶³

Nor is legislation authorizing spendthrift protection for self-settled trusts a positive development. It's not at all clear why we allow spendthrift trusts to stand between trust beneficiaries and their creditors. Our departure in this respect from English law⁶⁴ outraged Professor Gray, for example.⁶⁵ Why we would also allow the settlor to retain an interest in a trust while denying the settlor's own creditors access to the retained beneficial interest is even harder to fathom.⁶⁶ Certainly the possibility that such a settlor can thereby almost literally have his or her cake and eat it too⁶⁷ only sharpens the perception that the law favors the haves, particularly those who have the means to pay for the finest legal representation, over everyone else.⁶⁸

As Professor Friedman shows, the changes that have resulted in the decline of the rule against perpetuities and the advent of the self-settled spendthrift trust do not reflect considered judgments about good policy.⁶⁹ Why Professor Friedman is so hesitant to criticize these developments is

62. See UNIF. TRUST CODE § 803, 7C U.L.A. 600 (2000) (setting forth the trustee's duty of impartiality in "investing, managing, and distributing the trust property, giving due regard to the beneficiaries' respective interests"); RESTATEMENT (THIRD) OF TRUSTS § 79 & cmt. d (2007) (requiring the trustee's impartiality in investing, protecting, and managing the trust estate and in consulting with the trust beneficiaries).

63. See William J. Turnier & Jeffery L. Harrison, *A Malthusian Analysis of the So-Called Dynasty Trust*, 28 VA. TAX REV. 779, 779 (2009) ("Administrative and tax costs are likely to reduce the yield on such trusts to a level where inflation, rising expectations, and an ever growing band of beneficiaries are typically assured to outpace the ability of the trust to deliver the benefits anticipated by trust settlors."). For a different take, see Eric Rakowski, *The Future Reach of the Disembodied Will*, 4 POL. PHIL. & ECON. 91, 123 (2005) (predicting that a perpetual private trust will be administered "more like a professionally managed pension account" or a "garden-variety mutual fund").

64. See FRIEDMAN, *supra* note 2, at 122 ("[T]he spendthrift trust doctrine has no real counterpart in England.").

65. See JOHN CHIPMAN GRAY, *RESTRAINTS ON THE ALIENATION OF PROPERTY*, at iii–xiv (2d ed. 1895) (lamenting the rise of the spendthrift trust in the United States).

66. No doubt many physicians yearn for relief from the threat of malpractice liability. Some of the marketing for asset-protection trusts seems clearly aimed at them. See, e.g., ALASKA TRUST COMPANY, <http://www.alaskatrust.com/> (featuring a photo of a physician in the center of the home page screen). Might the rise of the asset-protection trust be the revenge of the trust lawyers on the tort lawyers?

67. See Elena Marty-Nelson, *Offshore Asset Protection Trusts: Having Your Cake and Eating It Too*, 47 RUTGERS L. REV. 11 (1994).

68. For more on self-settled spendthrift trusts, see *infra* notes 143–44 and accompanying text.

69. See, e.g., RESTATEMENT (THIRD) OF PROP.: WILLS AND OTHER DONATIVE TRANSFERS ch. 27, intro. note at 126 (Tentative Draft No. 6, 2010) ("The policy issues . . . have not been seriously discussed in the legislatures."); MADOFF, *supra* note 4, at 6 (contending that, while greater and greater rights are granted to the dead, these changes occur "with little attention paid to the costs imposed on the living").

unclear. It is not that he is uncomfortable sharing his opinions. He complains that the law of inheritance “is surely still too rigid,” that it still requires “too many formalities,” and that it is “still too bureaucratic and complex.”⁷⁰ He complains bitterly about McCarthyism.⁷¹ And he loves to jab conservatives and Republicans.⁷² But when it comes to evaluating the war on the rule against perpetuities or the campaign for self-settled spendthrift trusts, he is missing in action.

In fairness, it may be that Professor Friedman thinks that neither the decline of the rule against perpetuities nor the advent of the self-settled spendthrift trust is all that bad—or all that big a deal. Toward the end of his discussion of the rule against perpetuities, for example, he spends a fair amount of time arguing that perpetual accumulation trusts probably won’t “suck into their maws a big share of the country’s wealth.”⁷³ I agree. The courts have always been leery of long-term accumulation trusts.⁷⁴ Moreover, only the truly nutty are interested in consigning their assets indefinitely to accumulation trusts. Normal folk want their assets to continue benefitting themselves and their families. It is the perpetual trust for the benefit of the settlor’s family or the self-settled spendthrift trust that Professor Friedman should have worried more about. Unlike perpetual accumulation trusts, these are devices that are now reportedly being created in fairly large numbers.⁷⁵

More generally, Professor Friedman asserts, “Even without perpetual trusts, the very rich have a tendency to *stay* rich. . . . [T]he Vanderbilts and the Astors and the Duponts seem on the whole to be doing quite well, without perpetual trusts. . . . I doubt that perpetual trusts will make things much worse, or worse at all.”⁷⁶ Yet as Professor Friedman admits, not all of the “great” families continue to command anything like their old economic empires. We have all heard the old saying: “Shirtsleeves to shirtsleeves in

70. FRIEDMAN, *supra* note 2, at 32, 42–44, 179.

71. *See id.* at 151–52 (mentioning congressional investigations into tax-exempt charitable organizations based on unfounded suspicions of Communist ties).

72. *See, e.g., id.* at 53 (“Conservatives wring their hands and worry about the decline of the West, and the decay of traditional values, and so on.”); *id.* at 152 (“The right wing was especially suspicious of the social sciences.”); *id.* at 169 (“This fact has not escaped the beady eyes of conservatives.”); *id.* at 177 (“The Republican party beat the drums . . .”).

73. *Id.* at 136–39.

74. *See, e.g.,* *Thellusson v. Woodford*, (1798) 32 Eng. Rep. 1030 (H.L.) 1030–33, 1036–38, 1044–45 (appeal taken from Eng.) (construing a will to avoid impermissible accumulation); *Trusts of Holdeen*, 403 A.2d 978, 978–81 (Pa. 1979) (invalidating provisions for accumulation of trust income for 500 to 1,000 years).

75. *See* MADOFF, *supra* note 4, at 82 (“[D]ynasty trusts have become part of a standard estate plan for wealthy individuals.”). For similar observations, see CHESTER, *supra* note 3, at 24–27; Dukeminier & Krier, *supra* note 42, at 1316; Sitkoff & Schanzenbach, *supra* note 42, at 410–11; Ray D. Madoff, *America Builds an Aristocracy*, N.Y. TIMES, July 12, 2010, at A19.

76. FRIEDMAN, *supra* note 2, at 135; *see also* Rakowski, *supra* note 63, at 122 (“Allowing trusts to continue to be formed, now without limits on their duration, seems unlikely to make matters worse . . .”).

three generations.” And even if some of the “great” families do remain quite well-off, notwithstanding our meager efforts at reining in their wealth, that fact hardly argues for making it easier still for them to realize their dynastic ambitions. It may well be that the very rich simply are different, i.e., that it is exceedingly difficult to stifle their dynastic efforts. This, too, is hardly an argument in favor of encouraging those who aren’t *very* rich to try to keep up with the Vanderbilts.⁷⁷ It is one thing to admit that the law may not always have had its desired effect on the very rich; it is quite another to sit idly by while the law openly signals approval of dynastic behavior and encourages it in far broader segments of society.

Others, like Professor Madoff, openly worry about the dangers of perpetual private trusts:

[A]lthough these trusts operate largely outside the public view, like spores in a horror movie, they are posed to fundamentally transform the face of the United States by creating a new aristocracy made up of individuals who have access to large amounts of untaxed wealth to meet their every need and desire while being immune from claims of creditors.⁷⁸

She notes that exploitation of the GST exemption, currently \$5 million (\$10 million per couple), is only the tip of the iceberg.⁷⁹ Typical GST planning calls for “leveraging” the GST exemption: The settlor funds the GST-exempt trust (i.e., a dynasty trust) with assets expected to appreciate dramatically over time. One way to do this is to fund the trust with a deeply discounted minority interest in a family business. Even if the settlor uses cash, the trustee may purchase insurance on the life of the settlor (or someone else). Upon the death of the insured, the trust ordinarily ends up with vastly more than the premiums paid, at no additional income tax or transfer tax cost. Other methods of leveraging include using trust assets as start-up capital for a business enterprise or to purchase, at a substantial discount, a minority interest in a family business. Leveraging thus enables a family to commit to the dynastic enterprise vastly more than the amount of the current GST exemption. Moreover, because the magic of the GST exemption is strongest

77. Even today, the GST exemption is “only” \$5 million. I.R.C. §§ 2631(c), 2010(c) (West 2010). These days, a fair number of people can afford to create irrevocable trusts of up to \$5 million to exploit an overly generous tax loophole. Very few such folks, however, bear any meaningful resemblance to the Vanderbilts. Likewise, one does not need \$5 million to play these games. Almost all estate planners eagerly ply their trades for clients worth considerably less. Moreover, as word-processing skills and familiarity with the Internet increase, more and more people will create their own governing instruments online. Soon, perpetual private trusts of all shapes and sizes will begin to surface. It stands to reason that changes in the law are much more likely to affect the behavior of those who are nearer the margin of the planning envelope than that of the Vanderbilts. One of the most insidious aspects of these changes may be that they encourage people at any number of economic levels to aspire to dynastic behavior.

78. MADOFF, *supra* note 4, at 76.

79. *Id.* at 82–85.

when the property remains in trust indefinitely, a family with dynastic objectives can amplify the exemption's effect by making as few distributions as possible. Say that a beneficiary needs a principal residence or wants a vacation home. Why shouldn't the trustee purchase, own, and maintain the dwelling, allowing the beneficiary to live there rent free? The trust would thus retain the exempt property, and as the dwelling appreciated, so would the trust. Alternatively, if a beneficiary wanted to start a business, might not the trustee provide the funding and retain most, if not all, of the equity? Again, the trust would retain the exempt assets, and if the business prospered, the dynasty would too.

Professor Chester, as well, is clear that good policy does not include perpetual private trusts:

Locking up trust funds in perpetuity through generations of a family smacks of aristocracy. . . . Although experts differ . . . it is hard to deny that, at least symbolically, the widespread allowance of such trusts by the legal system seems antidemocratic. My own conclusion is that the Rule against Perpetuities has for centuries struck a reasonable balance between the interests of wealthy families and of society as a whole and thus should be retained in some form.⁸⁰

In addition, dynasty trusts may not be good for their beneficiaries.⁸¹ Even where the rule against perpetuities still limits the duration of private trusts, many thoughtful settlors impose substantial limitations on what they make available to their beneficiaries. Some impose age limitations, requiring their beneficiaries to be twenty-five or thirty-five years old, or older, before receiving their shares. Others impose conditions requiring their beneficiaries to get a college education, lead a life of sobriety, marry the right sort of spouse, or adhere to the right religion, in order to receive, or to continue receiving, their shares. Yet others consciously limit that which their beneficiaries can receive, either by dissipating it prior to death or by leaving it to others, such as charity or a surviving spouse. Why we would encourage the creation of dynasty trusts designed to blindly benefit endless generations, when so many people believe that doing so is not even in their own children's best interest, is not clear.

In the penultimate topical chapter, Professor Friedman takes up charitable gifts and foundations, which we have long permitted to be perpetual. He finds that they, too, pose little threat to an egalitarian society. For a number of reasons, he argues, gifts to charity involve substantially less

80. CHESTER, *supra* note 3, at 116; *see also* Rakowski, *supra* note 63, at 122 ("I confess to a lingering feeling that long-lived trusts are symbolically pernicious . . .").

81. *See* MADOFF, *supra* note 4, at 8, 82 ("Many wealthy people, including Andrew Carnegie and Warren Buffet[i], believe that it is not in their children's best interest for them to be given so much wealth that they don't need to work."); Ascher, *supra* note 55, at 99 (collecting similar viewpoints).

Dead Hand control than one might otherwise suppose.⁸² First, the best-advised settlors usually articulate their charitable purposes in exceedingly broad terms.⁸³ Second, *cy pres* has long allowed courts to reformulate a charitable trust's charitable purpose when the original purpose becomes illegal, impossible, impracticable, or now, wasteful.⁸⁴ Third, although the settlor's friends and family are ordinarily in charge at the beginning, others, much less familiar with the settlor and much less sympathetic to his or her wishes, inevitably replace them.⁸⁵ Professor Friedman also seems to conclude that a vast network of charitable foundations and their staggering wealth are a nice counterbalance to governmental and corporate power.⁸⁶

Professor Madoff takes a substantially dimmer view of perpetual charitable trusts. First, she says, it is all too easy to overestimate the benefit they confer over time.⁸⁷ One might think that a never-ending stream of payments would inevitably amount to a *huge* sum. Given the present value of money, however, a dollar now is worth significantly more than a dollar later. In fact, a buck consigned to a perpetual charitable trust is incapable of *ever* producing the same bang as a buck given outright. Second, the perpetual charitable trust "is founded on an assumption that people can make intelligent decisions about the use of resources in the distant future. . . . Does it really make sense for current policy to be dictated by plans established by someone living in the year 200?"⁸⁸ Third, the availability of perpetual charitable trusts "encourages saving for tomorrow (and the next century and next millennium) instead of spending for today. . . . [I]n the pursuit of perpetuity, fewer resources are being devoted to these pressing issues."⁸⁹ Fourth, "[t]he biggest beneficiaries are the trustees, who are paid large trustee salaries, and the supporting institutions, such as banks and other financial services companies, who are paid fees for managing this accumulating wealth."⁹⁰ Fifth, the tax incentives for charitable giving, be it for purposes of the income tax, the estate tax, or the gift tax, take the form of a deduction.⁹¹ Thus, the value of the incentive increases with the donor's wealth. Indeed, in the case of the estate tax and the gift tax, it is only the truly wealthy who

82. FRIEDMAN, *supra* note 2, at 169.

83. *See id.* at 142, 155, 160 (noting that large foundations typically have sweeping mandates).

84. *Id.* at 153.

85. *Id.* at 168–69.

86. *Id.*

87. MADOFF, *supra* note 4, at 105–06.

88. *Id.* at 106–07.

89. *Id.* at 107.

90. *Id.*

91. *Id.* at 109–12; *see also* I.R.C. §§ 170, 2055, 2522 (West 2010) (granting deductions for income, estate, and gift tax purposes).

receive any incentive at all.⁹² This suggests that the sorts of charities that receive the greatest tax subsidies may well differ dramatically from those that a broader slice of society might have chosen.⁹³ Fact seems to bear out this hypothesis given the propensity of the wealthy to support causes like art museums, opera houses, symphonies, Ivy League universities, a handful of elite private schools, and NPR. Moreover, the quantitatively unlimited nature of the deduction⁹⁴ makes the estate tax, quite literally, a “voluntary tax.”⁹⁵ That Congress would allow the wealthy—no matter how wealthy—thus to deny the government any share of their wealth, simply by leaving it to charities of their own choosing, seems remarkable, in and of itself. Even worse,

Charitable dollars do not fund necessary public expenditures, such as the costs of national defense or Social Security. In addition, private dollars cannot create programs that have a truly transformative effect on society, such as the GI Bill and investment in public education. Moreover, decisions about how charitable dollars are spent are made by the wealthy individual instead of through the political process. In this way, reliance on private charity as opposed to public tax revenues further undermines the strength of the democratic form of government.⁹⁶

It is hard to find fault with any of Professor Madoff’s criticisms. On the other hand, each of Professor Friedman’s observations also seems entirely accurate.

Professor Madoff is quite right when she notes that the legal status of the charitable trust has not always been secure in this country.⁹⁷ Since the dawn of the twentieth century, however, it has been.⁹⁸ Moreover, from their very beginnings, the income tax, the estate tax, and the gift tax have provided generous incentives for charitable giving, both outright and in trust. For better or for worse, then, charities and charitable trusts are so firmly embedded in both our law and our society that it is almost impossible to envision life without them. Few would deny that, taken together, charities and charitable trusts do an immense amount of good. In addition, the Dead Hand

92. See *id.* § 2010(c)(3) (West 2010) (setting forth, for estate tax purposes, a basic exclusion amount of \$5 million); *id.* § 2505(a)(1) (fixing gift tax unified credit at the same level as estate tax unified credit).

93. MADOFF, *supra* note 4, at 111–12.

94. I.R.C. § 2055.

95. See MADOFF, *supra* note 4, at 109 (noting that the unlimited charitable deduction allows complete avoidance of the estate tax). For the source of the label in the text, see George Cooper, *A Voluntary Tax? New Perspectives on Sophisticated Estate Tax Avoidance*, 77 COLUM. L. REV. 161, 221 (1977).

96. MADOFF, *supra* note 4, at 70.

97. See *id.* at 91–95.

98. See, e.g., 5 SCOTT, FRATCHER & ASCHER, *supra* note 40, § 37.1.3 (chronicling the gradual acceptance of charitable trusts by American courts and legislatures).

control inherent in charitable trusts pales in comparison with that inherent in perpetual private trusts. We the living should, therefore, strongly prefer charitable to private Dead Hand control. In that light, the unlimited charitable deduction makes perfect sense. To be sure, allowing the wealthy to opt out of the transfer tax system by devoting their wealth to charity substantially limits the government's ability to raise revenue. But raising revenue is only one of the objectives of the transfer taxes. Another is to break up concentrated wealth, and the transfer taxes accomplish this objective every time they divert wealth to a charitable trust. For all of these reasons, my own gaze is on the more recent and less securely embedded extensions of the Dead Hand, such as the decline of the rule against perpetuities and the advent of the self-settled spendthrift trust. These are developments that are hard to describe as positive for anyone other than the financial-services industry and those with dynastic aspirations.

Professor Friedman's final topical chapter deals with taxes. It is the shortest and least thoughtful. It does little more than note that "death taxes" have come down dramatically in the last few years. It does, however, begin the drum beat, which continues in the concluding chapter, that our collective attitude toward wealth, and toward dynastic wealth in particular, is not what it used to be.⁹⁹ As I have already indicated, I do not believe that the war on the rule against perpetuities or the campaign for self-settled spendthrift trusts says anything about our collective attitude toward wealth. Nor am I convinced that the recent decline in the death taxes proves that we are any less hostile to dynastic wealth than we used to be. That Californians in the 1980s voted to repeal a death tax that only a tiny portion of them could ever even have hoped to pay¹⁰⁰ surely says considerably more about our collective resentment of taxes and our widely shared suspicions that both government and government spending are thoroughly out of control¹⁰¹ than about our attitude toward wealth. I would predict that, when and if we ever pull our heads out of the sand and begin to make the tough choices that the Greeks, the French, the British, and the Irish now face, we as a country will promptly decide, once again, to soak the rich—and *especially* their heirs. Thus, it strikes me that news of the demise of the transfer taxes may be exaggerated. Under the terms of the very Act that "repealed the death tax," the federal estate tax was always scheduled to reappear in 2011.¹⁰² In fact, Congress recently struck

99. FRIEDMAN, *supra* note 2, at 175.

100. *Id.* at 176.

101. How else can we interpret the staggering Republican gains in the November 2010 election? It would be beyond foolish to conclude that those who voted Republican cherish chronic deficits well in excess of \$1 trillion. Yet if the candidates for whom they voted stood for anything, surely it was for "no new taxes." It is, therefore, hard to avoid the conclusion that a great many such voters voted against Big Government and government spending.

102. Economic Growth and Tax Relief Reconciliation Act of 2001, Pub. L. No. 107-16, § 901, 115 Stat. 38, 150.

from the books the statutory language that had implemented the silly one-year repeal.¹⁰³

As to taxes, others have been considerably more thoughtful. Professor Chester, for example, makes a number of important points about the role of an estate tax in limiting the Dead Hand. Some are obvious, but one in particular bears repeating: “[T]he battle over ‘death taxes’ is in reality a battle over the amount of power and control a decedent can still exercise after death. To the extent we weaken the taxation of wealth at death, we weaken any notion that ‘[the] earth belongs . . . to the living.’”¹⁰⁴ Plainly, anything that weakens the estate tax strengthens Dead Hand control. Yet over the last three decades, the federal “death tax” has been under siege. Professor Madoff, too, does not content herself with recounting its decline. “By failing to tax transfers of wealth,” we give up “an opportunity to raise revenue that could be used for a number of purposes.”¹⁰⁵ “Failing to tax transfers of wealth at death,” she continues, “promotes and nurtures an aristocratic class—individuals with enormous amounts of wealth and power achieved not because of their talents or effort but solely because of the luck of their birth.”¹⁰⁶ We should care, she says, because of the disproportionate power of the wealthy in government evidenced, for example, by the number of multimillionaires we elect to Congress.¹⁰⁷ “[W]hen there are gross disparities in wealth, there is more likely to be a mismatch between the interests and perspectives of those who govern and those who are governed.”¹⁰⁸

The opening shot in the Reagan Revolution’s war on the death tax involved the marital deduction, of all things. In the Economic Recovery Tax Act of 1981,¹⁰⁹ the marital deduction, which had previously been limited to 50% of “adjusted gross estate,”¹¹⁰ became quantitatively unlimited.¹¹¹ It also became *much* cheaper. Previously, it had been available only for transfers that conferred on the donee spouse power to control the ultimate disposition of the transferred assets.¹¹² After ERTA, however, the marital deduction was

103. Tax Relief, Unemployment Insurance Reauthorization, and Job Creation Act of 2010, Pub. L. No. 111-312, § 301(a), 124 Stat. 3296.

104. CHESTER, *supra* note 3, at 94 (quoting 7 THOMAS JEFFERSON, JEFFERSON’S WORKS 454 (Monticello ed. 1904)).

105. MADOFF, *supra* note 4, at 68.

106. *Id.*

107. *Id.* at 69.

108. *Id.* at 70.

109. Pub. L. No. 97-34, 95 Stat. 172.

110. See I.R.C. § 2056(c)(1) (West 1954) (repealed 1981).

111. See *id.* § 2056 (1982).

112. A transfer of property outright obviously qualified. So did a transfer of property in trust, if the donee spouse was the sole income beneficiary and the donee spouse had a general power of appointment over the trust assets. *Id.* § 2056(b)(5) (West 1954). A third possibility was the so-called estate trust, in which the donee spouse did not need to be entitled to the trust income currently if the terms of the trust provided that what was not distributed to the donee spouse would

also available for "qualified terminable interest property" (QTIP).¹¹³ Essentially all that was required of a QTIP trust was that the donee spouse be entitled to all of the income currently. This slashed the cost of the marital deduction.¹¹⁴ After ERTA, the donor spouse could retain control over the ultimate disposition of the transferred assets by designating the remainder beneficiaries either outright or in further trust. Not surprisingly, QTIP trusts became instant best sellers. Not only did they allow the donor spouse to retain ultimate control, but because the marital deduction was quantitatively unlimited, it was no longer necessary to resort to the charitable deduction to achieve the much-sought-after zero-tax estate. A husband, for example, could leave his entire estate, no matter how large, via a QTIP trust for the benefit of his wife, remainder in continuing trust to his children and then to his grandchildren, and be confident that at his death no federal estate tax would be due.¹¹⁵

The utility of these changes in day-to-day estate planning was immediately apparent.¹¹⁶ Professor Chester has shown, however, that these changes were particularly important for those with dynastic aspirations.¹¹⁷ Prior to ERTA, given the quantitatively limited nature of the marital deduction, the only way to "zero out" a sizable estate was to resort to the charitable deduction. That meant that even in the case of a married donor, as much as half of the estate was necessarily dissipated, passing either to charity or as taxes. Moreover, given the qualitative limitations on the marital deduction, the half that went to the donee spouse was *also* "lost," in the sense that it was the donee spouse who controlled the ultimate disposition of the assets. Under the best of circumstances, the donor spouse could only hope that, when the time came, the donee spouse would share the donor spouse's dynastic aspirations. In the worst-case scenario, the donee spouse would remarry and produce children by another, in which case even the dynasty's bloodline was at risk. After ERTA, the entire estate, no matter how large, could escape taxation via the marital deduction. Resort to charity was necessary only when the donor had no spouse or did not want to make a large

be accumulated and that whatever remained at the donee spouse's death would go to the donee spouse's estate. Rev. Rul. 68-554, 1968-2 C.B. 412.

113. I.R.C. § 2056(b)(7) (1982).

114. The entire value of the QTIP trust is instead includible in the donee spouse's gross estate. *Id.* § 2044.

115. Actually, less than the entire estate usually went via the QTIP trust. Part went via a "credit-shelter bypass trust" designed to ensure full utilization of the husband's unified credit. There would still be no tax due at the husband's death. The point of the second trust was to avoid inclusion in the wife's gross estate and thus to minimize the tax burden on her death.

116. See, e.g., Mark L. Ascher, *The Quandary of Executors Who Are Asked to Plan the Estates of the Dead: The Qualified Terminable Interest Property Election*, 63 N.C. L. REV. 1, 2 (1984) (noting contemporaneously that "[i]f elected by the executor, QTIP status allows an estate to claim the federal estate tax marital deduction for a devise of a life estate to the testator's spouse with the remainder passing to persons of the testator's choosing").

117. CHESTER, *supra* note 3, at 94, 102-03.

spousal transfer. By using a QTIP trust, the donor spouse could, at the relatively meager cost of granting the donee spouse a life interest, ensure both a zero-tax estate and perpetuation of the entire estate down the donor spouse's own bloodline. Thus, it was ERTA, far more than the early Massachusetts case law that Professor Friedman highlights,¹¹⁸ that set the table for dynastic estate planning.

And then there is the assertion that the transfer taxes have never been effective at breaking up concentrations of wealth.¹¹⁹ Such assertions have always struck me as dubious. How on earth can we know how much more concentrated wealth would be today if there had never been an estate tax? Professor Chester offers an irrefutable response: "Whether or not [the estate tax] accomplish[es] much in the way of redistribution, such a tax . . . [is] important as a symbol. A nation that professes so strongly to value equal opportunity cannot long be seen as allowing large fortunes to pass between generations completely unscathed."¹²⁰ Moreover, though the transfer taxes have never generated a large portion of the government's revenue, they generate far more than they cost to administer. So they could be a reliable source of additional revenue, especially as the baby boomers, with their massive wealth, begin to die.¹²¹ With annual deficits in the trillions of dollars, surely we ought not leave *this* stone unturned.

In his concluding chapter, Professor Friedman argues that whether the Dead Hand is "getting weaker or stronger" is a "fairly complex" story.¹²² Plainly, we have turned our backs on Thomas Jefferson. How else can we explain the fact that we now defer even more to the Dead Hand than the English? Though the English have never permitted spendthrift trusts, we have done so for more than a century. In England trust beneficiaries have always been able to terminate trusts by mutual consent, but since the turn of the last century in the United States, trust beneficiaries have had to wait until each of the settlor's material purposes has been fulfilled. In both countries the death taxes have sometimes seemed stiff, at least on paper, but, for the lucky families of American decedents who died in 2010, the federal estate tax was only for those who chose to pay it.¹²³ Though both the English and we have long understood the need to limit the duration of private trusts and have relied on the rule against perpetuities to do so, there is now in approximately half the states no effective limit on trust duration. And as though permitting spendthrift trusts were not bad enough, several states now permit settlors to create trusts for themselves while denying their own creditors

118. See *supra* text accompanying notes 31–33.

119. See *supra* text accompanying note 76.

120. CHESTER, *supra* note 3, at 98.

121. *Id.* at 104.

122. FRIEDMAN, *supra* note 2, at 181.

123. See *infra* note 146.

access. In short, we Americans seem to be creating for ourselves a law of inheritance even more amenable to dynastic wealth than that of the English, against whose aristocratic traditions we once chafed. Professor Friedman concludes from all of this that we as a country are no longer uncomfortable with wealth in general or with dynastic wealth in particular. I very much hope he is wrong and have offered at each point in Professor Friedman's analysis an alternative and less far-reaching explanation. As to the most recent events, i.e., the war on the rule against perpetuities and the campaign for self-settled spendthrift trusts, Professor Friedman seems to admit that there is a different explanation:

But this trend might be somewhat misleading. In the late nineteenth century the law of trusts did evolve in ways that favored rich individuals, rich dynasts In the late twentieth century, and into the twenty-first, the law smiled much more on rich *institutions*: notably the banks and trust companies. These institutions formed a powerful lobby. They were a strong and focused interest, and they were usually able to get their way. Individuals, even rich individuals, were a more diffuse interest. It is a maxim of political science that, in the legislative halls, even a small, focused interest beats out a larger but diffuse interest.¹²⁴

In other words, the recent changes in the law of trusts, which may well favor those with dynastic ambitions, were not enacted at the behest of those best situated to take advantage of them. They were enacted at the behest of the financial-services industry. They say little, if anything, about our collective attitude toward wealth generally or toward dynastic wealth in particular. Nor do they say anything at all about good policy. Indeed, they were often enacted after consultation with only a small bar committee composed primarily of estate planners and probate practitioners, many with close ties to the financial-services industry.¹²⁵

Ultimately, Professor Friedman seems to believe that worries about the extent of the Dead Hand are overblown:

The brutal fact remains: the dead are definitively dead. The dead "control" beyond the grave only insofar as living people let them do so. In the long run, the dead run nothing. Even in a supposed perpetual trust, charitable or not, the dead hand reigns but does not rule. Like modern kings and queens, its power ebbs away. The evolution of the *cy pres* doctrine; the decay of the Claflin doctrine; the

124. FRIEDMAN, *supra* note 2, at 181.

125. Professor Madoff recounts the nightmarish tale of how Alaska became a trust haven. Basically, a pair of brothers, one a prominent New York estate planner and the other the founder of an Alaska trust company, decided to make it so while on a fishing trip together in Alaska. MADOFF, *supra* note 4, at 81. For the company's version of the story, see Douglas J. Blattmachr, *The Genesis of Alaska Trust Company*, ALASKA TRUST COMPANY, <http://www.alaskatrust.com/index.php?id=57>.

hegemony of professional managers in the large foundations—these all demonstrate that, practically speaking, the living rule the dead, not vice versa.¹²⁶

With respect, that argument proves much too much. Of course, we *could* wrest *everything* from the dead. But we never have. More importantly, we seem recently, and for all the wrong reasons, to have ceded to the dead even more sway. This is a trend that demands reexamination.

In fact, others have begun such a reexamination. Professor Chester, for example, has concluded that there are substantial implications for our national identity in the increased reach and longevity of the Dead Hand:

The genius of our system, or at least the secular tradition that has dominated it, is that, whatever the frailties of one generation, this tradition can easily accommodate the ethos of each succeeding generation. The current financial elite, however, is trying to cement into place legal rules that will perpetuate its hegemony over the generations that will follow.¹²⁷

And again, “[T]he ultimate question for policy makers is whether property tied up for generations within a family, generally subject neither to voluntary alienation by beneficiaries nor the rights of creditors, is a good thing for society as a whole.”¹²⁸

Professor Madoff’s book, too, openly questions the increased reach of the Dead Hand. Even its subtitle, *The Rising Power of the American Dead*, leaves no doubt that Professor Madoff has evaluated the trend and found it wanting. It implies what should be obvious—that the rising power of the dead comes at the expense of the living. Nor does she shy away from saying this directly: There is “an inevitable trade-off. Whenever we grant rights to the dead, we invariably affect the living.”¹²⁹ More directly to the point,

By failing to tax inherited wealth and allowing the creation of perpetual private trusts, we allow the dead to impose their wishes on future generations. More troubling, we are allowing them to establish a new aristocracy made up of individuals who will have access to large amounts of untaxed wealth to meet their every need and desire while being immune from claims of creditors.¹³⁰

This, she concludes, “threatens our most fundamental values. . . . [Our] founders . . . consciously attempted to establish a nation that would be largely free from the strictures of the past. Over time, however, we have

126. FRIEDMAN, *supra* note 2, at 182–83.

127. CHESTER, *supra* note 3, at 39; *see also id.* at 115 (“[T]he creation and perpetuation of a hereditary aristocracy would contradict many of the ideas fundamental to the founding of America itself.”).

128. *Id.* at 115.

129. MADOFF, *supra* note 4, at 152.

130. *Id.* at 154–55.

forgotten what our founders knew, and today we are re-creating the very world from which they sought to distance themselves.”¹³¹ Professor Chester (I think)¹³² and I could not agree more.

Professor Madoff then asks why these changes have occurred. Mostly, she blames the political process. She mentions, for example, “the stealth nature” of the changes: “The larger picture has gone unnoticed because change has occurred within [discrete] areas of the law, and often at the state level.”¹³³ When the relevant committee of the local bar association recommends a package of proposed changes to the probate code, no bells begin to ring and no warning lights begin to flash. The committee states truthfully that it has vetted the proposed changes with all of the “relevant” groups, like the local bankers’ and accountants’ groups, and the legislature rubber-stamps the changes, probably without hearings. Professor Madoff adds,

Finally, the story of the American law of the dead would not be complete without recognition of the effect of money on legislation. It is significant that the areas in which American law has grown most dramatically . . . not only appeal to individuals’ desire to exert posthumous control but also appreciably benefit corporate interests. By using interests of the dead as a decoy, these entities have succeeded in enriching their own property interests. Although financial gain may be the driving force behind these changes, corporations are not the ultimate villains. Businesses are amoral, simply doing what our society tells them to do: maximize profit. The blame lies with legislators, who have responded to corporate demands even when they have not best served the needs of American society at large.¹³⁴

I would tell the story only a bit differently. It is not as though all of the banks and the entire estate planning community woke up one morning in the late 1990s and decided that they wanted to market dynasty trusts and asset-protection trusts. But some did.¹³⁵ In 1995, for example, Delaware repealed its rule against perpetuities as to personal property held in trust.¹³⁶ And in 1997, Alaska, followed almost immediately by Delaware, blessed self-settled spendthrift trusts.¹³⁷ What followed was a race to the bottom, with

131. *Id.* at 155.

132. See CHESTER, *supra* note 3, at 21 (“It would appear that the American fondness for dead hand control . . . may be enabling the creation of the very aristocracy that our country originally rebelled against.”); *id.* at 74 (“[A]ttempts to extend dead hand control . . . speak of a desire to endow a continuing aristocracy, the very institution this country’s founders revolted against.”).

133. MADOFF, *supra* note 4, at 155.

134. *Id.* at 155–56.

135. See *supra* note 125.

136. 70 Del. Laws 238 (1995) (codified at DEL. CODE ANN. tit. 25, §§ 503, 504 (2009)). In 1983, South Dakota had enacted similar legislation. 1983 S.D. Sess. Laws 523 (codified at S.D. CODIFIED LAWS §§ 43-5-1 to 43-5-9 (2010)).

137. See *infra* note 144.

legislature after legislature being asked to protect its own financial-services industry from out-of-state competition by enacting legislation that would permit it to market dynasty trusts, asset-protection trusts, or both.¹³⁸ In the ensuing rush, many state legislatures defaulted on their obligation to engage in any meaningful policy analysis.¹³⁹

For a long time in this country, only the courts and the legislatures had any real role in effectuating law reform. For nearly a century, however, there have been two highly influential and relatively independent additional sources of law reform: the American Law Institute (ALI) and the National Conference of Commissioners on Uniform State Laws (NCCUSL). The ALI produces the Restatements of the Law, and NCCUSL promulgates the Uniform Acts. Neither organization has yet enlisted in the war on the rule against perpetuities or volunteered for duty in the campaign for self-settled spendthrift trusts. Just this last May, at its annual meeting, the ALI officially repudiated the perpetual-trust movement: "It is the considered judgment of The American Law Institute that the recent statutory movement allowing the creation of perpetual or multiple-centuries trusts is ill advised."¹⁴⁰ At that same meeting, the ALI also approved a new approach to perpetuities, one that would, in general, limit the duration of Dead Hand control to two younger generations.¹⁴¹ NCCUSL's current offering on perpetuities remains the Uniform Statutory Rule Against Perpetuities (USRAP), which combines the traditional rule against perpetuities with a 90-year wait-and-see period. It is true that USRAP predates the perpetual-trust movement. Still, it is hard to imagine that NCCUSL would abandon the rule against perpetuities anytime soon given the fact that, at its high-water mark, USRAP was adopted by approximately half the states. Indeed, Professor Waggoner, who served as Reporter for both the Third Restatement of Property and USRAP, recently hinted that a new uniform perpetuities act tracking the new Restatement position may be in the offing.¹⁴² Likewise, after full-blown consideration, both the ALI and NCCUSL have officially refused to extend spendthrift

138. For other accounts, see *supra* notes 41–45 and accompanying text and *infra* note 144 and accompanying text.

139. See *supra* note 69 and accompanying text.

140. RESTATEMENT (THIRD) OF PROP.: WILLS AND OTHER DONATIVE TRANSFERS ch. 27, intro. note at 126 (Tentative Draft No. 6, 2010); see also Lawrence W. Waggoner, *Curtauling Dead-Hand Control: The American Law Institute Declares the Perpetual-Trust Movement Ill Advised 1 & n.1* (Univ. of Mich. Law Sch. Pub. Law & Legal Theory Working Paper Series, Working Paper No. 199, 2010), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1614934.

141. See RESTATEMENT (THIRD) OF PROP.: WILLS AND OTHER DONATIVE TRANSFERS § 27.1 (Tentative Draft No. 6, 2010); Lawrence W. Waggoner, *The American Law Institute Proposes a New Approach to Perpetuities: Limiting the Dead Hand to Two Younger Generations* (Univ. of Mich. Law Sch. Pub. Law & Legal Theory Working Paper Series, Working Paper No. 200, 2010), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1614936.

142. See Waggoner, *supra* note 41, at 5; Waggoner, *supra* note 140, at 11.

protection to trust settlors.¹⁴³ In short, the rigorous sort of policy analysis that the legislatures should have engaged in has begun.

It appears as well that the wind may be going out of the sails of both trends. The campaign for self-settled spendthrift trusts began in 1997 when Alaska and Delaware started competing for the asset-protection-trust business. More than thirteen years later, only ten additional states have joined them.¹⁴⁴ In contrast, during the same time frame, the war on the rule against perpetuities turned quite ugly, claiming casualties in approximately half the states. On this front, too, the carnage seems finally to have slowed. Perhaps, with clear guidance from both the ALI and NCCUSL and with the insights of academics such as those whose efforts we consider here, it will be easier in the future to turn back assaults in the remaining legislatures. In a number of legislatures these efforts have already gone down to defeat.¹⁴⁵ In

143. See UNIF. TRUST CODE § 505(a)(2) (2005); RESTATEMENT (THIRD) OF TRUSTS § 58(2) cmt. e (2003). By and large, the academic community, too, has rejected self-settled spendthrift trusts. See, e.g., Karen E. Boxx, *Gray's Ghost—A Conversation About the Onshore Trust*, 85 IOWA L. REV. 1195, 1201 (2000) (concluding that self-settled spendthrift trusts can do massive damage to our system of civil enforcement); Randall J. Gingiss, *Putting a Stop to "Asset Protection" Trusts*, 51 BAYLOR L. REV. 987, 988 (1999) (arguing that steps must be taken to eliminate the use of self-settled spendthrift trusts); Henry J. Lischer, Jr., *Domestic Asset Protection Trusts: Pallbearers to Liability?*, 35 REAL PROP. PROB. & TR. J. 479, 493–94 (2000) (stating that the self-settled spendthrift trust is a mixed bag from a policy perspective); Stewart E. Sterk, *Asset Protection Trusts: Trust Law's Race to the Bottom?*, 85 CORNELL L. REV. 1035, 1043 (2000) (highlighting the "vociferous opposition" to self-settled spendthrift trusts). For opposing viewpoints, see Robert T. Danforth, *Rethinking the Law of Creditors' Rights in Trusts*, 53 HASTINGS L.J. 287, 288 (2002); Adam J. Hirsch, *Fear Not the Asset Protection Trust*, 27 CARDOZO L. REV. 2685, 2686 (2006).

144. Currently, there are, more or less, twelve "asset-protection trust" states. Their statutes, arranged by effective date are ALASKA STAT. § 34.40.110 (1997); DEL. CODE ANN. tit. 12, §§ 3570–76 (1997); NEV. REV. STAT. §§ 166.010–170 (1999); R.I. GEN. LAWS §§ 18-9.2-1 to -7 (1999); UTAH CODE ANN. § 25-6-14 (LexisNexis 2003); MO. REV. STAT. § 456.5-505(3) (2004); OKLA. STAT. ANN. tit. 31, §§ 10–18 (West 2004); S.D. CODIFIED LAWS §§ 55-16-1 to -16 (2005); TENN. CODE ANN. §§ 35-16-101 to -112 (2007); WYO. STAT. ANN. §§ 4-10-505, 4-10-510 to -523 (2007); N.H. REV. STAT. ANN. §§ 564-D:1 to -D:18 (2008); 2010 Haw. Sess. Laws 182. Is it fair to point out that, by and large, these are among our least populous states? Or that none is a major money center? Might it be that the legislatures in these states are more susceptible to influence by a determined few? Or especially interested in attracting or retaining trust business?

145. Compare TEX. PROP. CODE ANN. § 112.036 (West 2010) (establishing the traditional version of the rule against perpetuities, subject to judicial reformation), with Tex. H.B. 1608, 77th Leg., R.S. (2001) ("An interest in a trust is not good unless it must vest, if at all, not later than 1,000 years after some life in being at the time of the creation of the interest." (alterations and emphasis omitted)). For a nice account of this incident, see STANLEY M. JOHANSON, JOHANSON'S TEXAS PROBATE CODE ANNOTATED XVII (2001 ed.). Georgia had a similar experience. See Verner F. Chaffin, *Georgia's Proposed Dynasty Trust: Giving the Dead Too Much Control*, 35 GA. L. REV. 1, 3, 9, 25–26 (2000) (arguing against proposed legislation). Compare GA. CODE ANN. §§ 44-6-200 to 44-6-206 (2010), with A Bill to Abolish the Rule Against Perpetuities, H.B. 1764, 1999–2000 Sess. (Ga. 2000) ("The rule against perpetuities is abolished."). Compare also N.Y. EST. POWERS & TRUSTS LAW § 9-1.1(b) (2008) (establishing a relatively traditional version of the rule against perpetuities), with An Act to Amend the Estates, Powers and Trusts Law in Relation to the Rule Against Perpetuities, A.B. 4924, 228th Leg., Reg. Sess. (N.Y. 2005) (providing that the rule against perpetuities would not apply to an interest in trust if the trustee had unlimited power to sell trust assets and one or more persons had unlimited power to terminate the trust).

the vast majority of states, spendthrift restrictions remain unenforceable as to the settlor's own interest, and, in about half the states, there are still reasonable limitations on the duration of private trusts.

Even on the tax front there is some room for optimism. Although Congress fiddled for nearly a decade after setting fire to the estate tax and seemed throughout much of 2010 to sanction the repeal, the estate tax was always scheduled to reappear in 2011. Indeed, when Congress eventually got around to reconsidering the fate of the estate tax, it wiped from the books the statutory language that had implemented the repeal, thereby retroactively reimposing the estate tax, albeit with staggeringly high exemption amounts and rates significantly lower than those with which we have become familiar.¹⁴⁶

Thomas Jefferson knew a little something about aristocracy. He therefore did his best to ensure that we would remain free of the threats it poses to egalitarianism and republican government. That is why he reminded us that the earth belongs to the living and urged us to keep it that way. I am glad he does not know how far we have recently strayed. For in a society that cherishes testamentary freedom, as shown not only by the words of our rhetoric but also by the force of our laws, the estate tax and the rule against perpetuities have until lately been our primary brakes on the accumulation and propagation of dynastic wealth. Yet both have recently been curtailed. Fortunately, it would not be hard to get back on the right path.¹⁴⁷ All we really need are sensible rules that end trusts after a reasonable period of time, allow legitimate creditors reasonable access to debtors' wealth, and impose a reasonable tax at death.¹⁴⁸

146. Tax Relief, Unemployment Insurance Reauthorization, and Job Creation Act of 2010, Pub. L. No. 111-312, § 301(a), 124 Stat. 3296. Presumably to avoid litigation over the constitutionality of a retroactive reimposition of the estate tax, Congress allowed the executors of decedents who died in 2010 to opt out of the estate tax. *Id.* § 301(c). The downside of this election was that the estate remained subject to the modified carryover basis rules that accompanied estate tax repeal. *Id.*

147. If Congress were to limit the duration of the GST exemption to two generations, it could kill or wound two birds with one stone. The transfer taxes would be a bit more effective, and the face that launched the war on the rule against perpetuities would be much less beguiling. *See supra* note 41.

148. *See* CHESTER, *supra* note 3, at 41 (arguing for "pragmatic legal rules" that would end trusts after a reasonable period, redistribute some wealth through taxes, and allow legitimate creditors access to one's assets).

* * *

Notes

A Proposed Wealth Redistribution System Based on the Underlying Premise of Revenue Sharing in American Pro Sports*

I. Introduction

The system of wealth redistribution¹ in the United States is unequal. It is unequal because each class does not have an opportunity to gain. The method of rectifying wealth inequality must be fair, must empower each group in the system, and must give each group an equal opportunity to benefit. Rather than taking from the rich² and giving to the poor, the method should combine the efforts of both rich and poor—contributors and beneficiaries—for the benefit of both classes and the economy as a whole. The ideal system is one that apportions, through taxes, a fund to create businesses that employ low-income beneficiaries, with their wages being the assistance they need.

Wealth redistribution is important because the income gap between rich and poor in the United States is at its widest in our recorded history.³ Similarly, wealth redistribution is a popular topic in American professional sports leagues⁴ because revenue sharing plays a prominent role in their

* I would like to thank Professor Mechele Dickerson for her help and advice. I appreciate the hard work of the Texas Law Review staff and Editorial Board, especially Sarah Hunger and Omar Ochoa. I am grateful to my family, especially Edy, for their encouragement and guidance. Finally, I would like to thank Nolan Ryan for returning the Texas Rangers to glory, and I would like to thank Carlos Slim for his inspiration.

1. *Wealth redistribution* may be used interchangeably with the term *welfare* or any other term indicating attempted rectification of wealth inequality.

2. Rather than attempt to precisely define the terms, *rich* and *wealthy* are used colloquially in this Note; they have the same denotation as would be found in a newspaper article or in a political speech. See, e.g., Frank Rich, Op-Ed, *Who Will Stand Up to the Superrich?*, N.Y. TIMES (Nov. 13, 2010), <http://www.nytimes.com/2010/11/14/opinion/14rich.html> (“Americans don’t hate rich people. They admire and often idolize success.”).

3. Ira Boudway, *The Rich Get Richer . . . and You Know the Rest*, BUSINESSWEEK (Sept. 30, 2010), http://www.businessweek.com/magazine/content/10_41/b4198033845016.htm. Wealth redistribution is also a popular topic because of the recent debate regarding tax cuts for the wealthy. Lori Montgomery, *In Push to Let Bush Tax Cuts Expire, Democrats to Focus on Narrowing Income Gap*, WASH. POST (Oct. 13, 2010), <http://www.washingtonpost.com/wp-dyn/content/article/2010/10/13/AR2010101305004.html>. On a similar note, more Americans are now living in poverty than in any other year since the Census Bureau poverty-line estimates were first published. Boudway, *supra*.

4. In an effort to focus the scope of the analysis, I will only discuss the National Football League and Major League Baseball.

administration.⁵ Professional sports leagues are addressing revenue disparity that has created a growing gap between low-revenue and high-revenue teams.⁶ High-revenue teams are often more competitive (that is, they succeed more in both the regular season and the playoffs than low-revenue teams), leading to actual league revenues less than their potential (i.e., what revenues would be if all teams were able to compete at an equally high level).⁷

Wealth inequality between those with means and those without means is not new. The issue has permeated American society since before the United States became an independent nation.⁸ Yet this important problem persists. The goal of wealth distribution, regardless of the means, should be equality of opportunity—the ability of each person to pursue her own objectives without being prevented from doing so by arbitrary obstacles.⁹ Proponents of policies to redistribute wealth through means other than the free market argue that “[a]t some point inequality in outcomes becomes so great that the quintessential American promise of equality of opportunity becomes unattainable.”¹⁰ Equality of opportunity, however, is attainable, and this Note proposes a system to counter inequality that benefits not only those on the lower end of the income spectrum but also those on the higher end.

5. See Symposium, *Panel III: Restructuring Professional Sports Leagues*, 12 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 413, 437 (2002) (explaining how revenue sharing impacts the direction taken by professional sports leagues and their investors). Throughout this Note, *revenue sharing* refers to the sharing of revenues among teams in the same professional sports league, not to be confused with revenue sharing between team owners and players.

6. See RICHARD C. LEVIN ET AL., MAJOR LEAGUE BASEBALL, THE REPORT OF THE INDEPENDENT MEMBERS OF THE COMMISSIONER'S BLUE RIBBON PANEL ON BASEBALL ECONOMICS 11 (2000), available at http://mlb.mlb.com/mlb/downloads/blue_ribbon.pdf (condemning MLB's revenue disparity and competitive imbalance as “unacceptable”).

7. See *id.* at 12 (explaining how MLB's sub-optimal financial performance is due in part to teams' revenue disparities and the resulting inequality in competitiveness and success).

8. See MICHAEL B. KATZ, IN THE SHADOW OF THE POORHOUSE: A SOCIAL HISTORY OF WELFARE IN AMERICA x (2d ed. 1996) (discussing local welfare measures that were utilized during colonial times). See generally Carole Shammas, *A New Look at Long-Term Trends in Wealth Inequality in the United States*, 98 AM. HIST. REV. 412 (1993) (examining wealth inequality in the United States from colonial times to the late twentieth century).

9. MILTON FRIEDMAN & ROSE D. FRIEDMAN, FREE TO CHOOSE: A PERSONAL STATEMENT 128 (2d ed. 1980).

10. JASON FURMAN ET AL., THE BROOKINGS INST., ACHIEVING PROGRESSIVE TAX REFORM IN AN INCREASINGLY GLOBAL ECONOMY 10 (2007), available at http://www.brookings.edu/~media/Files/rc/papers/2007/06globalization_furman/200706bordoff_summers.pdf. The authors, arguing to redistribute wealth using progressive taxation, iterate that equality of opportunity becomes unattainable because “[t]he very first thing . . . society's wealthy try to buy . . . is a head start for their children.” *Id.* (quoting J. Bradford DeLong, *Inequality on the March*, PROJECT SYNDICATE (Jan. 30, 2007), <http://www.project-syndicate.org/commentary/delong55/English>). However, if a head start for their children is the *very first* thing they buy, it is difficult to see how progressive taxation will change the circumstances of this inevitability.

Professional sports leagues, in an effort to benefit all teams, employ revenue sharing to increase the competitive balance of the league.¹¹ Optimally, even those who are net losers in revenue sharing (those who contribute a portion of their revenues to the fund that is then distributed to low-revenue teams) benefit from increased competition in the league, which leads to increased profits in the long run.¹² The United States should address wealth inequality using the same premise, substituting the goals of equality and fairness for competition. Just as each team in a professional sports league is supposed to obtain a monetary benefit in the long run from increased competition created by revenue sharing, all parties to a system used to rectify wealth inequality should receive a monetary benefit.

II. Background

A. History and Overview of Wealth Redistribution in the United States

Wealth redistribution, the vehicle for reducing wealth inequality, has historically been driven by a desire to relieve despair, maintain social order,¹³

11. See Clay Moorhead, Note, *Revenue Sharing and the Salary Cap in the NFL: Perfecting the Balance Between NFL Socialism and Unrestrained Free-Trade*, 8 VAND. J. ENT. & TECH. L. 641, 642 (2006) (discussing how the NFL uses revenue sharing to increase profits and competitiveness); see also Mark Maske, *Early Theme of NFL Season Is Parity*, WASH. POST (Oct. 17, 2010), <http://www.washingtonpost.com/wp-dyn/content/article/2010/10/16/AR2010101600325.html> (“The [NFL]’s revenue-sharing system was designed to keep franchises on relatively even financial footing . . .”). The NFL also seeks to create competitive balance on the field through a salary cap, which is supposed to keep teams from spending such a high amount on players’ salaries that other teams are unable to keep up. See generally *id.* MLB also employs a luxury tax with the goal of creating competitive balance in the league by putting a soft cap on salaries. See Matthew Ryan McCarthy, Note, *Revenue Sharing in Major League Baseball: Are Cuba’s Political Managers on Their Way over Too?*, 7 VAND. J. ENT. L. & PRAC. 555, 566 (2005) (explaining that the luxury tax—McCarthy calls it a “competitive balance tax”—applies to the amount spent in excess of a spending limit on team salaries). The tax is in addition to revenue sharing and is imposed on teams that spend more than a specified amount on player salaries in a given year. Neil deMause, *Does Baseball Need a Salary Cap?*, in *BASEBALL BETWEEN THE NUMBERS: WHY EVERYTHING YOU KNOW ABOUT THE GAME IS WRONG* 227, 228 (Jonah Keri ed., 2006). Teams exceeding their league-appointed budget are taxed at a rate beginning at 22.5% that becomes increasingly higher for repeat offenders. *Id.* The luxury tax is relatively inconsequential, however, because it is rarely used. *Id.* The luxury tax creates a soft cap on salaries, as opposed to a hard cap, since it does not impose an actual limit on team payrolls. *Id.* Rather, it is termed a soft cap because the consequences of the luxury tax are *intended* to keep team payrolls below a certain amount. *Id.*

12. This is the optimal result; however, revenue-sharing systems in professional sports leagues are imperfect. The proposed system is based on the premise of revenue sharing in professional sports leagues rather than on the results. This premise asserts that all parties should obtain a quantifiable benefit.

13. A recurring theme in history is revolt by the lower classes. One reason it may be less of a concern in contemporary America is that a percentage of all social classes maintain some notion of the “American dream”: that they have the ability to advance, succeed, and be prosperous through their own efforts. See John Zogby, *The American Dream Is Still Strong*, FORBES.COM (Jan. 29, 2009), http://www.forbes.com/2009/01/28/american-dream-polling-opinions-columnists_0129_john_zogby.html (reporting that a plurality of every single demographic group, equivalent to 56% of likely voters, believe “they can realize the American Dream”).

enforce discipline, regulate the labor markets, and allow political mobilization.¹⁴ During the American colonial period, counties, towns, and parishes handled wealth redistribution locally.¹⁵ Adam Smith commented on the importance of wealth redistribution in his influential 1776 treatise, *An Inquiry into the Nature and Causes of the Wealth of Nations*, in which he opines that “[n]o society can surely be flourishing and happy, of which the far greater part of the members are poor and miserable.”¹⁶ By the nineteenth century, states attempted to standardize welfare.¹⁷ The turn of the twentieth century brought high levels of wealth inequality. The highest concentrations existed in 1905–1906 and 1929, a result of industrial conglomerates and the financial trusts of their executives.¹⁸ The top 1% held approximately 45% of the nation’s wealth during these peak years.¹⁹

The Great Depression and the resulting New Deal policies brought about the emergence of modern welfare.²⁰ Social Security and publicly supported job programs were created to assist the unemployed, the elderly, disabled persons, and widows with children.²¹ Much of this assistance arose as part of the 1935 Social Security Act which initiated a federal financial-aid program for those who were deemed needy, but thought to be deserving of aid.²² The federal government succeeded in producing a uniformity of assistance through these programs that the states had yet to perfect.²³

In the following thirty years, social-insurance programs grew to serve government workers and veterans, and Congress passed new programs and amendments to create public housing and the National School Lunch Program.²⁴ In 1964, amid the social volatility of the Civil Rights Movement, a new era of welfare expansion occurred with the government establishing Medicare, Medicaid, and the Food Stamp Program.²⁵ The government created Medicare to provide for the medical needs of Americans over the age of sixty-four, while Medicaid was intended to give grants to states in order for

14. KATZ, *supra* note 8, at xi.

15. *Id.* at x.

16. ADAM SMITH, *AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS* 115 (The Electric Book Co. 1998) (1776).

17. KATZ, *supra* note 8, at x.

18. KEVIN B. PHILLIPS, *WEALTH AND DEMOCRACY: A POLITICAL HISTORY OF THE AMERICAN RICH* 122, 123 chart 3.5 (2002).

19. *Id.* at 122.

20. Hugh Heclo, *The Politics of Welfare Reform*, in *THE NEW WORLD OF WELFARE* 169, 170 (Rebecca M. Blank & Ron Haskins eds., 2001).

21. Spreng, *supra* note 20, at 287.

22. Heclo, *supra* note 20, at 171.

23. *Id.* at 171–72.

24. *Historical Development*, SOCIAL SECURITY ADMINISTRATION 5–6, tbl.6, <http://www.ssa.gov/history/pdf/histdev.pdf>.

25. Gertrude Schaffner Goldberg, *More than Reluctant: The United States of America*, in *DIMINISHING WELFARE: A CROSS-NATIONAL STUDY OF SOCIAL PROVISION* 33, 35 (Gertrude Schaffner Goldberg & Marguerite G. Rosenthal eds., 2002); SOCIAL SECURITY ADMINISTRATION, *supra* note 24, at 5–6, tbl.6.

them to give low-income and low-resource individuals medical assistance.²⁶ Between 1959 and 1973, the poverty rate in the United States was reduced by half, while real hourly wages rose each year.²⁷ The Earned Income Tax Credit (EITC) began in 1975, a refundable tax credit that generally supplements the wages of families classified as the working poor.²⁸ The 1980s brought a much more welfare-averse administration to Washington, D.C., which attempted to rid the United States of welfare-state status.²⁹ The Reagan Administration ended many programs that previously attempted to increase wealth for the lower classes, although Social Security continued in a constricted form.³⁰ The percentage of total income in the United States earned by the top 1% progressively increased during the 1980s and 1990s.³¹ The percentage of wealth held by the top 1% by the end of the 1990s almost reached pre-World War II heights, surpassing every other major Western nation.³²

The return of income disparity coincided with the reduction of marginal income-tax rates from almost 90% in the 1950s to 33% at present.³³ In 2004, the wealthiest 1% of the United States earned a before-tax income equal to that of the bottom 45%.³⁴ The top one-tenth of the top 1% received before-tax income equal to that of the bottom 28%, a population 280 times larger.³⁵ In 2009, census records showed that the top 20% held 84% of wealth in the United States.³⁶ The top 20% of households, which earned more than \$100,000 per year, reaped 49.4% of all household income.³⁷ The bottom 20%, which earned less than \$20,000 per year, reaped only 3.4% of all household income.³⁸ The ratio of earnings between the top 20% and the bottom 20% is almost double what it was when the Census Bureau began tracking the statistic in 1967.³⁹

26. SOCIAL SECURITY ADMINISTRATION, *supra* note 24, at 5–6, tbl.6.

27. Goldberg, *supra* note 25, at 39, 43.

28. *Id.* at 46. The program became quite successful. For example, it reduced childhood poverty about 14% to 21% in 1996 and raised hourly income between 34% and 40% for families with children in 1998. *Id.*

29. *Id.* at 36.

30. *Id.*

31. PHILLIPS, *supra* note 18, at 122–23, 123 chart 3.5.

32. *Id.*

33. David R. Francis, *Is the US System Rigged for the Rich?*, CHRISTIAN SCI. MONITOR (Oct. 18, 2010), <http://www.csmonitor.com/Commentary/David-R.-Francis/2010/1018/Is-the-US-system-rigged-for-the-rich>.

34. FURMAN ET AL., *supra* note 10, at 7.

35. *Id.*

36. Francis, *supra* note 33.

37. Boudway, *supra* note 3.

38. *Id.*

39. *Id.*

However, the top income-earning percentiles of Americans pay a disproportionately large percentage of taxes. The top 1% of income earners paid roughly 36.7% of federal income taxes and 25.3% of all federal taxes in 2004.⁴⁰ The highest 20% of income earners that year paid 67.1% of all federal taxes.⁴¹ By contrast, families comprising the lowest 40% of income earners, those making below \$36,300, generally did not pay any federal income tax in 2004 and received money from the government, mainly due to the earned-income tax credit.⁴²

B. History and Overview of Revenue Sharing in American Professional Sports

Sports are big business. Between 1989 and 2008, the average rate of return for premier American sports clubs surpassed that of the United States stock market.⁴³ Twenty American professional sports franchises are currently valued at more than one billion dollars each.⁴⁴ All are National Football League (NFL) teams except for the New York Yankees of Major League Baseball (MLB).⁴⁵

Although competition between teams is a central feature of the leagues,⁴⁶ working together, at times, is also important for their future success.⁴⁷ Accordingly, sports leagues were instituted to further this objective.⁴⁸ The leagues developed systems of revenue sharing to distribute income from high-income teams to low-income teams. Part of revenue sharing is giving the visiting team a share of the ticket sales for each game.⁴⁹ One might think that the highest revenue teams only distribute rather than benefit from revenue sharing, or that better teams draw more fans to home games than away games.⁵⁰ Due to ticket-sale sharing, however, when a high-

40. Edmund L. Andrews, *Bush Tax Cuts Offer Most for Very Rich, Study Finds*, N.Y. TIMES, Jan. 8, 2007, at A16.

41. *Id.*

42. *Id.*

43. Marc Edelman, *Why the "Single Entity" Defense Can Never Apply to NFL Clubs: A Primer on Property-Rights Theory in Professional Sports*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 891, 891–92 (2008).

44. Tom Van Riper, *The Most Valuable Teams in Sports*, FORBES (Jan. 13, 2010), http://www.forbes.com/2009/01/13/nfl-cowboys-yankees-biz-media-cx_tvr_0113values.html.

45. *Id.*

46. Teams compete on the field to attract fans, for ticket sales, for intellectual property, and for players and management personnel. *Am. Needle, Inc. v. Nat'l Football League*, 130 S. Ct. 2201, 2212–13 (2010).

47. *See id.* at 2213 (noting the common interest of NFL teams in promoting the NFL brand).

48. *Restructuring Professional Sports Leagues*, *supra* note 5, at 419–20 (describing how individual football teams came together to form a league).

49. MARK CONRAD, *THE BUSINESS OF SPORTS: A PRIMER FOR JOURNALISTS* 12 (2009); *see also* DAVID GEORGE SURDAM, *THE BALL GAME BIZ: AN INTRODUCTION TO THE ECONOMICS OF PROFESSIONAL TEAM SPORTS* 132–34 (2010) (chronicling the history of revenue sharing in professional sports, including the sharing of gate receipts from games).

50. SURDAM, *supra* note 50, at 134.

revenue team⁵¹ plays an away game, more fans go to the game, and the high-revenue team benefits from receiving a portion of the ticket sales for that game.⁵² For example, in the era following World War II, the New York Yankees and the Brooklyn Dodgers drew large numbers of fans to their away games, resulting in the teams incurring net benefits from revenue sharing.⁵³

MLB first implemented a limited amount of more formal revenue sharing with the 1996 Collective Bargaining Agreement.⁵⁴ The league further instituted revenue sharing in accordance with a study known as “The Blue Ribbon Report.”⁵⁵ In 2002, “[t]he Commissioner’s Blue Ribbon Panel on Baseball Economics . . . was formed to study whether revenue disparities among clubs [were] seriously damaging competitive balance, and, if so, to recommend structural reforms to ameliorate the problem.”⁵⁶ The panel concluded that “large and growing revenue disparities” were causing a “chronic competitive imbalance” in the league, despite the limited revenue sharing instituted in 1996.⁵⁷ It postulated that for teams to have high payrolls, an element that the panel found increasingly necessary for teams to succeed,⁵⁸ the teams needed to share between forty and fifty percent of all member clubs’ local revenues.⁵⁹ Players agreed, voicing their opinion that the absence of competitive balance was the biggest issue facing the league.⁶⁰

Total gross MLB revenues reached \$7 billion in 2010,⁶¹ yet the league is still experiencing revenue-sharing troubles. High-revenue teams want to en-

51. High-revenue teams usually garner high revenues because the teams are good and/or are popular, causing fans and spectators to attend their games and purchase their merchandise.

52. SURDAM, *supra* note 50, at 134.

53. *Id.*

54. LEVIN ET AL., *supra* note 6, at 1.

55. *Id.*; see also McCarthy, *supra* note 49, at 561 (explaining that changes made regarding revenue sharing and the luxury tax coincided with the recommendations of the Blue Ribbon Report).

56. LEVIN ET AL., *supra* note 6, at 1.

57. *Id.*

58. *Id.* at 8. A telling example of this is that the four highest paid players in MLB (all of whom play for the New York Yankees) were paid a combined salary in 2010 in excess of the total payroll of the bottom two teams, the Pittsburgh Pirates and the San Diego Padres, combined. MLB Salaries Database, *Top Player Salaries*, CBSSPORTS.COM (Apr. 4, 2010), <http://www.cbssports.com/mlb/salaries/top50>; MLB Salaries Database, *Team Payrolls*, CBSSPORTS.COM (Apr. 4, 2010), <http://www.cbssports.com/mlb/salaries/teams>. Considering that better players are generally paid more because their services are valued more highly by teams, teams with low payrolls cannot usually afford to pay better players. Consequently, low-payroll teams are generally unable to compete at the same level as teams with high payrolls.

59. LEVIN ET AL., *supra* note 6, at 8. Local revenues may consist of revenues from concessions, parking, local advertising, promotions, signage, print advertising, local sponsorship agreements, stadium clubs, luxury box income, Internet operations, and sales of programs and novelties. See *infra* notes 76–79 and accompanying text.

60. See LEVIN ET AL., *supra* note 6, at 11 (citing a *Baseball Weekly* player survey from the year 2000).

61. Pat Lackey, *MLB Revenue-sharing Dips Slightly in 2010*, AOLNEWS (Dec. 22, 2010), <http://www.aolnews.com/2010/12/22/mlb-revenue-sharing-down-slightly-in-2010/>.

sure that less competitive teams invest a portion of the money they receive through revenue sharing into acquiring good players.⁶² By acquiring good players, low-revenue teams should become more competitive, increasing the league's revenues and its reputation. The Pittsburgh Pirates are an example of how this can occur.⁶³ Since the Pirates have the lowest payroll in MLB, revenue sharing has allowed the team to spend the most money in the past three years on drafting players.⁶⁴ Better players entertain higher demand and require larger salaries, so it appears the Pirates are attempting to use revenue-sharing funds to increase the team's competitiveness on the field. This and other case studies will determine how effective revenue sharing is in increasing the competitiveness of low-revenue teams in the long run.

The NFL currently has the most extensive revenue-sharing system of any professional sports league, with over 90% of revenues being shared between franchises.⁶⁵ Revenue sharing in the NFL began in 1961 when Commissioner Pete Rozelle convinced team owners that by working together through combining resources and distributing profits, the total product of the NFL would be more valuable than it would be in separate pieces.⁶⁶ The first implementation of the NFL revenue-sharing system consisted of team owners selling the television broadcasting rights of the league as a package and evenly splitting the proceeds.⁶⁷ This agreement created a baseline financial balance between teams, which led, in turn, to a competitive balance between teams that continues to make the NFL incredibly popular.⁶⁸

The courts have echoed the need for teams to work together. In 1978, the United States Court of Appeals for the D.C. Circuit stated that "the ultimate success of the league depends on the economic *cooperation* rather than the economic competition of its members."⁶⁹ In a later case, the Ninth Circuit noted that professional football teams cannot profit without having other teams to play, saying that "in effect, all team revenue is jointly produced."⁷⁰ Due to revenue sharing, the financial success of professional sports teams hinges upon the monetary success of other teams in the same

62. SURDAM, *supra* note 50, at 137.

63. Ken Rosenthal, *Next Labor Deal Must Improve Balance*, FOX SPORTS: MLB (Sept. 8, 2010), <http://msn.foxsports.com/mlb/story/rosenthal-next-labor-deal-must-improve-competitive-balance-090710>.

64. *Id.*

65. Nathaniel Grow, Note, *There's No "I" in "League": Professional Sports Leagues and the Single Entity Defense*, 105 MICH. L. REV. 183, 193-94 (2006).

66. Moorhead, *supra* note 11, at 642.

67. *Id.*

68. *Id.*

69. Smith v. Pro Football, Inc., 593 F.2d 1173, 1195 (D.C. Cir. 1978) (MacKinnon, J., concurring in part and dissenting in part).

70. L.A. Mem'l Coliseum Comm'n v. Nat'l Football League, 726 F.2d 1381, 1402 n.1 (9th Cir. 1984) (Williams, J., concurring in part and dissenting in part).

league.⁷¹ Teams are thus incentivized to encourage the competitive development of other teams.

A great majority of the revenue presently shared in the NFL consists of revenue from the sale of television broadcasting rights and ticket sales, termed “Defined Gross Revenue.”⁷² As in MLB, ticket sales are divided for revenue sharing.⁷³ Unlike in MLB, visiting NFL teams’ shares are “pooled and shared equally among the 32 member clubs.”⁷⁴ The business model of NFL teams, however, is in the process of shifting due to the changing attitudes of owners.⁷⁵ The most profitable teams in the league are widening the revenue gap between high-revenue and low-revenue teams by capitalizing on sources of local revenue, which is unshared. The ability to exploit sources of local revenue is a product of a team’s market size and stadium ownership.⁷⁶ Team owners are beginning to identify the revenue benefits of owning the stadiums their teams play in, especially the profits from concessions, signage, luxury-box income, and stadium parking.⁷⁷ This priority and focus on unshared local revenue is dangerous for the NFL and for low-revenue teams because it risks undermining the revenue-sharing system, and consequently, the competitive balance of the league.⁷⁸ To combat this, the NFL adopted Resolution G-3 in 1999 to promote stadium construction, partly through a supplemental revenue-sharing pool.⁷⁹ By helping low-revenue teams build stadiums, the league hopes to help these teams acquire more local revenue and to keep the revenue gap from widening.⁸⁰ Owners of high-revenue teams with significant local revenue, like Jerry Jones, owner of the Dallas Cowboys, are currently fighting to reduce revenue sharing or eliminate it

71. Grow, *supra* note 66, at 194.

72. Moorhead, *supra* note 11, at 656.

73. *Id.*

74. NAT’L FOOTBALL LEAGUE, CONSTITUTION AND BYLAWS OF THE NATIONAL FOOTBALL LEAGUE, NFL RES. G-1 at 2001-2 (2001).

75. Moorhead, *supra* note 11, at 660–61.

76. *Id.* at 672.

77. See Heather J. Lawrence et al., *An Examination of Luxury Suite Ownership in Professional Sports*, 1 J. VENUE AND EVENT MGMT. 1, 1 (2009) (“Luxury suite revenue accounts for an average of \$9.8 million per professional sports venue annually in the United States.”); Moorhead, *supra* note 11, at 661 (describing the increased focus of NFL teams on stadium revenue, which is not subject to the league’s revenue-sharing program).

78. Moorhead, *supra* note 11, at 663–64. Want of local revenues has also increased the possibility of team owners seeking to move their teams to different cities in which they can build new stadiums or find more profitable deals with already-built stadiums, a phenomenon known as “franchise free agency.” *Id.* at 664; Don Nottingham, *Keeping the Home Team at Home: Antitrust and Trademark Law as Weapons in the Fight Against Professional Sports Franchise Relocation*, 75 U. COLO. L. REV. 1065, 1067–68 (2004). Franchise free agency has led Congress to propose legislation and has caused conflicts in court, while decreasing revenue in certain cases due to teams moving to cities with smaller television markets. Moorhead, *supra* note 11, at 664–65; Nottingham, *supra*, at 1074–79.

79. NAT’L FOOTBALL LEAGUE, *supra* note 75, at 1999-2; Moorhead, *supra* note 11, at 673–75.

80. Moorhead, *supra* note 11, at 673.

altogether out of the belief that the system hurts high-revenue teams.⁸¹ If revenue sharing in the NFL is curtailed or terminated, only time will tell how such action affects the competitive balance of the league.

III. The Successes and Failures of the Current Wealth Redistribution Systems in Professional Sports and in the Economy

A. *Wealth Redistribution in American Professional Sports Leagues*

The premise of revenue sharing in professional sports leagues seeks an ideal goal: monetary benefit for the teams contributing money to the revenue-sharing pool and for the beneficiaries who receive money from the revenue-sharing pool. Proponents of revenue sharing assert that a team with significantly lower revenues than other teams simply cannot compete without redistribution of team revenues.⁸² While this is generally true, there are exceptions, such as in the year 2010, when three of the eight MLB teams in the playoffs had total payrolls amounting to less than those of half the teams in the league.⁸³ High-revenue teams theoretically benefit from increased competition because in MLB, the money teams receive through revenue sharing must be spent to further the teams' performance on the field.⁸⁴ Increasing the ability of all teams in a league to compete increases the competitive balance of a league. The increase in the quality of a league greatly contributes to an overall increase in net value of franchises because it grows the fan base and increases the amount that potential business partners and fans are willing to pay to be a part of the industry. Increased competition is also necessary to compete against other forms of entertainment and leads to increased viewership,⁸⁵ which in turn leads to an increase in correlating factors such as merchandise sales, ticket sales, and media contracts.⁸⁶

81. TJ Sanders, *Enjoy This Season of Football NFL Fans, Next Year's Lockout Is Almost Here*, BLEACHER REPORT (Oct. 12, 2010), <http://bleacherreport.com/articles/489022-enjoy-this-season-of-football-nfl-fans-next-years-lockout-is-almost-here>; see Moorhead, *supra* note 11, at 663 (quoting Jones as supportive of an unshared-revenues system and arguing that unshared revenues allow teams to build new stadiums).

82. LEVIN ET AL., *supra* note 6, at 17.

83. See 2010 *Final Standings*, MLB.COM, http://mlb.mlb.com/mlb/standings/?ymd=20101003&tcid=mm_mlb_standings (2011) (reporting that the Tampa Bay Rays, the New York Yankees, the Minnesota Twins, the Texas Rangers, the Philadelphia Phillies, the Atlanta Braves, the Cincinnati Reds, and the San Francisco Giants made the playoffs); 2010 *Team Payrolls*, CBSSPORTS.COM, <http://www.cbssports.com/mlb/salaries/teams> (2011) (reporting that the Cincinnati Reds, the Tampa Bay Rays, and the Texas Rangers had team payrolls amounting to less than those of half of the teams in the league in 2010).

84. MAJOR LEAGUE BASEBALL, COLLECTIVE BARGAINING AGREEMENT 112 (2006).

85. See Erik M. Peterson & Arthur A. Raney, *Reconceptualizing and Reexamining Suspense as a Predictor of Mediated Sports Enjoyment*, 52 J. BROAD. & ELEC. MEDIA 544, 554 (2008) (concluding, based on a study, that viewers get the most enjoyment from watching "extremely close" and "rather close" sports games).

86. See Michael Smith, *TV Money Arriving Just in Time for Kentucky, Calipari*, STREET & SMITH'S SPORTS BUS. J. 1, 1 (2009) (quoting an IMG marketing executive as saying, "You think

Revenue sharing in American professional sports leagues may have certain imperfections. For example, the amount of money that high-revenue teams would be keeping if they were not forced to share revenue might outweigh the quantifiable benefit they reap from revenue sharing. While this may be true, leagues need low-revenue teams to be successful.⁸⁷ Without a sufficient number of successful teams, the owners of low-revenue teams might no longer be able to make money and their teams would fail. Over time, even the high-revenue teams would be unable to continue playing if they had no other teams to compete against.⁸⁸

Revenue sharing may also be criticized on the basis that it does not create perfect competition, which would be the essence of equality in sports. Perfect competition, however, is not the goal. Rather, the goal is competitive balance—giving each team in a league the opportunity to sufficiently compete and succeed. Perfect competition is the equivalent of equality of outcome in economic justice terms, in which everyone ends up with the same results. This is not the goal in sports. The objective of sports is to see who can overcome the competition and win, which perfect competition would eliminate.

Economics professors Yang-Ming Chang and Shane Sanders describe a “moral hazard” problem with revenue sharing.⁸⁹ They stress that lower-revenue teams may pocket the money they receive through revenue sharing instead of using it to increase their teams’ talent and consequently improve their teams’ play on the field.⁹⁰ MLB attempts to combat this possibility in its collective bargaining agreement by providing that “each Club shall use its revenue-sharing receipts . . . in an effort to improve its performance on the field.”⁹¹ While “improve its performance on the field” is sufficiently vague to be manipulated in many ways, the collective bargaining agreement also requires that a club receiving funds from revenue sharing “report on the

about sponsors, donor contributions, the pressure on ticket sales, premium seating, merchandise sales. It all goes up or down based on how the team is doing.”). An increase in the quality of lower-revenue teams is beneficial to low-revenue and high-revenue teams because more people will pay to watch competitive games live or on television, which increases ticket sales and the amount television networks will pay to broadcast the games. *Cf. id.* (explaining that ticket sales and payments from television networks will rise along with a team’s prominence).

87. See Moorhead, *supra* note 11, at 658–59 (asserting that, while some NFL owners may object to revenue sharing, it is necessary for the economic success of the league because the current economic inequality of the teams has threatened the survival of low-revenue teams).

88. See *id.* at 669 (“[I]t is impossible to ignore that the value of an individual franchise is completely dependent on the success of the League as a whole because without the League, each individual franchise would be worthless.”).

89. Yang-Ming Chang & Shane Sanders, *Pool Revenue-Sharing, Team Investments, and Competitive Balance in Professional Sports: A Theoretical Analysis*, 10 J. SPORTS ECON. 409, 409 (2009).

90. *Id.*

91. MAJOR LEAGUE BASEBALL, *supra* note 85, at 112.

performance-related uses to which it put its revenue-sharing receipts.”⁹² MLB reserves the right to penalize any club that violates these obligations.⁹³ Though the league has not technically penalized a team under these provisions, it recently showed its intent to enforce them by effectively reprimanding the Florida Marlins for using money acquired through revenue sharing to pay off debt.⁹⁴

Some owners argue that a minimum payroll would keep revenue-sharing beneficiaries from using the money they receive for expenses that are not performance related.⁹⁵ However, this puts low-revenue teams in small markets at a disadvantage because they have less ability to recoup their expenses through local revenues in ways that big-market teams can.⁹⁶ Additionally, a minimum payroll is detrimental to teams that are successful by making efficient management decisions that keep costs low, such as employing good scouting and signing talented players to inexpensive contracts.

Another proposed change to revenue sharing in professional sports is to distribute revenue-sharing funds in an amount no higher than the amount that an owner invests in his team each year.⁹⁷ Initially, this sounds like it might increase competitive balance because it is an incentive for owners to invest in their teams. Upon closer examination, however, the idea creates an irreconcilable problem for team owners that start without significant liquid assets.⁹⁸ These owners need funds from revenue sharing to improve their teams, make them competitive, and increase profits. Only upon making profits may these owners have enough money to significantly invest in their teams.

While revenue sharing certainly helps the competitive qualities of sports leagues, it is imperfect as a system of redistribution. Care must be taken to protect against revenue sharing's flaws when implementing into society the revenue-sharing premise that both beneficiaries and contributors should benefit. Society, which differs from the closed environment of professional sports leagues in many ways, also requires specialized adaptations for its own unique challenges.

92. *Id.* Some argue that increasing payroll and player development satisfy this requirement. Rosenthal, *supra* note 64.

93. MAJOR LEAGUE BASEBALL, *supra* note 85, at 112.

94. Rosenthal, *supra* note 64; *Statement from MLBPA, Florida Marlins and MLB Regarding Marlins' Compliance with Revenue-Sharing Provisions of the Basic Agreement*, MLB.COM (Jan. 12, 2010), http://mlb.mlb.com/news/article.jsp?ymd=20100112&content_id=7906016&vkey=news_mlb&fext=.jsp&c_id=mlb.

95. Rosenthal, *supra* note 64.

96. See Andrew E. Bortock, Note, *The Faux Fix: Why a Repeal of Major League Baseball's Antitrust Exemption Would Not Solve Its Severe Competitive Balance Problems*, 25 CARDOZO L. REV. 1069, 1096-97 (2004) (discussing the disparity in local revenues in MLB).

97. See McCarthy, *supra* note 49, at 567.

98. One reason for this is that an owner may have recently spent a significant portion of his liquid assets to purchase the team.

B. Wealth Redistribution in American Society as a Whole

The current method for redistributing wealth in the United States is a progressive taxation system structured so that the effective tax rate increases more than proportionately as the tax base increases. The central purpose of taxation is to supply the government with the money it needs to pay for its chosen spending, including the funding of welfare programs.⁹⁹ Redistributing wealth based on income percentile is flawed for a number of reasons.

First, such a system can act as a disincentive for an individual to work hard and earn more.¹⁰⁰ If extra effort expended to receive more benefits only results in benefits to others, individuals may lose the drive to perform at their best and to work to the fullest extent possible. This is especially true for someone who is on the brink of moving into a higher tax bracket by earning more income. Rather than lose an extra percentage of her income, she may believe the economically sound decision is to not work the extra time to earn the extra money, even if she was initially planning on working those hours.

Redistribution of wealth through progressive taxation also unequally confiscates the fruits of individual labor. A progressive tax system treats someone with a high income differently than someone with a low income by taking away a greater percentage of her earnings, effectively penalizing her for being more successful. Redistributing wealth through a progressive taxation system also narrows an individual's ability to make economic and financial decisions. High taxes consume large amounts of one's earnings, requiring that a higher percentage of one's income be spent on necessities. Consequently, earners have less discretionary income. In totality, this policy is restrictive of personal liberty.

Jason Furman, Lawrence Summers, and Jason Bordoff claim that progressive tax systems are appropriate because "high-income households have a greater ability to pay [taxes], . . . so their well-being is less affected by . . . an extra \$1,000 in taxes" and because a progressive tax system "can mitigate the impact of sharp losses in income, thus acting as a form of insurance [for the government]."¹⁰¹ If the highest income tax bracket is raised to a possible rate of forty-five percent, however, the difference is not likely to just be an extra \$1,000 in taxes. Rather, individuals will lose almost half of any amount of money they earn that surpasses a certain threshold. The argument that a progressive system acts as insurance may be correct, but it is not the appropriate policy. It is unequal for high-income individuals to bear the burden of insuring the government. If the government is to be

99. FURMAN ET AL., *supra* note 10, at 4.

100. See Joel Slemrod, 2002 Erwin N. Griswold Lecture Before the American College of Tax Counsel: "The Dynamic Tax Economist", 56 TAX LAW. 611, 613 (observing that higher marginal tax rates, resulting from tax progressivity, "provide . . . disincentives to earn more income").

101. FURMAN ET AL., *supra* note 10, at 10.

equally representative of each citizen, each citizen should have the same responsibilities to the government.

Furman, Summers, and Bordoff make a number of good points for why progressive taxation is the correct choice to remedy inequality.¹⁰² First, they assert that progressive taxation is better than industrial policies and direct market interventions because rather than trying to change the before-tax distribution of income, which harms the economy, the market is allowed to “maximize the total pie,” so that upon taxation, gains are spread more widely.¹⁰³ Second, they point out that progressive taxation can be targeted effectively to “avoid[] diverting scarce government resources to households who do not really need the assistance.”¹⁰⁴ Third, they argue that progressive taxation occurs on a grand enough scale, which allows it to make significant headway in reducing inequality.¹⁰⁵ Fourth, they point out that progressive taxation works instantaneously.¹⁰⁶ Despite these favorable aspects of progressive taxation systems, inequality remains a cornerstone of the scheme and the possibility of decreased production is real. Consequently, the large scale, instantaneous results, which would be a benefit of an appropriate policy, cannot be considered beneficial in regard to the inappropriate policy of progressive taxation. As an inappropriate policy with instantaneous effects, the detrimental impact is immediate.

Furthermore, some argue that taking larger amounts of money from high-income individuals hurts economic growth generally and stymies economic recovery currently.¹⁰⁷ In 2010, Republicans and some Democrats expressed concern that failing to extend the Bush tax cuts might slow economic recovery.¹⁰⁸ Those who would be taxed a higher percentage of their income were the ones with enough spending ability to boost the economy.¹⁰⁹ While some on the political left did not believe that the increased income

102. *Id.* at 11.

103. *Id.*

104. *Id.* at 11–12.

105. *Id.* at 12.

106. *Id.*

107. 151 CONG. REC. 10,868 (2005) (statement of Sen. Richard Shelby).

108. See Ben Adler, *Republicans Hold Senate Ransom for Rich Tax Cut*, WWW.NEWSWEEK.COM (Dec. 1, 2010), <http://www.newsweek.com/blogs/the-gaggle/2010/12/01/senate-republicans-hold-national-security-for-rich-tax-cut-ransom.html> (“Republicans have been demanding that the Bush tax cuts, due to expire, be fully and permanently extended.”); Jeanne Sahadi, *New Option: Don’t Tax the Rich. Tax the Really Rich*, CNNMONEY.COM (Oct. 13, 2010), http://money.cnn.com/2010/10/13/news/economy/bush_tax_cuts_possible_compromise/index.htm (stating Senate Banking Chairman Chris Dodd and Senator Blanche Lincoln, both Democrats “[wanted] to see” and were “amenable” to extending tax cuts for households making below \$500,000 and \$1 million, respectively).

109. See Jeanne Sahadi, *\$700 Billion Too Much? Why is \$3 Trillion OK?*, CNNMONEY.COM (Sept. 26, 2010), http://money.cnn.com/2010/09/23/news/economy/bush_tax_cuts_affordability/index.htm (“The theory for extending any of the tax cuts is this: If we don’t, the economic recovery could be thrown into reverse. Once the cuts expire, everyone’s tax bill will go up, and households will have less to spend.”).

taxes were wholly aimed at wealth redistribution, they agreed that wealth redistribution was among such policies.¹¹⁰

Even though Congress and President Obama decided to continue tax cuts for wealthy citizens, some Americans are less in favor of these tax cuts as the relevant income percentile increases.¹¹¹ They might argue, as Furman, Bordoff, and Summers did when they claimed “high-income households have a greater ability to pay [taxes], . . . so their well-being is less affected by . . . an extra \$1,000 in taxes,”¹¹² that people in the top percentile of income earners are excessively rich, so some of their money should be redistributed. The core of this argument is that those relinquishing income will not be harmed nearly as much, if at all, as those receiving the money will benefit. This view, however, is without reasoning. There is no logic in the statement that the more money someone has, the less they care about losing some of it. Additionally, despite the high incomes of the top percentile of income earners,¹¹³ the individuals at the top of this percentile skew the numbers to make the statistics sound more outlandish than they actually are. The top one-thousandth percentile of income earners, for instance, was composed of 145,000 individuals earning an average of three million dollars each in 2002.¹¹⁴ This is the same reason that, despite the fact that in 2009 the United States had the eighteenth-highest income per capita in the world (based on purchasing power),¹¹⁵ one would be misguided in saying that everyone is wealthy.¹¹⁶

Reducing wealth inequality using a tax system in which rates change automatically, according to income distribution, has been suggested.¹¹⁷ This seems to be a more efficient means of achieving wealth redistribution through taxation; however, it is inappropriate in the same manner as progressive taxation without automatically adjusting rates. Decreased production may occur, and a percentage of high-income individuals will be unequally treated by having to bear the burden of insuring the government

110. See, e.g., FURMAN ET AL., *supra* note 10, at 7–13 (extolling the virtues of progressive taxation as a remedy for wealth inequality).

111. See Sahadi, *supra* note 114 (describing the various proposals for raising the tax-cut income threshold to as high as \$1 million).

112. See *supra* note 103 and accompanying text.

113. See *supra* subpart II(A).

114. David Cay Johnston, *Richest Are Leaving Even the Rich Far Behind*, N.Y. TIMES, June 5, 2005, at 1.

115. WORLD BANK, GROSS NATIONAL INCOME PER CAPITA 2009 1 (2010), available at <http://siteresources.worldbank.org/DATASTATISTICS/Resources/GNIPC.pdf>.

116. The lowest 40% of income earners in America earn less than \$36,300 annually. See *supra* note 40 and accompanying text.

117. See FURMAN ET AL., *supra* note 10, at 13 (“Some scholars . . . would design the tax system such that rates automatically change as the distribution of income changes, to offset a certain percentage of the change in inequality.”).

and by having to contribute a disproportionate amount to the government compared to their fellow citizens.

Tax credits, or “tax expenditures,” are another proposed solution.¹¹⁸ Tax credits differ in that unlike deductions, which change in amount according to tax bracket, each credit is worth the same amount regardless of bracket.¹¹⁹ All uniform tax credits would do, however, is significantly decrease after-tax prices for high-income families, who would purchase anyway, rather than low-income families, who need more subsidies to make purchases they would not otherwise make.¹²⁰

The effects of being a beneficiary of wealth redistribution are also important. While proponents claim that there are no serious negative effects and that the benefits allow beneficiaries to survive,¹²¹ detractors claim that the beneficiaries are receiving money based on no work, which breeds laziness and inefficiency.¹²² The Tax Policy Center estimated that close to 47% of Americans would pay no federal income taxes in 2010.¹²³ This statistic evokes the criticism that these non-taxpaying Americans have less of an incentive to vote or work in favor of efficient government than do Americans who pay taxes because it is not their earnings that the government is using—they are not invested.

The American welfare state *is* incomplete. Rather than moving further on the spectrum and turning our “semiwelfare state” into a welfare state,¹²⁴ we can move forward by dismantling unappealing aspects of the welfare state to become a state in which everyone has an equal opportunity to benefit by pulling her own weight. Today’s economic challenges are complex and varied. The United States faces high unemployment, slowed growth, and a large deficit.¹²⁵ We must implement a system to overcome these challenges; otherwise, we face the risk of allowing wealth inequality to destabilize the

118. *See id.* at 24–26 (defining “tax expenditure” and recommending its role in the U.S. system of taxation).

119. *Id.*

120. *See id.* at 24 (“[S]ince deductions reduce the after-tax price more for high-income families than for low-income families, they generally produce too much added consumption by the former and too little by the latter.”).

121. *See* JOEL BLAU & MIMI ABRAMOVITZ, *THE DYNAMICS OF SOCIAL WELFARE POLICY* 38 (2004) (“[A]ccess to income and services outside the market enables people to survive . . .”).

122. *See* Katz, *supra* note 8, at ix (presenting the view that welfare “erodes the work ethic, retards productivity, and rewards the lazy”).

123. Robertson Williams, *Why Nearly Half of Americans Pay No Federal Income Tax*, TAX FACTS FROM THE TAX POLICY CENTER, 1149 (June 7, 2010), http://www.taxpolicycenter.org/UploadedPDF/412106_federal_income_tax.pdf. Other than actual income, reasons nearly half of Americans do not pay federal income taxes include credits, deductions, and exemptions. *Id.* This does not mean, however, that these people will not pay any taxes. Some will pay different forms of taxes, such as Medicare or Social Security payroll taxes. *Id.*

124. KATZ, *supra* note 8, at x.

125. John Sullivan, *Slow Growth for 2011, No Double-Dip Recession, Say Leading Economists*, ADVISORONE (Nov. 22, 2010), <http://www.advisorone.com/article/slow-growth-2011-no-double-dip-recession-say-leading-economists-1>.

country and create social unrest and insecurity.¹²⁶ The government is often slow to recognize and respond to changes in the economy.¹²⁷ The proposal I describe below will be administered by those who have shown an ability to run successful businesses and to understand markets, providing the system with the ability to remedy wealth inequality more quickly than the current welfare system.

IV. A Proposed Wealth Redistribution System Based on the Underlying Premise of Revenue Sharing in American Professional Sports

A. *The Details of the System*

The ideal method of wealth distribution is the creation of a system in which high-income individuals contribute an amount of money, through their already progressively higher tax bracket,¹²⁸ to create businesses in which low-income individuals can work. In addition to funding the creation of businesses, the contributions will be used to pay the low-income individuals' wages. High-income Americans will not be forced to participate in the system. If an individual does not want to partake, she may continue to pay progressively structured taxes as designated by the current wealth-redistribution scheme.¹²⁹ Some high-income individuals will not be allowed to participate. In an effort to increase the profitability of the businesses, only those who can show a profitable business history will be allowed to take part. High-income individuals that can show a profitable business history and decide to join the system will propose, design, and administer the created businesses, acting as executives and making critical management decisions.

Mid-level business managers will be members of the class of low-income beneficiaries. The opportunity to manage will empower the beneficiaries by placing them in leadership positions, give them the opportunity to earn higher wages, and provide them with quality professional training and experience for opportunities they encounter after elevating

126. UNITED NATIONS HUMAN SETTLEMENTS PROGRAMME, STATE OF THE WORLD'S CITIES 2008/2009: HARMONIOUS CITIES 56 (2008). According to the United Nations study, social unrest results from a perception of limited prospects in poor areas. *Id.* This perception often arises from inadequate health and education, an effect of wealth inequality. *Id.* Social unrest is detrimental to the economy because it creates disincentives to invest and requires a higher budget for internal security. *Id.* at 56–57.

127. See Kathleen E. Lange, Note, *The New Antifraud Rule: Is SEC Enforcement the Most Effective Way to Protect Investors from Hedge Fund Fraud?*, 77 *FORDHAM L. REV.* 851, 899–900 (2008) (asserting that “government regulators . . . are notoriously slow to respond to rapid market innovation”).

128. The criticisms of progressive taxation stated above are inapplicable here because those being taxed at a higher rate have the opportunity and incentive to regain the money paid, and more. See *supra* subpart III(B).

129. Without this alternative, no incentive exists for high-income individuals to participate in the proposed system.

themselves from a low-income level. Low-income Americans do not have to participate in the system; however, they must join the system in order to earn from it. Necessary participation in return for the fruits of one's labor does not push the boundaries of the Thirteenth Amendment's protection against slavery and involuntary servitude because no one is forced to take part in the system. If an individual chooses to participate, however, she enjoys many freedoms. Freedoms for beneficiaries include the ability to choose which company to work for based on individual interests and abilities, and the ability to use earned wages as one deems appropriate.

Profits from the businesses' success and growth will be used to make sure all beneficiaries are paid, all contributors are reimbursed, and contributors are compensated for the time and effort they give in overseeing the businesses. First, wages will be paid. Second, the contributors will be paid back for their contribution. Third, a portion of excess revenues will be shared by the workers. Fourth, contributors running the businesses will be paid a portion of excess revenues for their time and effort. The goal is for low-income beneficiaries to earn as much or more money through jobs in the proposed program as they would be making through welfare and a job, combined. Accordingly, the decision to participate in the system should be easy to make, even for individuals who are already employed.

Since beneficiaries of the system will be maximizing wealth through the profit-apportionment structure, they should be able to pull themselves and their families from poverty. It is only fair that once someone has pulled herself or her family out of poverty, another person in an unfortunate situation should be able to enter the profitable employment position. The threshold for this transition will be above the current threshold for receiving welfare to ensure that the person who pulled herself out of poverty is not in danger of falling back into poverty. The former beneficiary of the system will then be able to use her new experience and skills to gain employment and be successful.

To prevent any quantifiable benefit being taken away from high-income contributors, money paid to them will not be taxed as income. This will incentivize individuals with high incomes to participate and dedicate themselves to a business in the proposed system. The more profitable their company is, the more non-taxable income they will earn.

The class of contributors may include private and public companies. Companies will be able to participate in the system as contributors as long as they follow the rules, particularly those pertaining to profit apportionment. Companies may only participate, however, through the creation of wholly owned subsidiaries. This restriction ensures that all businesses in the system are evaluated in the same manner. Keeping a uniform structure will also simplify oversight of the program.

Oversight will be conducted by an agency created to supervise the program.¹³⁰ The agency will require each business to annually report its earnings and profit apportionment. The decision will be made, based on these reports, whether the business is fulfilling its duty to sufficiently compensate its employees in a manner that would make welfare unnecessary for them. If a business in the system is not sufficiently compensating employees or is unreasonably apportioning profit distributions, the business will receive one warning. If the compliance issues are not fixed within six months of the warning, the business will be shut down and will not be allowed to participate in the program.

The agency will also review each proposed business in an effort to prevent competition, making sure businesses are not competing in ways that negatively affect other businesses in the program. The question of how much competition is too much will be addressed when a business is proposed and, for approved businesses, on a yearly basis thereafter. Competition decisions will be made on a case-by-case basis, providing that each case's unique circumstances are fully reviewed. The "gray area" of what constitutes too much competition might lead to litigation based on contributors believing that their business will not be overly competitive with other businesses in the system. In an effort to preempt this litigation, which would be wholly unproductive and wasteful to the program, contributors who are proposing a business will be allowed one appeal of a competition ruling and may resubmit their proposal periodically. The appeal will be in place of adjudication in the courts. If the agency finds that an operating business is engaging in detrimental competition, mediation will be imposed to foster bilateral agreement to remedy the situation.

This system is based on the premise of revenue sharing in American professional sports, where the goal is quantifiable monetary benefits for all parties involved. Because everyone benefits if the revenue-sharing system runs correctly, all interested parties should, in theory, be satisfied with the outcome. Professional sports leagues are unique in that each one is a regulated business environment, not subject to certain forms of regulation,¹³¹ and without external intrusions. Due to the nature of professional sports leagues, teams receiving money through revenue sharing may profit without helping the competitive balance of the league, taking without contributing to the

130. Examples of similar entities include the Government Accountability Office, which works to "support . . . Congress in meeting its constitutional responsibilities and to help improve the performance and ensure the accountability of the federal government," and the Public Company Accounting Oversight Board, which exercises "external and independent oversight" of public companies in the United States. *About GAO*, GAO, <http://www.gao.gov/about/index.html> (2011); *About the PCAOB*, PCAOBUS, <http://pcaobus.org/About/Pages/default.aspx> (2011).

131. For example, professional sports leagues have consistently used revenue sharing as the basis for the defense that leagues are single entities and consequently are not subject to antitrust laws. See McCarthy, *supra* note 49, at 556–59 (chronicling the legal history of MLB's antitrust exemption).

league as a whole.¹³² The proposed system is without this unique, yet potentially detrimental feature, since in the proposed system, one cannot take without contributing. The NFL's current difficulty with team owners focusing on local unshared revenue demonstrates why society is a fertile environment for this solution.¹³³ Conversely to the NFL, under the proposed system, there is only one source of returns for the high-income contributors—profits from the businesses. Consequently, the contributors' sole focus will be making the created enterprises successful, which will benefit the low-income job and wage beneficiaries as well as the other high-income contributors who will share in the revenues of these enterprises. The fact that the system gives the high-income contributors a stake in helping the low-income members of society is very important. Without incentives to help low-income citizens, high-income citizens will become disinterested and will not work to make the businesses profitable.

B. *Why This System is Preferable to Others*

The primary reason that the proposed plan is preferable to other possible systems of wealth redistribution is that everyone reaps financial benefits. The plan also confers numerous additional social and economic benefits on society. Current welfare programs, which sustain recipients, are often not sufficient to extract individuals from poverty.¹³⁴ Unlike current programs, the proposed system can effectively pull members of society out of poverty by providing them with opportunities, not just to sustain themselves, but to prosper. The system gives low-income individuals experience and training they otherwise might not get. It boosts the economy by pumping in money and by creating jobs. While some of this money may enter the economy regardless, the required amount for the program *must* enter the economy.

Unlike large corporations, which focus on retaining their share of the market in which they do business, this system will bring innovation into the economy. The businesses created will need to grow to profit. To accomplish this, the businesses will have to focus on creativity, which will open new markets and create market share in existing markets.

The proposed system avoids the potential conflicts arising from a plan in which the government purchases social services from private entities, juxtaposing government frugality and intended efficiency with private profit making.¹³⁵ Since profit making will be the goal of both parties, confusion

132. See *supra* notes 91–96 and accompanying text.

133. Moorhead, *supra* note 11, at 660–61. Previously, NFL team owners maintained a collective mentality that equality and shared revenue were beneficial to the league. *Id.* at 660.

134. *Time on Welfare and Welfare Dependency: Hearing Before the Subcomm. on Human Res. of the H. Comm. on Ways & Means*, 104th Cong. (1996) (statement of LaDonna Pavetti), available at <http://www.urban.org/publications/900288.html> (detailing the prevalence of welfare recidivism).

135. See KATZ, *supra* note 8, at x (“[F]ranchising encourages the confusion of service with profit making and removes important public tasks—and a lot of money—from public oversight and scrutiny . . .”).

that often arises in exchanges between the government and private entities will not be an issue. Rather, a possible problem arises if the goal of profit making causes the program to endure too much risk.¹³⁶

A key aspect of the proposed system is that, unlike many other wealth-redistribution programs, it will actually benefit the economy through job creation and increased production. Since many of the low-income beneficiaries of the system may be unskilled, manufacturing has the potential to be a significant industry in which the proposed businesses engage. Rather than being a burden or difficulty within the system, uneducated and unskilled workers will thus be a cornerstone in the system. If a contributor determines that training for a certain position or skill is in line with her business model, however, the uneducated or unskilled worker will get the extra benefit of such training in addition to the monetary benefits of employment. Either way, the system will actively tackle unemployment and industrial underutilization, and will reduce outsourcing to other countries.

C. *Rebutting Criticisms of This System*

1. *Moral Hazard Problem.*—A common critique of welfare and revenue sharing generally is the moral hazard problem described by economics Professors Chang and Sanders, in which welfare recipients may choose not to cycle the money they receive back into the economy, or sports teams may choose not to invest the money they receive to help performance on the field.¹³⁷ In turn, those funding the distributions, and the economy, do not benefit.¹³⁸ With the proposed solution, this is not an issue. The contributors benefit from the businesses' production, as does the economy. Moreover, the recipients of the wages are not pressured to spend their wages and will not be maligned for failing to recycle the money they receive through the economy. Nor are the beneficiaries limited in how they can use the assistance they receive, as is the situation with food stamps and Medicaid.¹³⁹ Distribution is in the form of wages—cold, hard cash—so recipients may spend or save their money as they see fit.

2. *Welfare Trap.*—Another possible issue with a revenue-sharing system is the “welfare trap”—the possibility that receiving welfare will be

136. See *infra* subpart IV(C).

137. See *supra* notes 88–93 and accompanying text.

138. See Chang & Sanders, *supra* note 91, at 423 (describing baseball teams that did not invest their entire revenue-sharing receipts into on-field performance).

139. See Tom Curry & JoNel Aleccia, *Holes in the Safety Net: Medicaid Falls Short Just As Some Need It Most*, MSNBC.COM (Aug. 11, 2009), http://www.msnbc.msn.com/id/32127373/ns/health-health_care/ (describing various state restrictions and Medicaid patients' struggles to find qualifying doctors); *Eligible Food Items*, U.S. DEP'T OF AGRIC. (Oct. 5, 2010), <http://www.fns.usda.gov/snap/retailers/eligible.htm> (listing eligible and ineligible SNAP purchases).

more beneficial than getting a job because getting a job causes a net decline in combined income and benefits.¹⁴⁰ This is similar to a sports team choosing to spend very little in order to receive revenue-sharing funds, rather than choosing to spend a lot in an effort to win because spending in an effort to win reduces net profits. Under the proposed system, there can be no welfare trap problem. The potential beneficiaries can only decide whether they want the job—the money the beneficiaries receive is contingent on employment. The welfare trap problem is thus transformed into an incentive because the potential beneficiary must participate to benefit.

3. *Crowding Out Other Businesses.*—Businesses operating outside of the system, such as those in existence prior to the system's creation, may argue that businesses it creates will crowd markets, forcing existing businesses to close. This is only true if the existing businesses are operating inefficiently. Businesses created as a part of the system will benefit other businesses in the system and the economy as a whole by forcing existing businesses to innovate in an effort to be competitive. The creation of more businesses, inside and outside of the system, will also be spurred. New businesses will need to emerge to satisfy the buyers and sellers of businesses in the proposed system, as will manufacturers and suppliers. These companies will be able to work with one another, while creating jobs in various industries.

4. *Sharing Excess Revenues with Contributors and Beneficiaries Alike.*—A possible criticism of the system's profit apportionment is that the excess revenues of the businesses should be shared amongst the contributors, rather than given to workers based on performance, because this maintains the equality mentality of the system. Reward based on performance, however, is appropriate, because it incentivizes the beneficiaries to be successful.

5. *Contributors Lacking Motivation.*—Other wealth redistribution methods, such as taxation without duties, do not give the contributors a quantitative monetary benefit. Consequently, critics might conceivably point out that contributors will be careless and risky in the decisions they make as overseers because they have nothing to lose. This will not be the case. Since contributors will be compensated for their time and effort overseeing the businesses and will receive a portion of excess revenues, they will be incentivized to make smart financial decisions that lead to profitability.

140. See *Out of Sight, Out of Mind*, ECONOMIST (May 28, 2000), <http://www.economist.com/node/13747897> (explaining one reason that efforts to reduce poverty fail is that at some point, people who leave welfare to find work discover that "work does not pay").

V. Conclusion

The best way to remedy wealth inequality in the United States is to implement a system that seeks to attain what revenue sharing seeks to attain in American professional sports leagues: quantitative monetary benefits for all parties involved. The system proposed is efficient because it will react quickly to changes in the markets, it allows all types of income earners to keep the fruits of their labor, it empowers low-income citizens, it incentivizes productivity and efficiency, and it does not reduce individual liberty like the many systems and proposed solutions that take but do not give back. It is, in essence, the best capitalistic solution for wealth inequality in our country today.

F. Gibbons Addison

Third-Party Assistance in Determining Obviousness*

I. Introduction

Obviousness, despite being the newest of the patentability requirements,¹ has been described by patent law experts as the “ultimate condition of patentability,”² the “final gatekeeper of the patent system,”³ and “the heart of the patent law.”⁴ Yet despite the significance of the obviousness requirement in patent law, determining obviousness in a practical, yet technically and legally consistent manner during patent prosecution and litigation remains an elusive goal.⁵

In response to these challenges, this Note proposes the introduction of a third-party “obviousness specialist” that would bring more consistency to the obviousness determination in patent litigation while remaining within the doctrinal bounds established by the U.S. Supreme Court and the Court of Appeals for the Federal Circuit.

Part II of the Note presents a statement of the problem and briefly discusses the background of the obviousness question. Part III discusses the proposal for an obviousness specialist and situates the obviousness specialist role within the concepts of patent law. Part IV addresses the practical implementation options for the proposed obviousness specialist role. Part V presents arguments for the use of an obviousness specialist in light of the implementation options. Part VI concludes.

II. Statement of Problem

The challenge of making consistent determinations of obviousness is present at all levels of the patent system. Patent examiners, the Board of Patent Appeals and Interferences, the Federal Circuit, and the U.S. Supreme

* My thanks to the Texas Law Review editors and staff for their hard work in preparing this Note. I would also like to thank Professors Jane Cohen, Henry T.C. Hu, Robert Peroni, and Sean Petrie for their support, guidance, and wisdom in all of my endeavors at the law school and beyond. Finally, to my wife Jen and children Atticus and Wren—thank you for your continued love and support, even when my work runs long into the evening.

1. The concept of obviousness was first expounded in *Hotchkiss v. Greenwood*, 52 U.S. (11 How.) 248 (1851), and later added to the Patent Act in 1952. Pub. L. No. 82-593, 66 Stat. 792, 798 (1952).

2. NONOBVIOUSNESS—THE ULTIMATE CONDITION OF PATENTABILITY (John F. Witherspoon ed., 1980).

3. ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, PATENT LAW AND POLICY: CASES AND MATERIALS 612 (4th ed. 2007).

4. FED. TRADE COMM’N, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY, ch. 4, at 2 (2003) (quotations omitted).

5. See Gregory Mandel, *The Non-Obvious Problem: How the Indeterminate Nonobviousness Standard Produces Excessive Patent Grants*, 42 U.C. DAVIS L. REV. 57, 61 (2008) (noting that legal experts have struggled with the nonobviousness standard for over 150 years).

Court are all called upon to navigate the vague waters of what is “obvious.” Even though Congress,⁶ the U.S. Patent & Trademark Office (USPTO),⁷ and the courts⁸ have done their best to provide guidelines and a legal standard to delineate the boundaries of when an invention is or is not obvious, the standard is often easier to state than to apply. While inconsistent determinations of obviousness during patent examination are not socially optimal due to the negative externalities of “bad” patents,⁹ many of the determinations made by examiners during the patent application process will be inconsequential due to the high percentage of patents that are not commercially viable and are never practiced or contested.¹⁰

On the other hand, obviousness determinations made in the course of patent litigation often represent very real, very large, and very costly disputes¹¹ for the parties involved and may have broader market impacts in terms of product availability and pricing depending on the outcome of the litigation. As a result, developing patent litigation practices that provide greater consistency and certainty may provide large benefits, particularly as the more consistent litigation outcomes shaped by improved decision-making practices filter down into the patent examination procedures. Thus, based on the observations that obviousness remains a challenging concept for legal experts¹² and that many patent lawyers believe patent trials produce

6. 35 U.S.C. § 103(a) (2006). The statute states:

A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Id.

7. U.S. PATENT & TRADEMARK OFFICE, MANUAL OF PATENT EXAMINING PROCEDURE, EXAMINATION GUIDELINES FOR DETERMINING OBVIOUSNESS UNDER 35 U.S.C. 103 § 2141, at 2100-114 (8th ed., rev. 6 2007) [hereinafter MANUAL OF PATENT EXAMINING PROCEDURE].

8. *See, e.g.,* *Graham v. John Deere Co.*, 383 U.S. 1, 17–18 (1966) (laying out a multipart test for determining obviousness).

9. For example, “bad” patents may affect innovative activity at the margin if an inventor feels his idea treads too closely to an issued patent. Even though the questionable patent could potentially be invalidated on an obviousness challenge, the inventor may not wish to incur this risk and thus abandon his efforts.

10. The patent failure rate is 99.8%, meaning that roughly 3,000 out of 1.5 million are commercially viable. Karen E. Klein, *Avoiding the Inventor’s Lament*, BUSINESSWEEK ONLINE (Nov. 10, 2005), http://www.businessweek.com/smallbiz/content/nov2005/sb20051109_124661.htm (quoting USPTO Director of Public Affairs Richard Maulsby).

11. *See, e.g.,* Bart Showalter, Presentation at the American Intellectual Property Law Association Mid-Winter Conference: Cost of Patent Litigation (Jan. 25, 2008), *available at* http://www.aipla.org/Content/ContentGroups/Speaker_Papers/Mid-Winter1/20083/Showalt (indicating that for a patent case with more than \$25 million at stake the cost of litigating the case approached \$5 million in 2007).

12. Mandel, *supra* note 5, at 61.

“unpredictable,” “irrational,” and “unfair results,”¹³ improving the process of evaluating obviousness in litigation would be one valuable step towards improving patent-trial outcomes.

The doctrinal positioning of obviousness as a mixed question of fact and law is at the base of the challenge for determining obviousness during patent-invalidity litigation. Most recently, the U.S. Supreme Court, in *KSR Int'l Co. v. Teleflex Inc.*,¹⁴ affirmed that the “ultimate judgment of obviousness is a legal determination.”¹⁵ However, in the same opinion, the Court also reaffirmed that the legal determination of obviousness was to be informed by the objective, long-standing factual inquires set forth in *Graham v. John Deere Co.*¹⁶ Thus, a court tasked with answering the legal obviousness question must make its determination by reference to “several basic factual inquiries,” including (1) “the scope and content of the prior art,” (2) the “differences between the prior art and the claims at issue,” and (3) “the level of ordinary skill in the pertinent art.”¹⁷ “Against this background, the obviousness or nonobviousness of the subject matter is determined.”¹⁸ The Court in *KSR* further muddied the waters by reinforcing that these three factors must be supplemented by the consideration of a range of secondary factors that may provide additional indications of whether an invention was obvious, including whether the invention was “obvious to try.”¹⁹ Notably, and possibly indicative of the Court’s “expansive and flexible approach”²⁰ in *KSR*, the Federal Circuit has found patents in nonpharmaceutical cases to be obvious at a rate double that of the cases where the court has held patents nonobvious.²¹

Before beginning to consider how to best answer the obviousness question in patent litigation, it is helpful to begin by reviewing how *obviousness* is defined in patent law doctrine. While a textual analysis of § 103 of the Patent Act appears to set forth a simple comparison between

13. Paul R. Michel & Michelle Rhyu, *Improving Patent Jury Trials*, 6 FED. CIR. B.J. 89, 90 (1996).

14. 550 U.S. 398 (2007).

15. *Id.* at 427 (citing *Graham v. John Deere Co.*, 383 U.S. 1, 17 (1966)).

16. *Id.* at 406–07 (reiterating the objective standard set forth in *Graham*).

17. *Graham*, 383 U.S. at 17.

18. *Id.*

19. *KSR*, 550 U.S. at 421.

20. *Id.* at 415.

21. Robert H. Reis, *Lessons to Learn from Post-KSR Pharmaceutical Obviousness Decisions*, LANDSLIDE, Nov.–Dec. 2009, at 38. There were thirty nondrug precedential Federal Circuit decisions on obviousness through May 31, 2009. *Id.* Of these, the Federal Circuit found obviousness 67% of the time and nonobviousness 33% of the time. *Id.* Interestingly, the numbers are nearly inverted when looking at drug-related cases, with the Federal Circuit finding obviousness only 38% of the time and nonobviousness 62% of the time over a sample size of thirteen precedential decisions. *Id.*

prior art and the claims of the invention by a knowledgeable practitioner,²² a more nuanced look illustrates that the comparison is not so simple.

The first portion of the obviousness doctrine that remains unclear is at which point in the “invention” process the obviousness determination should be made. Specifically, patent doctrine recognizes separate constructs for the time of conception of an invention and the point at which the invention is considered either actually or constructively reduced to practice.²³ However, existing obviousness doctrine provides little guidance on how to address the obviousness cases at the two ends of this spectrum—the invention in which the initial concept is nonobvious but the reduction to practice is simple following conception,²⁴ and alternatively, the invention in which the concept is obvious but the reduction to practice is nonobvious.²⁵ While current doctrine appears to require that the obviousness determination be considered at a single point in time, this approach does not correlate well with the realities of the development of some types of innovations.²⁶

This indeterminacy in the obviousness standard has several consequences. Following the Court’s holding in *KSR*, it appears the Court was focused primarily on “assessing the inventive leap” based upon the identified differences between the prior art and the claims at issue.²⁷ Yet observers of Federal Circuit precedent following the *KSR* decision report that the subsequent case law has been unclear as to how much of the obviousness inquiry is to be directed at conception, reduction to practice, or to some combination thereof.²⁸ While the Federal Circuit has historically attempted to provide some guidance by indicating that the obviousness inquiry should consider whether the invention would have a reasonable expectation of success, this standard fails to fully answer the question of which point in time in the inventive process a fact finder is expected to consider obviousness.²⁹ As a result, juries in different patent trials may be given very different instructions on obviousness that may direct their attention in different ways when considering obviousness at conception or at reduction to practice.

22. See 35 U.S.C. § 103(a) (2006) (paraphrasing the standard set forth in the statute).

23. See, e.g., MANUAL OF PATENT EXAMINING PROCEDURE, *supra* note 7, at § 2138.05 (outlining the concept of “reduction to practice” in patent law).

24. See Mandel, *supra* note 5, at 61 (suggesting that Post-It notes are an example of this situation).

25. See *id.* (using the example of the incandescent light bulb patent scenario in which the concept of using a carbonized filament was obvious but determining which specific filament to use was nonobvious).

26. See *id.* (providing some examples of real inventions that were alternately obvious and nonobvious at different stages of their development and reduction to practice).

27. Jeanne C. Fromer, *The Layers of Obviousness in Patent Law*, 22 HARV. J.L. & TECH. 75, 83 (2008).

28. *Id.*

29. *Id.*; see also *In re O’Farrell*, 853 F.2d 894, 903 (Fed. Cir. 1988) (discussing an invention that was held obvious because the prior art suggested a reasonable expectation of success).

This indeterminacy is not lost on the laypersons that serve on patent-trial juries. Jurors are generally one-time participants in the litigation process.³⁰ Because of the singular nature of their jury service, jurors have no opportunity to build an understanding or common viewpoint that would allow them to build on prior experience and reasoning as a supplement for an indeterminate obviousness standard in the manner that experienced judges are able to develop ancillary considerations when employing an indefinite standard in multiple cases over time.³¹ Thus, in order to allow patent-trial juries to deliver consistent and correct factual determinations, the standards for what they are being asked to consider must contain definitive boundaries. For the obviousness analysis, this includes providing guidance as to the specific points in time at which they are to consider the obviousness factual questions. Without such definitive guidance, there is a risk that some jurors will rely on their own sense of “gee whiz” intuition about whether the concept would have appeared nonobvious at the time they believe is most important, while others, if they are able to separate the realms of conception and reduction to practice, may evaluate the question of obviousness in the appropriate light of these points in time.³²

Beyond the potential confusion over the indefinite extent of the obviousness doctrine, the determination of obviousness as a mixed question of law and fact has fueled an ongoing debate over whether a judge, a jury, or some combination of the two should be responsible for making the obviousness determination.³³ Some suggest that a standard jury, composed of a mix of citizens with only average education and intelligence, cannot properly decide complex technical issues such as obviousness.³⁴ Other commentators insist that jurors are likely to be no more or no less competent in deciding technical issues than a judge with a generalist background.³⁵ While an

30. Kimberly A. Moore, *Judges, Juries, and Patent Cases—An Empirical Peek Inside the Black Box*, 99 MICH. L. REV. 365, 389 (2000).

31. *Id.*

32. See Rochelle Cooper Dreyfuss, *Are Business Method Patents Bad for Business?*, 16 SANTA CLARA COMPUTER & HIGH TECH. L.J. 263, 270 (2000) (describing how judges are more likely to protect patents covering fields that the judges do not understand). Though Professor Dreyfuss attributes the “gee whiz” factor to judges in patent trials, there is no reason to suspect that jurors would not also be subject to the same biases in attempting to distinguish novel or nonobvious advancements in light of their own predispositions.

33. Though addressed in this Note in the context of patent law, the challenge of mixed questions of law and fact is not limited to patent law and is a long-standing and often-debated topic. See, e.g., Francis H. Bohlen, *Mixed Questions of Law and Fact*, 72 U. PA. L. REV. 111 (1924) (describing the challenges of dealing with mixed questions of law and fact in the context of negligence claims).

34. See Michel & Rhyu, *supra* note 13, at 90–91 (discussing the argument that juries are unable to adequately resolve complex cases); Moore, *supra* note 30, at 369–74 (discussing a number of views of perceived juror incompetence).

35. Michel & Rhyu, *supra* note 13, at 90–91; see also Moore, *supra* note 30, at 373–74 (describing the view that juries may be better equipped to decide technical issues than judges due to the need for judges to split their attention among many cases on the docket).

argument could be made that judges in particular districts may have developed a greater expertise in dealing with technical issues in patent trials (e.g., Judges Ward and Davis in the Eastern District of Texas),³⁶ these isolated specialists are not common and the existence of such specialist judges, outside the confines of an “expert judge” system, may raise broader equal protection of law questions.

Independent from the question of whether the judge or the jury is making the obviousness determination, operationalizing the *Graham* test factors set forth by the Federal Circuit and the Supreme Court remains a significant challenge. While juries are often asked to place themselves in the proverbial shoes of another person at another time in tort litigation, the three-part *Graham* analysis³⁷ asks the fact finder—whether judge or jury—to go much further. Unlike the general “reasonable person” standard set forth in tort litigation, the “person having ordinary skill in the art” standard used in the obviousness determination asks the fact finders not only to place themselves into a scenario in which their perspective must be that of a person with a particular set of technical knowledge that is likely foreign to the individual, but furthermore, to predict what that hypothetical skilled person might have considered creative at some point in the past.³⁸

Both commentators and judges have stated that this type of analysis is unlikely to be consistent or reliable. As stated by Judge Learned Hand, “Courts, made up of laymen as they must be, are likely either to underrate, or to overrate, the difficulties in making new and profitable discoveries in fields with which they cannot be familiar”³⁹ Similarly, Professor Mandel states, “Such a judgment [whether an advance would be obvious from the perspective of another] is epistemically impractical. Due to the ‘curse of knowledge,’ individuals are cognitively incapable of accurately making judgments from other individuals’ perspectives.”⁴⁰ Furthermore, Professor Dreyfuss provides additional circumstantial evidence to support this concern by reviewing the outcomes of three Supreme Court patent cases on the same day—the oft-cited *Graham*; *Calmar, Inc. v. Cook Chemical Co.*,⁴¹ about a spray can; and *United States v. Adams*,⁴² about a battery.⁴³ As Professor Dreyfuss points out, while there are few distinguishable differences in the

36. See, e.g., Yan Leychikis, *Of Fire Ants and Claim Construction: An Empirical Study of the Meteoric Rise of the Eastern District of Texas as a Preeminent Forum for Patent Litigation*, 9 YALE J.L. & TECH. 193, 205–08 (2007) (describing patent litigation in the Eastern District of Texas under the “patent ‘rock star’” Judge Ward and the “relatively recently” appointed Judge Davis).

37. See *supra* notes 17–18 and accompanying text.

38. See *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 420–21 (2007) (indicating that the obviousness determination requires consideration of the “creativity” of a person of ordinary skill in the art).

39. *Safety Car Heating & Lighting Co. v. Gen. Electric Co.*, 155 F.2d 937, 939 (2d Cir. 1946).

40. Mandel, *supra* note 5, at 59–60.

41. 383 U.S. 1 (1966) (consolidated in *Graham v. John Deere Co.*, 383 U.S. 1 (1966)).

42. 383 U.S. 39 (1966).

43. Dreyfuss, *supra* note 32, at 269–70.

three cases, the patents in two of the cases were invalidated for obviousness while the third was upheld.⁴⁴ The two invalidated patents were directed towards what may have appeared to be more mundane mechanical developments while the battery patent involving electricity may have been viewed as somewhat more mysterious. As a result, there is a question of just how well a fact finder can make an obviousness determination without regard to his or her own familiarity with the invention's underlying technical basis.⁴⁵

Overcoming this “curse of knowledge”—or otherwise termed “hindsight bias”—challenge is another well-known operational issue in attempting to make obviousness determinations.⁴⁶ Professor Mandel has written several articles on the impact of hindsight bias in patent law.⁴⁷ His empirical results illustrate that when a jury member is simply aware that an invention has been achieved, it will increase that individual's likelihood of finding that the invention was obvious.⁴⁸ Stated differently, when the jury is initially presented with the knowledge of an invention that has been made, they are more likely to find obviousness, as opposed to the situation of considering only the scope of the prior art without affirmative knowledge that an invention has been made in the art.

These operational issues arise regardless of whether the judge or the jury is making the obviousness determination. In the event that the judge is the primary decision maker for the obviousness question, the judge must attempt to perform this challenging cognitive task as an individual, without the benefit of multiple viewpoints or reasoned debate. On the other hand, a jury gains the advantage of being able to make an obviousness determination with the benefit of multiple perspectives on each of the *Graham* factual inquiries but without the judge's benefit of familiarity with this type of analysis. Fortunately for the jury, it has become common practice to use special-verdict forms to guide its decision process by focusing the jurors' attention on the relevant factual questions.⁴⁹ This should mitigate some, if not most, of the arguments that a judge is better equipped to make the

44. *Id.*

45. *Id.*

46. See *Graham*, 383 U.S. at 36 (stating that legal inferences that focus on economic and motivational rather than technical issues help judges resist the “temptation to read into the prior art the teachings of the invention in issue” and “guard against slipping into the use of hindsight” (quoting *Monroe Auto Equip. Co. v. Heckethorn Mfg. & Supply Co.*, 332 F.2d 406, 412 (6th Cir. 1964))).

47. Gregory N. Mandel, *Another Missed Opportunity: The Supreme Court's Failure to Define Nonobviousness or Combat Hindsight Bias in KSR v. Teleflex*, 12 LEWIS & CLARK L. REV. 323 (2008); Gregory N. Mandel, *Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational*, 67 OHIO ST. L.J. 1391 (2006); Gregory Mandel, *Patently Non-Obvious II: Experimental Study on the Hindsight Issue Before the Supreme Court in KSR v. Teleflex*, 9 YALE J.L. & TECH. 1 (2007) [hereinafter Mandel, *Patently Non-Obvious II*].

48. Mandel, *Patently Non-Obvious II*, *supra* note 47, at 15, 17–18.

49. See Kimberly A. Moore, *Juries, Patent Cases, & a Lack of Transparency*, 39 HOUS. L. REV. 779, 783 (2002) (noting that “nearly every patent case utilized special verdicts or interrogatories, rather than general verdicts, to query the jury about its decision-making”).

obviousness determination simply due to familiarity with the analytical process.

However, the jury may still fare no better in making a determination of obviousness because, despite having the benefit of multiple perspectives, it faces the same challenges of determining a level of creativity in a likely unknown technical field as discussed above. Additionally, a jury-led decision on obviousness may result in more troubling systemic legal issues. Depending on how the special-verdict questions are presented to the jury, there may be situations in which the obviousness determination contains conflicting factual and legal conclusions on obviousness or situations in which the jury ultimately may provide an answer only to the legal question of obviousness without providing definite answers to the *Graham* factual inquiries. Without a record of the detailed factual findings by the jury, the appellate review of the obviousness determination must rely on assumptions that the jury found a proper factual basis for its conclusion.⁵⁰ This results in a less than optimal scenario for the reviewing appellate panel as the jury verdict is virtually unreviewable and often requires the panel to review the entire record to determine whether there was substantial evidence to support the verdict.⁵¹

Additionally, the interrogatories and special-verdict forms used in many patent trials have not been well drafted in view of their purpose of allowing jurors to focus on answering the specific factual questions without regard to the effect of the answers on the legal outcome for the parties.⁵² Judge Moore of the Federal Circuit provides several examples of poorly drafted special verdict forms used in patent cases, including the following:

3. Has Guardian proved by clear and convincing evidence that claim 4 is invalid for obviousness? YES _____ (for Guardian) []
NO _____ (for PPG).⁵³

Not only does this type of question fail to provide any insight to the jury's findings on the factual inquiries required under *Graham*, it also undermines the purpose of a special verdict as a means for the jury to answer factual questions without regard to the effect of the answers on the parties.

Even well-intentioned special-verdict forms are not always a reliable method of eliciting proper jury review of obviousness. The Federal Circuit recently had to vacate a Delaware district court decision in which the jury was provided a special-verdict form regarding the obviousness of a certain

50. See, e.g., *Kinetic Concepts, Inc. v. Blue Sky Med. Grp., Inc.*, 554 F.3d 1010, 1021 (Fed. Cir. 2009) (“Thus, we must *assume* that the jury found that the prior art does not disclose ‘treating a wound with negative pressure’ within the meaning of the patents.” (emphasis added)), *cert. denied*, 130 S. Ct. 624 (2009).

51. Moore, *supra* note 49, at 791.

52. See *id.* at 783–84 (discussing the purpose of special verdicts).

53. *Id.* at 785. The question was used in *PPG Industries, Inc. v. Guardian Industries Corp.*, Civ. No. 94-1112 (W.D. Pa. 1994). Moore, *supra* note 49, at 785 n.30.

type of golf-ball construction.⁵⁴ Despite being given a special-verdict form to determine the obviousness of the eight patent claims, the jury returned a verdict that found one of the dependent claims to be invalid for obviousness and found the independent claim from which the dependent claim was derived to be valid.⁵⁵ This is an impossible condition in patent law due to the fact that a dependent claim can only add a supplemental limitation to the parent independent claim.⁵⁶ In other words, if the independent claim is valid, the dependent claims must be valid by definition. Not only does the district court's original verdict on obviousness illustrate that the jurors fundamentally misunderstood the basic structure of the claims they were considering, it also shows that even judges in well-regarded patent-litigation districts may get lost in the complex details of making an obviousness determination, even when the patent claims at issue involve the relatively simple subject matter of a golf ball constructed of three types of material.

These types of recurring technical and doctrinal challenges in obviousness determinations are exemplified by the petition for certiorari of the recent Federal Circuit decision in *Kinetic Concepts, Inc. v. Blue Sky Medical Group, Inc.*⁵⁷ (The petition caption uses the party names *Medela AG v. Kinetic Concepts, Inc.*)⁵⁸ The question presented in the petition was “[w]hether a person accused of patent infringement has a right to independent judicial, as distinct from lay jury, determination of whether an asserted patent claim satisfies the ‘non-obvious subject matter’ condition for patentability.”⁵⁹ The petition was the result of a challenging case involving a patent for treating wounds through the use of negative pressure as opposed to the more traditional method of applying positive pressure.⁶⁰ The question of obviousness was sent to the jury via an interrogatory that asked for a “yes” or “no” answer as to whether each of the thirty-seven claims were obvious.⁶¹ Each question was stated as follows:

Have the Defendants proven by clear and convincing evidence that the invention claimed in any of the following claims of the ‘643 patent would have been obvious to a person of ordinary skill in the art at the time the invention was made, in light of the prior art? In answering

54. *Callaway Golf Co. v. Acushnet Co.*, 576 F.3d 1331, 1343–45 (Fed. Cir. 2009).

55. *Id.*

56. Gene Quinn, *Callaway Golf Loses Jury Verdict at the Federal Circuit*, IPWATCHDOG (Aug. 15, 2009), <http://www.ipwatchdog.com/2009/08/14/callaway-golf-loses-jury-verdict-at-the-federal-circuit/id=4872/>.

57. 554 F.3d 1010 (Fed. Cir. 2009), *cert. denied*, 130 S. Ct. 624 (2009).

58. Petition for a Writ of Certiorari, *Medela AG v. Kinetic Concepts, Inc.*, 130 S. Ct. 624 (2009) (No. 09-198), 2009 WL 2509227.

59. *Id.* at i.

60. *See id.* at 4 (describing the district court proceedings).

61. *Id.* at 6.

Question No. 17, please refer to pages 25-34 and 38-45 of the Court's Instructions.⁶²

The instructions mentioned in the question included a listing of the *Graham* factors as well as additional details on basic patent-law doctrine such as what constitutes an invention having been "made."⁶³

The jury returned the verdict form having answered all thirty-seven questions "no," meaning that none of the claims were invalid due to obviousness.⁶⁴ The jury gave no additional explanation as to how it arrived at answers to all thirty-seven questions, nor did the jury provide any indication as to which facts were considered in making its decision.⁶⁵ The district judge in turn simply took the jury's answers, and, finding no reason to disturb the jury verdict, entered judgment that reflected the jury's finding of nonobviousness.⁶⁶ This is particularly notable in light of the Supreme Court's statement in *KSR* that the obviousness analysis "should be made explicit."⁶⁷ Not only was the obviousness analysis not explicit, but the court effectively turned the ultimate legal question of obviousness into a factual question to be answered solely by the jury by simply approving the jury verdict without further analysis. On appeal, the Federal Circuit was limited to assuming that the jury had found sufficient evidence because the interrogatories provided no factual basis for the verdict, and as such, the Federal Circuit upheld the jury's verdict of nonobviousness.⁶⁸

As a result, the *Medela AG* petition sought to have the Supreme Court establish a definitive set of guidelines on if, or when, a litigant can require the court to make an explicit legal finding on the question of obviousness independent of the jury. In setting forth their argument for the Court to grant certiorari, the petitioners highlighted one of the cumulative effects of the various challenges noted above—specifically an unresolved split in circuit court precedent on the power of the jury to decide the obviousness question.⁶⁹ Additionally, because of the limited and indeterminate guidance on how to effectively separate the legal question of obviousness from the

62. *Id.*

63. *Id.* at 5.

64. *Id.* at 6.

65. *Id.* at 6–7.

66. See *Kinetic Concepts, Inc. v. Blue Sky Med. Grp., Inc.*, 554 F.3d 1010, 1020 (Fed. Cir. 2009) (noting that the district court simply reviewed the jury's verdict for substantial evidence), *cert. denied*, 130 S. Ct. 624 (2009).

67. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 418 (2007).

68. *Kinetic Concepts*, 554 F.3d at 1021.

69. Petition for a Writ of Certiorari, *supra* note 58, at 11–17. The Federal Circuit, in consistently holding that the obviousness issue can be submitted to the jury, has explicitly rejected competing views of the Fifth and Seventh Circuits that allow subsidiary factual questions of obviousness to be submitted to the jury for the rendering of advisory verdicts, but reserve the legal determination of obviousness for the judge. *Id.* at 11, 13 (citing *Roberts v. Sears, Roebuck & Co.*, 723 F.2d 1324 (7th Cir. 1983) (en banc); *Sarkisian v. Winn-Proof Corp.*, 688 F.2d 647 (9th Cir. 1982) (en banc) (per curiam)).

underlying factual inquiries, the petition for certiorari, along with amici curiae briefs from a variety of large technology companies, argued that by providing additional procedural certainty to the obviousness determination, the outcome of invalidity litigation would be more uniform and predictable.⁷⁰

While this Note does not purport to answer the question posed in the petition for certiorari, there may be an alternative method that would improve the obviousness determination without resorting to a strict dichotomy of assigning the analysis to either the court or the jury.

III. Proposal

In response to the general challenges previously identified and against the backdrop of the *Medela AG* petition for certiorari, I propose the creation of a third-party obviousness specialist or “master” in invalidity litigation to assist in developing an obviousness determination that is both technically consistent and doctrinally correct.⁷¹

At a high level, the goal of the obviousness specialist is not to supplant either the judge or jury in making the ultimate obviousness determination. Instead, the obviousness master would assist the court and the litigants in establishing the boundaries of the obviousness factual inquiries. Although it does not appear that courts or litigants have attempted to utilize a third-party specialist for obviousness outside of the traditional use of expert witnesses during trial, there is support for employing new innovations in making the obviousness determination and support for developing a form of the obviousness-specialist role in both academic commentary and existing legal doctrine.

In order to situate the following discussion, a brief overview of the manner in which a third-party obviousness master would be implemented is necessary. The details of these forms of implementation will be more fully developed in Part IV. The concept of a third-party obviousness specialist could take two primary forms. The first is the establishment of a traditionally titled special-master role⁷² that would be involved in an “ordinary skill in the art” and prior art interpretation hearing modeled loosely after the

70. *Id.* at 28; Brief for Apple, Inc., Cisco Systems, Inc., Google, Inc., Microsoft Corp., Symantec Corp., and Yahoo!, Inc. as *Amici Curiae* Supporting Petitioners at 6, *Medela AG v. Kinetic Concepts, Inc.*, 130 S. Ct. 624 (2009) (No. 09-198), 2009 WL 2879049; Brief of Amici Curiae Intel Corporation, SAP America, Inc., and Shoretel, Inc., in Support of Petitioners at 5, *Medela AG v. Kinetic Concepts, Inc.*, 130 S. Ct. 624 (2009) (No. 09-198), 2009 WL 2979048.

71. Note that choosing the word *master* at this point in the discussion to describe this new role in litigation is intended to simplify nomenclature and is not intended at this point in the discussion to directly invoke the legal concept of a special master. A more thorough analysis of the implementation of an obviousness master in the model of a special master is included in Part IV, *infra*.

72. See, e.g., Don W. Martens, *Use of a Special Master for Patent Claim Construction*, KNOBBE, MARTENS, OLSON & BEAR, LLP 1 (2004), http://www.knobb.com/pdf/spec_master_patent_claim.pdf (noting the historical use of patent attorneys as special masters for resolving discovery disputes and assisting at trial).

procedural structure used for *Markman* claim-construction hearings.⁷³ The second form would utilize an obviousness master through the use of an expert witness that would be selected through a special double-blind selection process derived from social science principles and medical malpractice litigation concepts.⁷⁴ The academic and legal support for these two general approaches for an obviousness specialist is discussed below.

A. *Support in Literature*

A prime example of the support for considering new methods in conducting obviousness evaluations are published comments from former Federal Circuit Chief Judge Michel. In an article that is now more than ten years old, Judge Michel and his coauthor called for federal trial judges to employ innovative techniques in patent trials to increase the “rationality, predictability, and reliability” of juries.⁷⁵ The article discusses several different methods with which the court, in conjunction with the parties, can provide greater understanding to the jury of both the applicable rules and the technical content of the dispute in forms that are outside of the traditional avenues through which evidentiary material is presented.⁷⁶ The methods discussed include having the parties provide the jury with a jointly developed technical-term glossary and requiring the parties to create a tutorial-like video describing the relevant technology in the case.⁷⁷

Former Chief Judge Michel is not alone on the bench in his belief that judges should seek out additional assistance in understanding the intricacies of patent litigation. Chief Judge James Holderman in the Northern District of Illinois has also called for judges to more broadly employ special masters not only to assist in claim construction, but also to assist with other tasks in patent suits that require a technical background in order to make an informed recommendation.⁷⁸

The introduction of an unbiased third-party obviousness specialist to establish the factual boundaries of the prior art would appear to fit into Chief Judge Michel and Chief Judge Holderman’s theme of encouraging innovation at the trial court level in patent cases. Though the introduction of a third party to assist in determining the boundaries of skill in the art and prior art may be unconventional, it would seem to be a small step beyond what Chief Judge Michel has endorsed, provided of course that both parties were willing

73. See *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 391 (1996) (holding that the interpretation of a patent claim is an issue for the judge, not the jury).

74. See generally Christopher Tarver Robertson, *Blind Expertise*, 85 N.Y.U. L. REV. 174 (2010) (advocating a procedural system for civil litigation in which litigants and experts are separated by intermediaries).

75. Michel & Rhyu, *supra* note 13, at 104–05.

76. *Id.*

77. *Id.* at 104.

78. James F. Holderman & Halley Guren, *The Patent Litigation Predicament in the United States*, 2007 U. ILL. J.L. TECH. & POL’Y 101, 117–18.

participants. For example, extending his suggestion that the jury be provided a glossary of technical terms jointly prepared by the parties, the jury could be provided with an outline of the major features and characteristics of the relevant prior art along with a representative list of skills that a person skilled in the art would possess. The parties would retain control over the material provided to the jury but would employ a jointly selected specialist in the art to assist in the preparation of the outline. The outline would then be used by the jurors as a reference guide in their *Graham* factual analysis but would not supplant the traditional forms of testimony and evidence presented at trial.

This type of proposal exists with a caveat, however. While Chief Judge Michel might be amenable to such innovation and has been instrumental in actively developing Federal Circuit patent case law independent of congressional legislative action, it remains to be seen if his favorable view of trial court procedural innovation will continue to be endorsed by the Federal Circuit now that he has left the bench.⁷⁹

Other patent law commentators have suggested using a separate specialist to perform the obviousness analysis independent of the determination of the other patentability requirements during the initial examination of a patent application.⁸⁰ By separating the determination of obviousness from the evaluation of the other requirements for patentability, one specialist responsible for the analysis of all patentability requirements other than obviousness could provide a second specialist focused solely on the obviousness analysis with the information on the problem domain, the relevant prior art, and a description of the ordinary skill in the art.⁸¹ Using only this information, the obviousness specialist would remain unaware of the actual invention itself until after considering these materials, at which point the claimed invention would be analyzed. This process would avoid much of the hindsight-bias problem and could make the doctrinal determination of the obviousness of the invention a more objective analysis.⁸²

In the context of patent litigation, the use of a third-party obviousness specialist could similarly be employed as a secondary evaluation of the obviousness factual inquiry separate from the jury's determination. Either of the two basic forms of implementation for the obviousness specialist noted above could employ this concept by initially limiting the specialist's access to knowledge of the claimed invention until after a reasoned analysis of the prior art and skill in the art has been developed and provided to the court and both parties.

79. Mike Scarcella, *Federal Circuit Chief Judge Paul Michel Announces Retirement*, LAW.COM (Nov. 23, 2009), <http://www.law.com/jsp/law/LawArticleFriendly.jsp?id=1202435764527>.

80. Gregory N. Mandel, *Does Hindsight Bias Affect Obviousness Rulings?*, NAT'L L.J., Aug. 18, 2008, at S2.

81. *Id.*

82. *Id.*

B. Support in Legal Doctrine

A more challenging analysis is evaluating the doctrinal support for the use of a third-party obviousness specialist in patent-invalidity litigation, including ensuring that the use of the obviousness specialist does not erode litigants' Seventh Amendment right to trial by jury on questions of fact, enabling the valid evidentiary use of any materials or impressions developed by the obviousness specialist, and aligning the use of the obviousness specialist with existing Supreme Court and Federal Circuit precedent.⁸³ While the legal support for the use of an obviousness specialist would be nuanced based on which of the two implementation methods discussed in subpart III(A) was selected, the following discussion addresses a range of applicable legal doctrines relevant in varying measures to both methods.

The legal foundation for the use of an obviousness specialist in an official role as a special master is found in Rule 53 of the Federal Rules of Civil Procedure.⁸⁴ The grant in Rule 53 is quite broad, including the ability of the court to appoint a master to "perform duties consented to by the parties."⁸⁵ This broad grant of power is tempered by the Supreme Court's holding in *La Buy v. Howes Leather Co.*⁸⁶ that the use of special masters should be the exception as opposed to the rule.⁸⁷ Though not directly tied to Rule 53, the courts have further bolstered the belief that the use of special masters should be limited by generally refusing to recognize patent cases as representing a "complexity exception" that would remove the cases from the purview of a jury.⁸⁸ In doing so, the courts have indicated that simply attempting to inject "expertise" into the proceedings is not valid or proper under traditional jurisprudence.⁸⁹ However, other courts have held that when both parties consent to the use of a special master to develop findings in a highly technical matter, the master's findings will be generally accepted unless clearly erroneous.⁹⁰

83. See B.D. Daniel, *The Right of Trial by Jury in Patent Infringement Cases*, 28 REV. LITIG. 735, 772-74 (2009) (synthesizing a number of Supreme Court and Federal Circuit decisions that, on balance, appear to provide support for the proposition that the factual elements of invalidity defenses should remain with the jury).

84. See FED. R. CIV. P. 53 (authorizing court appointment of a special master in certain circumstances); Thomas L. Creel & Thomas McGahren, *Use of Special Masters in Patent Litigation: A Special Master's Perspective*, 26 AIPLA Q.J. 109 (1998) (commenting on the appointment and duties of special masters in patent litigation).

85. FED. R. CIV. P. 53(a)(1)(A).

86. 352 U.S. 249 (1957).

87. *Id.* at 253-54.

88. See, e.g., *SRI Int'l v. Matsushita Electric Corp. of Am.*, 775 F.2d 1107, 1127-29 (Fed. Cir. 1985) (cautioning against the creation of an exception to jury trials for patent cases, which would allow judges to "exercise their personal predilection by revising or repealing the Seventh Amendment").

89. *Id.* at 1129. The court goes on to note that in the case of patent validity, the reexamination process is the proper forum. *Id.* at 1129 n.9.

90. See, e.g., *Avco Corp. v. AT&T Co.*, 68 F.R.D. 532 (S.D. Ohio 1975). This case is instructive as it involved subject matter similar to what would be found in patent litigation. The

However, the foundation provided by Rule 53 only represents one perspective on the ability of the court and litigants to utilize a special master in a patent dispute. In order to more fully analyze the capacity to utilize an obviousness special master, we must examine the Court's *Markman* decision and the resulting impact on patent litigation.

In *Markman*, the Court declared that the activity of claim construction is solely a legal determination, thus removing any questions on claim construction from the purview of the jury.⁹¹ As a result of the *Markman* decision placing the full responsibility for claim construction on the judge as a matter of law, courts have chosen to make greater use of special masters to accurately construe the scope of patent claims. A recent study by the Federal Judicial Center noted that special masters were contemplated in roughly ten times more patent cases than in all civil and criminal cases generally.⁹² The statistics seem to reinforce the use of special masters in this role as the rate of appeal for complex patent cases that utilize special masters in comparison to those cases in which a special master was not used is nearly half (2.9% versus 5%).⁹³ Additionally, the overall reversal rate was 11.7% for those complex patent cases that were appealed in which a special master was not used, while the ultimate reversal rate for complex patent cases that utilized a special master was only 3.6%.⁹⁴ These favorable statistics provide a strong circumstantial argument for extending the use of special masters to perform similar technical interpretative tasks such as construing the boundaries of the relevant prior art necessary for the obviousness analysis.

A portion of the reasoning in the *Markman* opinion also provides textual support for the limited use of an obviousness master to evaluate portions of the *Graham* factual inquiries. While a portion of the Court's analysis focused on determining the historical precedent for which issues in patent law had traditionally been allocated to juries, the Court's opinion also expressed concern over uniformity in patent disputes. Of note, the Court stated, "Uniformity would, however, be ill served by submitting issues of document construction to juries."⁹⁵ This statement is particularly insightful considering

case involved a contract claim with regard to a highly technical dispute over a novel method of maintaining a communication network among microwave relay stations across the United States. *Id.* at 533–34. The master in this case presided over a seven-day trial that included over 200 exhibits and 1,200 pages of trial transcripts before submitting a recommendation to the court. *Id.* at 534.

91. *Markman v. Westview Instruments, Inc.*, 517 U.S. 370, 387–91 (1996).

92. JAY P. KESAN & GWENDOLYN G. BALL, FED. JUDICIAL CTR., A STUDY OF THE ROLE AND IMPACT OF SPECIAL MASTERS IN PATENT CASES 2 (2009), available at [http://www.fjc.gov/public/pdf.nsf/lookup/specmapa.pdf/\\$file/specmapa.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/specmapa.pdf/$file/specmapa.pdf) (reporting that the appointment of a special master was contemplated in 2.7% of patent cases compared to special masters being contemplated in 0.27% of civil and criminal cases overall).

93. Neil A. Smith, *Complex Patent Suits: The Use of Special Masters for Claim Construction*, LANDSLIDE, Sept.–Oct. 2009, at 36, 40 (citations omitted).

94. *Id.* at 40.

95. *Markman*, 517 U.S. at 391.

the scope of the prior art to be considered under a § 103 obviousness analysis.

By statute, the § 103 obviousness analysis requires the consideration of the materials described in § 102.⁹⁶ Section 102 in turn directs the obviousness analysis, in part, to a broad set of documents that includes patents, printed publications in the United States or foreign countries that describe the invention, and inventors' certificates in order to determine the scope of the prior art.⁹⁷ Thus, a close reading of this portion of the *Markman* opinion would seem to indicate that at least the portion of the obviousness analysis that requires ascertaining the scope of the prior art with respect to the listed types of documentation does not need to be submitted to a jury and could be assigned solely to the court, and thus in turn, to a special master, in the same manner as claim construction.

A potential challenge to the use of an obviousness master would be limiting the scope of the master's findings so as not to impinge on the factual findings reserved for the jury by the Seventh Amendment. In order to comply with existing Seventh Amendment jurisprudence, the basic role of the master could be designed in a manner similar to the responsibilities of the claim-construction special master in providing the court a set of definitions—or more properly termed, boundaries—of what the prior art could reasonably be considered to have contained at the time of the obviousness inquiry, whether that be at time of conception or reduction to practice. The contours of these boundaries could be argued by the parties and shaped in accordance with a *Markman*-like hearing, after which the arguments at trial and the resulting jury instructions would reflect the agreed upon prior art boundaries for performing the obviousness determination. The jury would still be responsible for answering the factual inquiries set forth by *Graham*, but would be assisted in its consideration of the prior art.

The alternative proposed method of employing an obviousness specialist is through the more traditional form of an expert witness. As a general matter, *Daubert*⁹⁸ provides significant leeway to the court in allowing expert testimony. In conjunction with Federal Rule of Evidence 702, and provided that the judge is satisfied that the testimony is relevant and rests on a reliable foundation, the judge may choose to admit testimony from an expert witness.⁹⁹ Additionally, the court has the option of appointing its own independent expert witness, in consultation with the parties, in order to assist in matters that the court deems appropriate.¹⁰⁰

A timely example of the specialized use of an expert witness as envisioned in this Note occurred in the recent case of *Monolithic Power*

96. 35 U.S.C. § 103(a) (2006).

97. *Id.* § 102(a)–(e).

98. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

99. *Id.* at 589–92.

100. FED. R. EVID. 706.

*Systems, Inc. v. O2 Micro International Ltd.*¹⁰¹ The Federal Circuit affirmed the court's ability to appoint an independent expert to assist in understanding complex technical issues in a manner that bears great similarity to the obviousness specialist proposed in this Note.¹⁰² While the power of courts to independently appoint an expert to assist in evaluating the technical issues of a case is well established, the power is very rarely exercised,¹⁰³ making the Federal Circuit's recent review of the practice an interesting, and perhaps important, precedent.

The *Monolithic* case involved DC/AC-converter circuitry used to power laptop computers.¹⁰⁴ Early in the case, the court expressed frustration over the technical complexities of the case and openly questioned whether the jury would be able to understand any of the technical issues.¹⁰⁵ Despite an acknowledgment that the use of a court-appointed expert may influence the jury, the court proposed the appointment of an independent expert to testify with respect to the "electrical engineering aspects" of the case.¹⁰⁶ Over the objections of one of the parties, the court and the litigants jointly selected the expert, Dr. Enrico Santi.¹⁰⁷ Dr. Santi spent over seventy-five hours analyzing the technical issues.¹⁰⁸ He produced a written report of his findings, was available for depositions, and met with counsel from both sides to prepare a technology tutorial for the court.¹⁰⁹ Dr. Santi opined and testified on both infringement and invalidity issues, including obviousness and statutory bars.¹¹⁰ Both parties also presented their own expert witnesses, and the jurors were provided instructions that they were not to afford the opinion of the independently appointed expert any greater deference than the litigant's experts.¹¹¹

101. 558 F.3d 1341 (Fed. Cir. 2009).

102. *See id.* at 1346–48 (holding that a court is able to appoint an independent expert if the court follows the requirements set out in Rule 702).

103. *See id.* at 1346–47 (listing a number of authorities that note the rare nature of courts exercising the power to appoint independent experts). The Federal Circuit's review cites a number of sources that describe the power of the court to appoint an independent expert, quoting at one point that the power is "virtually unquestioned." *Id.* at 1347 (quoting FED. R. EVID. 706 advisory committee's note). While the opinion expresses some concerns with the practice, it notes that local circuit court precedent governs the decision and that, provided there was no abuse of discretion by the court, the use of the expert was proper. *Id.* at 1348.

104. *Id.* at 1343–44.

105. *Id.* at 1345 ("I find this extremely difficult to understand. And the notion that a jury is going to understand it, to me, is foolishness. You can talk for months and the jury isn't really going to understand this in the sense of being able to make a reasoned, rational decision about it.") (quoting Hearing Transcript at 35:3–8, *Monolithic Power Sys. v. O2 Micro Int'l Ltd.*, 2007 WL 3231709 (N.D. Cal. 2006) (Nos. C 04-2000 CW, C 06-2929 CW)).

106. *Id.* (internal quotation omitted).

107. *Id.* at 1345, 1347.

108. *Id.* at 1347.

109. *Id.*

110. *Id.* at 1345–46, 1348.

111. *Id.* at 1347–48.

On appeal, the Federal Circuit considered the objecting party's arguments that the appointment of the independent expert was a violation of the litigant's Seventh Amendment right to a jury trial.¹¹² The Federal Circuit maintained that, provided the court followed the procedural rules of Rule 706, the use of the independent expert to opine on matters including obviousness and infringement was considered acceptable.¹¹³

IV. Implementation

The concept of an obviousness specialist in invalidity litigation could be implemented in two different forms. The first is to implement an obviousness master in a manner similar to that employed for a traditionally titled special master in claim-construction and interpretation hearings. The second is to develop the obviousness-master role via a specialized incarnation of an expert witness. The specialized nature of the obviousness expert witness would stem from the use of a double-blind selection method to choose the individual. Each form of implementation is discussed below.

A. *Obviousness Special Master*

While the authority to appoint an obviousness special master ultimately lies with the court, the initial discussion on whether to employ a special master should include each party to the dispute. Federal Rule of Civil Procedure 53(b) requires that when considering the use of a special master, each party must be allowed to suggest candidates for appointment.¹¹⁴

In making the determination of whether to appoint a special master to develop findings on the proper boundaries of relevant prior art and to characterize the general abilities of a person with skill in the art at the relevant point in time, the court and the parties should consider several key issues. First, how significant is the probable gap between the art in question and a standard jury's level of understanding? This question would be particularly useful for patent litigation occurring in docket locations with a large amount of high-tech industry in the surrounding area or in locations such as the Eastern District of Texas where the volume of patent litigation may subtly increase the relative sophistication of the juror pool. The more complex the field of innovation is in comparison to the expected sophistication of the expected jury pool, the higher the likelihood that an obviousness special master would be beneficial.

The second question the litigants and the court should consider is the timeline of the inventive activities in question. As noted above, the issue of hindsight bias can make obviousness determinations challenging and the further in the past the inventive activity occurred, the more challenging the

112. *Id.* at 1347.

113. *Id.* at 1347-48.

114. FED. R. CIV. P. 53(b).

hindsight-bias issue becomes. Thus, longer invention timelines would be better candidates for using an obviousness master.

Third, the parties should consider the perceived breadth of the subject matter that they believe would constitute the relevant field or fields of prior art. If both parties hold a similar view on the breadth of the prior art field, there may be less of a need for an obviousness expert due to the ability of the parties to argue the finer contours of the prior art through the trial. However, in cases where the parties have significantly divergent views of the relevant field or fields of prior art, it would be beneficial for both the court and the jury to obtain an informed, independent, and unbiased view of the prior art boundaries from a source that has practical familiarity with the technical areas. Simply allowing both parties to present their widely divergent views of broad-versus-narrow fields of art is likely to result in the factually insufficient or legally inconsistent results discussed in Part II.¹¹⁵

Fourth, the court and the parties should consider the actual and targeted timelines for the case. Utilizing special masters in patent litigation for claim construction can save the litigants both time and money by allowing portions of the case to move forward more quickly.¹¹⁶ Similarly, the use of a special master could save the parties a nontrivial amount of time in preparing and trying the invalidity portion of the case based on their ability to argue the obviousness factors with reference to a known set of boundaries developed by the special master. The parties also need to evaluate the proper point in the litigation to engage a special master. The earlier the obviousness special master is used, the earlier the parties gain valuable perspective on the strength of their obviousness positions.¹¹⁷

Once the parties do elect to employ an obviousness special master, the court and the parties must jointly agree on the selection of the individual. Once agreement is reached, the court, with the support and agreement of the parties, must develop specific requirements for the analysis that the obviousness master is to provide. These requirements should contain a format and a structure of the desired analysis that do not usurp the power of the jury to answer the *Graham* factual queries. The requirements should, however, allow the obviousness special master to develop a set of outer boundaries that would reasonably encompass the prior art at the time of the invention in question. Additionally, the obviousness master should be able to develop a set of general capabilities of a person with skill in the relevant art at the appropriate time and a description of the common tools and techniques that a person skilled in the art would have had access to.

In order for the obviousness master to develop a report of this information, the court and the parties must agree on an appropriate set of data

115. See *supra* notes 54–70 and accompanying text.

116. Smith, *supra* note 93, at 39.

117. See *id.* (discussing the implications of how the timing of claim construction may impact the parties' election to use a special master).

to be provided to the obviousness master. Keeping in mind the challenge of hindsight bias, the parties may or may not decide to provide the actual patent or patents at issue in the case. The level of detail in the materials provided to the obviousness master on the problem domain would likely be related to the gap between the two parties' views on the relative breadth of the prior art. If the difference in the parties' views on the scope of the prior art is small, the parties could provide the obviousness master with more detailed materials in order to allow the report to address finer distinctions in where the prior art boundaries should be set. If, however, the parties' views on the relevant scope of the prior art are widely divergent, the parties may need to provide more general information in order to allow the obviousness master more leeway in fashioning the boundaries.

Once the obviousness master provides his report, the court must review the report to ensure that it complies with the requirements and does not impinge on the parties' right to have the obviousness factual inquiries tried by the jury. Provided that the report is acceptable, all parties would receive a copy and would have the opportunity to lodge any objections to the report via a traditional hearing proceeding. As long as the court finds that the report is not unduly prejudicial, the findings of the obviousness master would be explained to the jury as part of its instructions in analyzing the obviousness question. The parties would be free to argue the conclusions of the special master to the jury and the jury would not be limited to considering only those materials determined by the obviousness master to be within the probable scope of the prior art. Any appeals on the question of obviousness could involve a review of both the procedural use of the obviousness master and the substantive content and implications of the master's conclusions.

B. Obviousness Expert Witness via Special Selection Process

Alternatively, and building on the approach used in the *Monolithic* case, the court could initiate a dialog early in the case-management process to discuss the potential use of an independent obviousness expert witness for cases in which the technical innovation may be particularly challenging. While it is likely that both parties will still wish to employ their own expert witnesses at trial, an *ex ante* opportunity to potentially receive an independent assessment of obviousness in their favor may be as significant a development to their potential success in trial as would be a favorable claim construction.

However, the simple appointment by the court of an independent expert, such as in *Monolithic*, may result in the jury giving too much deference to that witness's testimony based on a perceived imprimatur by the court.¹¹⁸ Based on the judge's comments that appear in the *Monolithic* opinion, the

118. Robertson, *supra* note 74, at 198.

judge was aware of this concern but felt that proper jury instructions would minimize the risk.¹¹⁹

One potential issue with the expert used in *Monolithic* is that the expert still had full knowledge of who the parties to the litigation were and what their respective interests were in the litigation. In order to gain the benefit of a truly unbiased but technically skilled obviousness expert witness, the court should draw upon the time-tested practice of blind selection as is used in social and scientific research.¹²⁰ Specifically, the use of a double-blind selection process—referred to as the “gold standard” in scientific research¹²¹—to select the obviousness expert would provide the court and the litigants the most objective determination of obviousness possible. The double-blind selection process would ensure that no affiliation or compensation biases exist in either direction between the obviousness expert and the litigants,¹²² unlike the party-aware expert used in *Monolithic*.

In order to make the double-blind selection of an obviousness expert operational, there would be a need to establish an intermediary that could play two roles: (1) maintainer and evaluator of the available experts, and (2) communication liaison between the parties and the expert once an expert has been engaged. The biggest challenge would be the identification of an appropriate intermediary that could maintain a roster of available experts in the various fields. The ideal candidates to serve as obviousness experts would be individuals who could fulfill the legal requirements to qualify as an expert witness and would have sufficient familiarity with the historical developments in the relevant technical field such that the individual could provide a correct and comprehensive evaluation of the appropriate boundaries for the content of the prior art at the time of invention and a set of characteristics that would best characterize the skill of the relevant individual in the art. Note that this description of a potential obviousness expert does not require the individual to have practiced in the art at the time in question, but merely requires a fluent understanding of the key development milestones and the standard process of innovation that was prevalent in the art.

While a judge could—and per Rule 702, likely would—be required to play the role of communication liaison between the expert and the court, the development of a representative list of candidates is likely beyond the capabilities and time constraints of the court. Instead, this matchmaker role

119. *Monolithic Power Sys., Inc. v. O2 Micro Int'l Ltd.*, 558 F.3d 1341, 1345–46 (Fed. Cir. 2009).

120. See Robertson, *supra* note 74, at 202–05 (discussing the concept of blinding in justice and science).

121. *Id.* (citation omitted).

122. *Id.* at 208. Robertson discusses the complete background for the double-blind selection process and notes that even an expert selected in a single-blind scenario may be susceptible to changing his opinion once he learns who the parties are. *Id.* at 209. A double-blind selection in which neither party knows the identity of the other is the only method that provides a near-ideal objective determination. *Id.* at 206.

would likely need to be coordinated between the judge and various professional-registration organizations¹²³ in order to ensure that the candidates were both willing and suitably qualified to be considered. Once the court and the parties had reached an agreement on the use of an independent expert, the parties would jointly specify a list of qualification criteria for the individual and the judge would submit these criteria to the appropriate licensing organization or society. Based on the available pool, the organization would return a list of possible candidates that satisfy some threshold quantity of the specified criteria. The parties would have the opportunity to review the basic anonymized academic and professional credentials of the candidates and, in consultation with the judge, would select one of the candidates, or if no agreement could be reached, either request a different pool of candidates or allow the court to revert back to the traditional process of selecting an independent expert. If a candidate is chosen, all costs for the use of the expert would be split evenly between the parties.

Assuming that an expert is blindly selected, the judge, working with the parties, would assemble a generalized package of information regarding the area or areas that the court and the parties believe will need to be considered in evaluating the obviousness of the patent. Information about the parties and the specific patent at issue would not be provided to the obviousness expert. The judge would provide instructions to the expert on the type and form of the analysis required. Any inquiries for clarification or additional data would be communicated to the judge who would consult with the parties to determine the proper response.

Upon completion of the analysis by the expert, the expert's report would be provided to the court and to both parties. At that point, both parties would have the option to hire the expert as a traditional expert witness to be used at trial. It is likely that the expert's report would be more favorable to one of the party's positions than the other and thus the expert would be more likely to be hired by that party. If a party elects to hire the expert, the jury would be informed of the independent nature of that expert's testimony while being cautioned that the independence of the expert's opinion should not be construed as determinative and should be considered in light of any other expert testimony in the trial. If neither party believes that the expert's analysis would be beneficial, it is possible that neither party will choose to hire the expert.

However, under Rule 706, the court may still retain the expert if the judge believes that the expert's analysis on the boundaries of the skill in the art and the scope of the prior art will be beneficial in a complex case.¹²⁴ This

123. These would likely include professional-licensing agencies for experts such as physicians, engineers, and architects and professional societies for many of the other scientific fields such as physics, chemistry, and biology.

124. See FED. R. EVID. 706(a) ("The court . . . may appoint expert witnesses of its own selection.").

may occur when the expert's opinion appears to favor neither party but does provide useful insights into the relevant field of art. If the court believes that the expert can be of particular assistance at the trial and that the use of the expert would be nonprejudicial to either party, the court could retain and utilize the independent obviousness expert in a manner consistent with the use of the expert in *Monolithic*.

V. Benefits and Challenges

Depending on which one of the two implementation methods is chosen, there are a variety of arguments that can be raised both in support of and against the use of a third-party obviousness specialist or master.

At a high level, a general argument can be made that the use of either an obviousness special master or a specially selected obviousness expert witness would work in favor of producing more consistent, uniform, and technically correct results to obviousness determinations. Because the obviousness specialist would be expected to possess deep knowledge of the processes and historical developments in the relevant technical area, the use of the independent specialist to define the general boundaries of the field of innovation would greatly assist the court and the jury in placing the parties' evidence and arguments within a well-defined and objective frame of reference.

Opponents could argue that this general framing of the scope of the relevant field of art by an independent obviousness specialist, while helpful, may end up being too helpful if the court or the jury assigns too much weight to the specialist's opinions. The use of the specialist could thus become outcome determinative for the litigation, subverting the standard procedures and protections embodied in civil litigation practice.

While this is a valid concern, the key is to focus on the scope of the obviousness specialist's activities to ensure that they do not tread too closely to the ultimate factual or legal questions that are assigned to the jury and the court, respectively. Both the court and the parties would have multiple procedural opportunities to shape the scope and depth of the specialist's work. The parties would also still be able to present arguments to the jury in an attempt to cast doubt on the work of the obviousness specialist. In fact, the specialist's work may actually assist the parties in the long run because any obviousness determination made using the specialist's input would intrinsically provide a reviewing appellate court a more robust record upon which to consider the appeal.

Additionally, the argument that the obviousness specialist's opinion might be outcome determinative is no more or less effective in this scenario than in the claim-construction scenario. While it can be conceded that the *Markman* decision may have influenced the litigation of patent cases due to the resulting claim-construction outcomes having a strong potential to influence the overall outcome of the litigation, obtaining third-party assistance in evaluating the question of obviousness should be no more influential. In

fact, due to the circumspect use of the obviousness expert witness or special master to opine only on the general boundaries of the skill and content in the prior art, the obviousness specialist's impact is at least two steps removed from the ultimate obviousness determination due to the requirement that both the court and the jury must still fulfill their duties in supplying the legal and factual responses to the obviousness question. If the obviousness specialist were asked to opine specifically on the final question of obviousness, such as what was done in *Monolithic*, the parties would have a much stronger argument on appeal that the use of the obviousness specialist was unduly prejudicial.

From a more practical perspective, the use of an obviousness specialist would remove some of the concern over the indefiniteness of the obviousness standard, particularly with respect to the jury. While a jury working alone may be unclear about how best to weigh the evidence of possible obviousness at the time of conception versus at the time of reduction to practice, the general perspective from the obviousness specialist would provide a backdrop against which the jury could better understand the process of innovation in the industry and thus better understand what might be produced by an ordinary creative person skilled in the art during the relevant time period.

The use of an obviousness specialist could also limit the impact of hindsight bias by providing a reasoned, independent perspective on the scope of innovation in the industry, which would remove some of the "a-ha!" moments from the court and jury as they begin to grasp the technology during the course of the trial. Alternatively, it could be argued that the obviousness specialist may actually increase the hindsight-bias problem due to the specialist's knowledge in the area that he is being asked to review. Such familiarity in the art may cause the specialist's opinions to be skewed toward a finding of obviousness. Again, however, the focus for the obviousness specialist is on a depiction of the overall state of the art at the time in question, not on reviewing the specific invention at issue. Even if the specialist is permitted to opine on the obviousness of the invention in question, there is a policy argument that a bias toward a finding of obviousness would be consistent with the constitutional and statutory goals of limiting patent protection to truly innovative advances in the art and that the review by an individual truly skilled in the art would better serve this goal than unskilled reviewers attempting to make the determination.

The use of an independent obviousness special master or expert witness also provides an additional truth factor to patent litigation.¹²⁵ By including an independent opinion in some portion of the obviousness determination, the court and the jury gain a proverbial guidepost that can be used to evaluate and navigate the "hired-gun" expert witnesses and the theories advanced by the parties. Provided that the court is diligent about controlling the

125. See Robertson, *supra* note 74, at 181-84 (discussing "truth-signals" with respect to the blind procedure).

placement of this philosophical guidepost, the additional objective-decision criteria should produce more uniform and logically consistent results than a jury that “when . . . flummoxed by the substance of highly technical testimony . . . evaluat[es] the credibility—and especially the impartiality—of the experts delivering testimony.”¹²⁶

Finally, the use of an independent obviousness specialist’s analysis would provide a practical foundation on which the court could develop effective special-verdict and interrogatory questions to guide the jury’s evaluation of the *Graham* factual inquiries. In the absence of consistent model forms for obviousness special verdicts, this may be the next-best option in developing forms that would help prevent legally inconsistent jury verdicts such as in *Callaway Golf Co.* More detailed, focused special-verdict forms would also assist the court in complying with the requirement in *KSR* that the obviousness analysis be explicitly made.

VI. Conclusion

The determination of obviousness in patent litigation remains fraught with challenges for both the court and the jury. In attempting to work with an indeterminate standard, lay persons, who are often not technically savvy, are asked to transport themselves to another place and another time in which they are assumed to possess both the technical knowledge and creativity necessary to determine if the claims of one particular invention would have been obvious to their hypothetical selves. While commentators and courts have recognized the numerous issues with expecting such a system to produce consistent and predictable results, there have been few actions taken to mitigate the challenge.

This Note has briefly described the use of an independent third-party obviousness specialist during invalidity litigation as one possible approach to minimizing this challenge. By providing a baseline perspective on the process and history of innovation in the technical art, the independent third-party obviousness specialist allows the parties, the court, and perhaps most importantly, the jury to focus on the primary task of navigating the defined waters of the invention and the most-relevant art without attempting to determine the location and validity of the invention in the larger, unmapped sea of innovation.

Gregory R. Baden

126. *Id.* at 232.

The Effect of the Parliamentary Voting System Act on the British Constitution *

Perhaps presciently, in 2009 Vernon Bogdanor noted, “A hung parliament, were it to occur, would almost certainly put the reform of the first-past-the-post electoral system at the centre of the political agenda.”¹ Subsequently, the 2010 British election for the House of Commons returned a hung Parliament with no party able to command an outright majority in the 650-seat House of Commons.² The election results gave the Conservative Party a plurality of 307 seats, with the Labour Party receiving 258 and the Liberal Democrats receiving 57.³ Eventually the Conservatives and Liberal Democrats formed a coalition government with David Cameron of the Conservatives as Prime Minister and Nicholas Clegg of the Liberal Democrats as Deputy Prime Minister.⁴ The formal agreement between the parties set out numerous legislative goals for the coalition, perhaps most importantly in the area of political and constitutional reform. In particular, as Bogdanor predicted, the coalition government pledged to “bring forward a Referendum Bill on electoral reform, which includes provision for the introduction of the Alternative Vote . . . as well as for the creation of fewer and more equal sized [parliamentary] constituencies.”⁵

The coalition fulfilled this promise by passing the Parliamentary Voting System and Constituencies Act (PVS Act).⁶ The House of Commons passed the PVS Act on November 2, 2010.⁷ The House of Lords subsequently passed the PVS Act on February 14, 2011, and the PVS Act received the Royal Assent on February 16, 2011.⁸ In short, the PVS Act (1) provides for a referendum—to be held between May 5, 2011 and October 31, 2011—on

* For my grandfather of blessed memory, David D. Becker. I am grateful to the staff and editors of the Texas Law Review, especially Sarah Hunger, Serine Consolino, Jamie France, Christopher Granaghan, and Anthony Arguijo, for their efforts in bringing this Note to publication. I also wish to thank Professor Emily Kadens and Professor Steve Bickerstaff for their helpful research suggestions. Finally, I am grateful to my family for their constant love and support.

1. VERNON BOGDANOR, *THE NEW BRITISH CONSTITUTION* 130 (2009).

2. *Election 2010: First Hung Parliament in UK for Decades*, BBC NEWS, <http://news.bbc.co.uk/2/hi/8667071.stm>.

3. *Election 2010: National Results*, BBC NEWS, <http://news.bbc.co.uk/2/shared/election2010/results/>.

4. See *THE COALITION: OUR PROGRAMME FOR GOVERNMENT* 7–8 (2010), available at http://cabinetoffice.gov.uk/sites/default/files/resources/coalition_programme_for_government.pdf (introducing the plan for a coalition government).

5. *Id.* at 27.

6. Parliamentary Voting System and Constituencies Act, 2011, c. 1 (U.K.).

7. 2 Nov. 2010, PARL. DEB., H.C. (2010) 888–92 (U.K.).

8. *Bill Stages—Parliamentary Voting System and Constituencies Bill 2010–11*, PARLIAMENT.UK, <http://services.parliament.uk/bills/201011/parliamentaryvotingsystemandconstituencies/stages.html>.

switching from the current first-past-the-post electoral system to the alternative vote,⁹ and (2) reduces the size of Parliament to 600 members, from the current 650, and redraws the electoral constituencies accordingly.¹⁰ Part I of this Note examines the effect of the PVS Act on the British Constitution by virtue of the referendum. Part II of this Note examines the constitutional effects of switching the electoral system to the alternative vote.¹¹ As the choice of an electoral system fundamentally reflects the constitutional values of society, it is the alternative-vote provisions, rather than the referendum included in the PVS Act, that represent a major reform of the British Constitution and a threat to parliamentary sovereignty.

I. The Effect of the PVS Act's Referendum on the British Constitution

A. *The British Constitution and the Sovereignty of Parliament*

In discussing the British Constitution, it is first important to clarify the sense in which Britain has a constitution. Many commentators have stated, for example, that the British Constitution is unlike other constitutions in that it is unwritten.¹² Upon further reflection, however, this statement is not entirely correct because it conceives of a constitution in a particular manner, namely one like the Constitution of the United States. To be sure, the British Constitution is quite unlike the U.S. Constitution in that the British Constitution is not “a single instrument promulgated at a particular point in time and adopted by some generally agreed authorisation procedure under

9. The alternative vote, also known as instant-runoff voting, is a system where voters can vote for multiple candidates, ranking those candidates in their order of preference. *Alternative Vote*, ELECTORAL REFORM SOC'Y, <http://www.electoral-reform.org.uk/article.php?id=55>. For a detailed description of the alternative vote, see *infra* section II(B)(2).

10. Parliamentary Voting System and Constituencies Act, 2011, c. 1, §§ 1, 8–9, 11 (U.K.). The provisions of the PVS Act that propose to redraw the constituencies represent a major change for the current House of Commons. These provisions reflect an emphasis on the individual right to vote insofar as the rules for drawing constituencies change to emphasize mathematical equality of districts in apportionment rather than geographic ties. Compare *id.* § 11 (indicating the new rules for determining parliamentary constituencies), with Parliamentary Constituencies Act, 1986, c. 56, §§ 1–6, sch. 2 (U.K.) (specifying the 1986 rules for redistribution of parliamentary seats). The constitutional effects of changing the method of drawing constituencies, however, are unclear and are beyond the scope of this Note.

11. In its analysis, this Note assumes that the referendum will return a result in favor of switching to the alternative vote. Even if the referendum returns a result against the alternative vote, however, this Note's analysis remains relevant because reform of the electoral system will likely remain on the British political agenda. In particular, the Liberal Democrats, as a significant third party, will continue to pressure the major parties to pass electoral reform. See BOGDANOR, *supra* note 1, at 131 (noting that the main aim of third parties in cases of hung parliaments has been to secure electoral reform).

12. See, e.g., ANTHONY KING, *THE BRITISH CONSTITUTION* 5 (2007) (“It is often said that, whereas most other liberal democracies have written constitutions, the British constitution is unwritten.”).

the title ‘constitution’ (or equivalent rubric such as ‘basic law’).¹³ Neither is the British Constitution “a list of statutes or instruments that have an entrenched status and can be amended or repealed only by a special procedure.”¹⁴ Together, these definitions conceive of a constitution as a single, special document that sets out the basic form of government. It is in exactly this sense that Britain does not have a written constitution.

Rather than a single document, the British Constitution is “the combination of legal and non-legal (or conventional) rules that currently provide the framework of government and regulate the behaviour of the major political actors.”¹⁵ This definition of a constitution is more general than the sense of a constitution as a single document, but in this general sense, every country has a constitution.¹⁶ While the British Constitution, like the U.S. Constitution, sets out the framework of government, the British Constitution is made up of the entire corpus of British law, in addition to other sources.¹⁷ Thus, the British Constitution is unwritten in the sense that it is uncodified.¹⁸

Of the many parts of the British Constitution, perhaps the most fundamental is the doctrine of the sovereignty of Parliament. Parliament is composed of “the king’s majesty . . . and the three estates of the realm; the lords spiritual, the lords temporal (who sit, together with the king, in one house) and the commons, who sit by themselves in another.”¹⁹ In more modern terms, these would be the Monarch, the House of Lords, and the House of Commons.²⁰ For Blackstone, “the power of making laws constitutes the supreme authority,”²¹ and the British Constitution grants that legislative power to Parliament.²²

It is important to emphasize here that the sovereignty of Parliament applies in two distinct ways. The first is, quite simply, that “Parliament can

13. Geoffrey Marshall, *The Constitution: Its Theory and Interpretation*, in THE BRITISH CONSTITUTION IN THE TWENTIETH CENTURY 29, 31 (Vernon Bogdanor ed., 2003).

14. *Id.*

15. *Id.*

16. *Id.*

17. See WILLIAM PALEY, *Elements of Political Knowledge* (“In England, the system of public jurisprudence is made up of acts of parliament, of decisions of courts of law, and of immemorial usages; consequently, these are the principles of which the English constitution itself consists . . . and by which every constitutional doubt and question can alone be decided.”), in THE WORKS OF WILLIAM PALEY, D.D.: ARCHDEACON OF CARLISLE 122 (1836).

18. KING, *supra* note 12, at 5.

19. 1 WILLIAM BLACKSTONE, COMMENTARIES *153.

20. See ROGER D. CONGLETON, PERFECTING PARLIAMENT: CONSTITUTIONAL REFORM, LIBERALISM, AND THE RISE OF WESTERN DEMOCRACY 301–03 (2011) (describing the King’s role in Parliament, in addition to defining “the lords spiritual” and “the lords temporal” and using “commons” synonymously with “House of Commons”).

21. 1 BLACKSTONE, *supra* note 19, at *52.

22. See *id.* at *160–61 (describing the legislative power of Parliament and the sovereignty of Parliament).

pass any law it likes.”²³ As Blackstone notes, quoting Sir Edward Coke, the power of Parliament “is so transcendent and absolute, that it cannot be confined, either for causes or persons, within any bounds.”²⁴ This is quite unlike the situation in the United States where there are affirmatively certain laws that Congress could not pass, such as ex post facto laws or bills of attainder.²⁵ Simply, this principle means that Parliament could pass or repeal any law, though it also implies that Parliament can rewrite the British Constitution at its whim²⁶ and, perhaps most importantly, that an earlier session of Parliament cannot bind a later session of Parliament.²⁷

The second way that the sovereignty of Parliament applies is that Parliament’s legal authority is supreme. Here, Blackstone declares that “what the parliament doth, no authority upon earth can undo.”²⁸ In effect, Parliament not only possesses legislative authority, but it is the sole source of that legislative authority under the British Constitution. Blackstone’s declaration posits a situation where no person or institution can question the legal authority of Parliament to take some action: the corollary of “parliamentary sovereignty is that there can be no formal legal or constitutional checks upon the power of government.”²⁹ Unlike the situation of the United States, there is no arbiter such as the Supreme Court that can declare an act of Parliament unconstitutional.³⁰ Indeed, to the extent that the British Constitution is merely the sum of the acts of Parliament, any act of Parliament must by definition be constitutional. This is, of course, not to say that any law Parliament passes is necessarily just, but merely to say that Parliament has the constitutional power to pass any law it chooses.³¹ In short, Parliament is sovereign in that it has complete legislative authority and that none has the power to question that authority.

The doctrine of Parliament’s sovereignty over legal matters, however, should not be confused with the notion of who holds political sovereignty

23. J. HARVEY & L. BATHER, *THE BRITISH CONSTITUTION* 10 (4th ed. 1977).

24. 1 BLACKSTONE, *supra* note 19, at *160.

25. U.S. CONST. art. I, § 9, cl. 3.

26. See A.V. DICEY, *INTRODUCTION TO THE STUDY OF THE LAW OF THE CONSTITUTION* 89 (E.C.S. Wade ed., 10th ed. 1959) (1885) (“There is under the English constitution no marked or clear distinction between laws which are not fundamental or constitutional and laws which are fundamental or constitutional.”).

27. See 1 BLACKSTONE, *supra* note 19, at *186 (noting that only Parliament can alter a previous act of Parliament).

28. *Id.* at *161.

29. BOGDANOR, *supra* note 1, at 15.

30. See KING, *supra* note 12, at 9 (“One consequence of the fact that Britain does not have a Constitution and that no distinction is made in British law between specifically constitutional matters and others is that the word ‘unconstitutional’ has no precise meaning in the U.K., if indeed it has any meaning at all.”).

31. Parliament could, for example, extend the duration between elections of the House of Commons or attain someone of treason. See, e.g., DICEY, *supra* note 26, at 46–48 (describing the constitutionality of Parliament’s passage of the Septennial Act and the power of Parliament to interfere with public and private rights).

under the British Constitution. As Dicey notes, “[t]he external limit to the real power of a sovereign consists in the possibility or certainty that his subjects, or a large number of them, will disobey or resist his laws.”³² Parliament, too, is subject to this fundamental limitation just as is any government. As a practical matter Parliament’s sovereignty is limited because “widespread resistance would result from legislation which, though legally valid, is in fact beyond the stretch of Parliamentary power.”³³ A perfect example here would be a bill by Parliament to suspend elections indefinitely—while there is no question that Parliament legally could pass such a law, the popular resistance that would result from such an act effectively ensures that Parliament would never attempt to do so. Further, the doctrine of parliamentary sovereignty makes no claim as to “the relative power of Parliament and the executive,” which includes the Prime Minister and the cabinet.³⁴ In modern British government, the executive, which is still an element of Parliament, tends to dominate the rest of Parliament in the sense that it controls what legislation is brought forward and enacted.³⁵ In both of these situations, it is the British electorate, through its capacity to resist unjust laws and through voting for the members of Parliament who ultimately choose the Prime Minister, that is politically sovereign. Nevertheless, it is not fundamentally inconsistent to maintain both that there are practical political limits to Parliament’s powers and that its authority over law and legislative matters remains supreme.

This Note, of necessity, takes a simplistic view of the doctrine of parliamentary sovereignty and its centrality to the British Constitution. To be sure, the modern reforms of the British Constitution, especially those concerning the Human Rights Act and Britain’s relationship to the European Union (EU), have led some commentators to question whether Parliament is actually still sovereign over legislative matters in Britain and whether the principle of parliamentary sovereignty still sits at the heart of the British Constitution.³⁶ These considerations are far beyond the scope of this Note. Going forward, this Note assumes that, prior to the enactment of the PVS Act, Parliament is the supreme legislative authority in the United Kingdom and that parliamentary sovereignty is the core principle of the British Constitution.

B. Bruce Ackerman’s Commentary on the PVS Act

One key element of the sovereignty of Parliament is that a law cannot be entrenched in such a way as would make it impossible for Parliament to

32. *Id.* at 76–77.

33. *Id.* at 79.

34. BOGDANOR, *supra* note 1, at 15.

35. *See id.* (noting the decline of Parliament vis-à-vis the government).

36. *See, e.g., id.* at 278, 282–84 (describing the European Communities Act and Human Rights Act and evaluating their effects on the British Constitution).

touch, alter, or repeal that law. An entrenched law would vitiate parliamentary sovereignty in the first sense as Harvey and Bather describe it. The question here is whether a law that is practically entrenched, in the sense that Parliament could overturn it but would likely never do so, would vitiate the parliamentary sovereignty in the same way as a law that Parliament could not legally change.

Here, Dicey argues “that the arrangements of the constitution are now such as to ensure that the will of the electors shall by regular and constitutional means always in the end assert itself as the predominant influence in the country. But this is a political, not a legal fact.”³⁷ So long as it is Parliament that is the sole source of the law, and so long as it is Parliament’s interpretation of the law that controls, Parliament remains legally supreme. This is so even if it is the people that are ultimately politically sovereign.

In a recent article in the *London Review of Books*, Bruce Ackerman, commenting on the PVS Act, argues the contrary: “As the new power centres increasingly assume a life of their own, statutory assertions of ultimate control by Westminster begin to wear thin.”³⁸ Ackerman argues that the PVS Act provides an example of “an alternative source of constitutional legitimacy: the referendum.”³⁹ As a historical precedent to the effect of the PVS Act, Ackerman cites the 1975 referendum on British membership in the European Economic Community (EEC).⁴⁰ For some commentators, the “yes” result in that referendum—that Britain should be a member of the EEC—effectively entrenched, for political purposes, the issue of British membership in the EEC.⁴¹ Ackerman’s argument, at the simplest level, is that the PVS Act would accomplish the same result. For Ackerman, the PVS Act strikes a blow against parliamentary sovereignty and in favor of popular sovereignty by holding a referendum on the issue of electoral reform.

On an intuitive level, Ackerman’s argument appears correct. Having, for example, granted devolved powers to the Welsh Assembly and the Scottish Parliament by referenda,⁴² the Parliament cannot take them back (without the consent of the Welsh and Scottish citizens, of course). Though Dicey would remind us that we are confusing political sovereignty with legal sovereignty, the solution to the question cannot be that simple. To argue that Parliament has the legal authority to take back the devolved powers of the Welsh Assembly and the Scottish Parliament but that it would never dare to

37. DICEY, *supra* note 26, at 73.

38. Bruce Ackerman, *At the Crossroads: Bruce Ackerman on the Surrender of Parliamentary Sovereignty*, LONDON REV. BOOKS, Sept. 9, 2010, at 32, 32.

39. *Id.*

40. *Id.*

41. SELECT COMMITTEE ON THE CONSTITUTION, REFERENDUMS IN THE UNITED KINGDOM, 2009-10, H.L. 99, ¶ 19 (U.K.).

42. *Id.* at 9–10 box 1.

do so is tantamount to stating that Parliament lacks that power. As a practical matter, the referenda on the devolved assemblies entrenches them beyond the reach of Parliament, and it is likely that the referendum on the alternative vote included in the PVS Act would do the same should the referendum succeed. Practically, this vitiates the sovereignty of Parliament in the sense that Parliament cannot pass any law it likes (at least regarding an electoral system), though not in the sense that Parliament's legal authority is unquestioned.

A closer examination of Ackerman's argument, however, reveals that the sovereignty of Parliament is not so clearly vitiated as would appear from our intuitive analysis. In his argument, Ackerman poses a hypothetical situation to test the "traditionalist legal fiction . . . that Parliament [may] merely us[e] the referendum as an 'advisory' device while maintaining its ultimate authority to make the final decision."⁴³ Ackerman describes a situation in which the Conservative Party takes power and pledges, among other things, to remove Britain from the EU,⁴⁴ and a constitutional crisis ensues on account of the popular support for British EU membership from the referendum on that issue in 1975.⁴⁵ Ackerman posits that another referendum may be necessary to resolve the issue as parliamentary power is unlikely to be able to do so alone.⁴⁶ To be sure, as was the case above with the intuitive analysis, the use of the referendum has placed the subject of EU membership beyond the scope of parliamentary power (absent a second referendum on the subject). This hypothetical does not show, however, as Ackerman argues, that popular sovereignty is an alternative source of legislative authority under the British Constitution; it merely shows that it is possible, for practical purposes, to entrench a law beyond the ambit of Parliament's legislative powers.

Ackerman continues with a second hypothetical that seizes on the common wisdom that the alternative vote will increase the representation of third parties, in this instance the Liberal Democrats and the nationalist parties, at the expense of the major parties, Labour and the Conservatives.⁴⁷ Such a situation makes hung parliaments, and therefore coalition governments, more likely.⁴⁸ Ackerman argues that, in addition, the alternative vote increases the likelihood of future referenda in such a way as

43. Ackerman, *supra* note 38, at 32.

44. The European Union is the successor to the European Economic Community. Michael Z. Wise, *A Capital of Europe? Brussels Is Primping*, N.Y. TIMES, Mar. 2, 2002, at B9.

45. Ackerman, *supra* note 38, at 32.

46. *See id.* (hypothesizing that if the majority party in Parliament were to attempt to remove the United Kingdom from the EU, some would insist that another referendum be held in light of the popular support of membership expressed in the 1975 referendum).

47. *Id.*

48. *See id.* (imagining a situation where a minor party wins enough votes to deprive the Conservative and Labour parties of the majorities that they have had for many years).

to erode parliamentary sovereignty over time.⁴⁹ Here, Ackerman posits a situation in which an election has returned a hung parliament.⁵⁰ One of the third parties, in his example the Greens, takes advantage of its ability to form a coalition government with one of the major parties by demanding significant concessions: a referendum on permanently banning the use of nuclear power.⁵¹

While the logic of Ackerman's hypothetical is understandable—this is how the Liberal Democrats were able to obtain a referendum on the alternative vote after all—Ackerman overstates his case here. First, Ackerman seems to accept as a matter of course that the third party and one of the major parties would form a coalition government and that the major party would acquiesce to the demand for a referendum, neither of which is certain. If a coalition agreement like that between the Conservatives and Liberal Democrats is evidence of a bargain, surely both parties enter into the agreement because they feel they have something to gain. If the major party feels the price of coalition government is too high, it will consider other options such as forming a different coalition or forming a minority government. Second, Ackerman claims these hypothetical scenarios demonstrate that Parliament is not using the referendum as “an ‘advisory’ device while maintaining its ultimate authority.”⁵² Ackerman's hypothetical scenarios and the historical experience of the referendum in Britain actually show, however, that Parliament uses the referendum as an advisory device.

Referenda are comparatively rare in the United Kingdom. There have been only nine referenda in the country since 1973, only one of which occurred at the national level.⁵³ In the case of each of these referenda, the process has been that Parliament placed the question before the electorate rather than the electorate acting as the originator of the law.⁵⁴ So, too, is that the case under Ackerman's hypothetical scenarios: it is Parliament, through forming a coalition government and through authorizing the referendum, that places the question before the people. In this sense, it is not that popular sovereignty “displaces” parliamentary sovereignty, as Ackerman alleges,⁵⁵ but that Parliament willingly relinquishes a portion of its legislative power

49. *Id.*

50. *Id.*

51. *Id.*

52. *Id.*

53. See SELECT COMMITTEE ON THE CONSTITUTION, *supra* note 41, at 9–10 box 1 (listing the national and regional referenda since 1973).

54. See *id.* ¶¶ 5–6, 142 (discussing the Political Parties, Elections and Referendums Act and noting that referenda held under the Act, of which there has been one, are also held “in pursuance of an Act of Parliament”); DAWN OLIVER, CONSTITUTIONAL REFORM IN THE UNITED KINGDOM 156 (2003) (“To date there has been specific legislation about each of the referendums that has been held.”).

55. See Ackerman, *supra* note 38, at 32 (“If [the referendum for the alternative vote] succeeds, it will mark the displacement of parliamentary sovereignty by popular sovereignty as the foundational principle of the British constitution.”).

and so places the subject of the referendum beyond its own reach. As a practical matter, Parliament is less sovereign than before in that it cannot pass any law it likes. Its position as the final source and arbiter of law, however, remains unaffected. For referenda to truly displace parliamentary sovereignty with popular sovereignty, it would require a system of popularly, rather than parliamentarily, initiated referenda, though that would reflect the state of affairs in California much more so than in Britain.⁵⁶ The constitutional effect of the referendum, therefore, leaves the sovereignty of Parliament neither in as complete a state as Dicey would suggest nor in as incomplete a state as Ackerman would suggest. The effect of the PVS Act as a referendum only partly diminishes the sovereignty of Parliament, and it does not completely displace parliamentary sovereignty for popular sovereignty.

The referendum, however, is not the only content of the PVS Act—the PVS Act also proposes to change the electoral system, which is a significant constitutional reform itself. Ultimately, the alternative-vote provisions of the PVS Act have a similar effect to that which Ackerman alleged in the case of the referendum. Changing the electoral system to use the alternative vote works toward the same result of displacing the sovereignty of Parliament in favor of popular sovereignty. In the case of the alternative-vote provisions, the PVS Act increases the importance of individuals, albeit slightly, in the lawmaking process vis-à-vis Parliament. Though the PVS Act does not completely displace parliamentary sovereignty, as was the case with the referendum, it hints at moving the source of law closer to the people.

II. The Effect of the Alternative Vote on the British Constitution

A. *The Purpose of Elected Government Under the British Constitution*

In order to understand the effects of the alternative vote on the British Constitution and British government, it is important first to describe the system and theory of government that exists under the current British Constitution. In order to describe the form of the government under the British Constitution, Blackstone begins with reference to the three basic forms of government: democracy, aristocracy, and monarchy.⁵⁷ Extolling the benefits of each form of government, Blackstone notes that “[d]emocracies are usually the best calculated to direct the end of a law; aristocracies to invent the means by which that end shall be obtained; and monarchies to

56. See *The Ungovernable State*, ECONOMIST, May 16, 2009, at 33, 33 (criticizing California’s direct-democracy system of governance). It is interesting to note here that the House of Lords Select Committee on the Constitution, in its report on referenda in the United Kingdom, discussed the use of citizen-initiated referenda. The Committee declined to endorse the use of such referenda in Britain. See SELECT COMMITTEE ON THE CONSTITUTION, *supra* note 41, ¶¶ 130, 140 (denying the proposal to allow citizen-initiated referenda).

57. 1 BLACKSTONE, *supra* note 19, at *49.

carry those means into execution.”⁵⁸ By this assertion, Blackstone means that democracies will tend to direct their laws toward the public good, that aristocracies will tend to craft their laws wisely, and that monarchies will be able to carry their laws out swiftly and decisively.⁵⁹ Blackstone also recognizes that each form of government has certain disadvantages as well: that democracies are “frequently foolish in their contrivance”; that aristocracies are less concerned with the public good than democracies and are less powerful and swift than monarchies; and that monarchies can be oppressive and can be directed for private, rather than public, gain.⁶⁰ For Blackstone, the strength of the British Constitution lies in the fact that the executive branch is vested in the hands of the Monarch and that the legislative branch combines the three forms of government by vesting power in Parliament, which includes the Monarch, the House of Lords, and the House of Commons.⁶¹

While Blackstone acknowledges that democracy is an essential element of the British Constitution and of Parliament, its purpose under the British system is very limited. Specifically, democracy is the source of public virtue, and democracy is incorporated into Parliament in order to ensure that it crafts laws for the public good. It is affirmatively not the purpose of democracy to ensure that the laws enacted by Parliament comport with the will of the people in the sense that each law should be as the electorate desires it. Blackstone notes that “every member [of the House of Commons] though chosen by one particular district, when elected and returned, serves for the whole realm.”⁶² For Blackstone, the ultimate purpose of each member of the House of Commons is “not particular, but general; not barely to advantage his constituents, but the *common* wealth.”⁶³ Blackstone concludes that the member “therefore . . . is not bound . . . to consult with, or take the advice of, his constituents upon any particular point, unless he himself thinks it proper or prudent so to do.”⁶⁴

Blackstone’s analysis emphasizes that the will of the individual constituents is of less importance than the good of the commonwealth—of the public as a whole. This application of democracy makes logical sense given Blackstone’s earlier assertion that democracies are sources of public virtue but are subject to foolishness. Representation, therefore, is the tool the British Constitution uses, in Blackstone’s judgment, to obtain the socially beneficial aspects of democracy without succumbing to its faults.

58. *Id.* at *50.

59. *See id.* at *49–50 (stating that democracies create laws with the intentions of public virtue, aristocracies create wise laws, and monarchies are the most powerful).

60. *Id.*

61. *Id.* at *50–51.

62. *Id.* at *159.

63. *Id.*

64. *Id.*

To be sure, it is slightly anachronistic to rely upon Blackstone's view of the British Constitution when the PVS Act will affect the modern British Constitution. Nevertheless, certain aspects of Blackstone's view of democracy survive in the British Constitution to this day, as historical experience and commentators indicate. While describing the doctrine of parliamentary sovereignty, Dicey turns to the case of the Septennial Act,⁶⁵ which will also serve to illustrate Blackstone's view of parliamentary democracy with an historical example. Under the Septennial Act, the sitting Parliament prolonged its own term in office by extending the terms of the members from three to seven years.⁶⁶ To Dicey, the Act serves as proof of parliamentary sovereignty, but it also "proves to demonstration that in a legal point of view Parliament is neither the agent of the electors nor in any sense a trustee for its constituents."⁶⁷ Dicey's thought to connect Blackstone's principle of representation to parliamentary sovereignty is intuitively correct here—the more that representatives must abide by the opinions and views of the constituents, the more it is the constituents, rather than the members of Parliament, who are the source of law. While it is not certain that the Parliament in 1716 passed the Septennial Act with the best interests of its constituents in mind—though the argument can be made that they did⁶⁸—Dicey and Blackstone would argue that this was not required so long as Parliament was acting for the good of the commonwealth.

A variant of this idea remains important today and influences the way commentators and politicians see the purpose of democracy under the British Constitution. In the modern day, "[t]he legal doctrine of the sovereignty of Parliament has . . . come to legitimize a political doctrine, the doctrine that a government enjoying an overall majority in the House of Commons should enjoy virtually unlimited power."⁶⁹ When Bogdanor speaks about the power of government, he is referring to the power of the political party to which the Prime Minister and the cabinet belong. This is to say that representation, in the sense that Blackstone and Dicey would understand it, has come to justify the powers of the Prime Minister and the cabinet in the modern Parliament. Under this view, the purpose of British elections "is not to get a new House of Commons but either to confirm the Government in power or to obtain a different one."⁷⁰ Expressed differently, the goal of the election is not "mathematical accuracy" but rather to capture the "swing" in the country's preference from one party's government to another party's government.⁷¹ In

65. DICEY, *supra* note 26, at 44–45.

66. *Id.* Parliament enacted similar legislation in the early twentieth century and during World War II. *Id.* at 45 n.1.

67. *Id.* at 47–48.

68. *See id.* at 44–45 (noting the justification of the Septennial Act on the grounds of statesmanship, expediency, and preservation of tranquility).

69. BOGDANOR, *supra* note 1, at 15.

70. HARVEY & BATHER, *supra* note 23, at 72–73.

71. *Id.* at 73.

effect, the election serves not as a poll of individual voters on what particular policies they would prefer but as a poll of the commonwealth as a whole to see which government is preferable. Thus, this notion of elections reflects Blackstone's view of representation through its emphasis on the commonwealth over the individual.

Further, the voting system in Britain acts to reinforce this view. To the extent that power in Parliament is concentrated in the hands of the Prime Minister and the cabinet, the British Constitution is "a front-bench constitution."⁷² This is to say that "[l]arge majorities [in the House of Commons] also mean that the government does not need to be responsive to pressure from other parties or even from its own backbenchers, as it is not at risk of losing votes."⁷³ Thus, to the extent that the British electoral system tends to return large majorities in the House of Commons, it reinforces the constitutional importance of Blackstone's view of representative democracy and Dicey's view of the sovereignty of Parliament. It is precisely these elements of the British Constitution that the PVS Act will change through the enactment of the alternative vote.

B. The Alternative Vote Under the PVS Act

1. *The Current First-Past-the-Post Election System.*—Currently, the United Kingdom operates under a first-past-the-post election system. Under this system, the United Kingdom is divided into constituencies, currently numbering 650,⁷⁴ with each constituency returning a single member.⁷⁵ Voters are permitted to vote for only a single candidate during elections, and the candidate receiving the plurality of votes in each constituency is elected.⁷⁶

The Liberal Democrats, as a third party within the current British political system, would stand to gain, at least somewhat if not substantially, as a result of reforming the first-past-the-post electoral system. The United Kingdom's current system of elections, like that used in the United States, tends to award seats in the legislature to various political parties disproportionately in comparison with the percentage of votes that party actually won in the election.⁷⁷ For example, for elections occurring between 1979 and

72. *Id.*

73. OLIVER, *supra* note 54, at 139.

74. *Election 2010: National Results*, *supra* note 3.

75. Parliamentary Constituencies Act, 1986, c. 56, § 2, sch. 2 (U.K.).

76. *First-Past-the-Post*, ELECTORAL REFORM SOC'Y, <http://www.electoral-reform.org.uk/article.php?id=54>.

77. See RON JOHNSTON & CHARLES PATTIE, PUTTING VOTERS IN THEIR PLACE: GEOGRAPHY AND ELECTIONS IN GREAT BRITAIN 270 (2006) (noting the disproportionality of the first-past-the-post electoral system). In the case of British elections, Johnston and Pattie discuss a statistical relationship known as the cube law, which suggests that the ratio of the percentage of seats won by two parties in the House of Commons will tend to be the cube of the ratio of the percentage of votes won by each party. *Id.*

2005 in the United Kingdom, the ratio of the percentage of seats won by the majority party in the House of Commons to the percentage of votes won by that party in the election varied from a low of 1.22 in the 1979 election to a high of 1.53 in the 2001 election.⁷⁸

By contrast, the first-past-the-post system tends to hurt third parties—over that same period, the ratio of the percentage of seats won in the House of Commons by the Liberal Democrats to the percentage of votes the party won in elections has never reached 0.5.⁷⁹ This is the case because the first-past-the-post system tends to encourage large numbers of wasted votes—because a first-past-the-post system is a “winner-takes-all”⁸⁰ system, any votes for a losing third-party candidate in a given district are worthless in that they do not affect the outcome of the election.⁸¹ Further, third parties like the Liberal Democrats must also deal with the problem of tactical voting. Tactical voting is voting behavior where a voter who prefers one party that is likely to lose in an election votes instead for a different party in the hopes of making the vote effective.⁸² Thus, in constituencies where “Labour was best placed to defeat the Conservatives, tactical voting defectors from the Liberal Democrats may have made this feasible.”⁸³ Tactical voting in this sense does not deprive parties of effective votes—the votes lost due to tactical voting would not have created a win for that party in a given constituency in the first place. Tactical voting, however, may create difficulties for third parties in that it may alter voters’ preferences over time in such a way that voters will avoid voting for third parties given the perception that a vote for a third party is likely to be ineffective.⁸⁴ In any case, the current electoral system operates such that election results can be highly disproportional to the votes cast, especially to the detriment of minority parties.

2. *The Proposed Changes in the PVS Act.*—The PVS Act provides for a referendum, to take place between May 5, 2011 and October 31, 2011, that asks voters if the United Kingdom should switch from the current first-past-the-post system of elections to the alternative-vote system.⁸⁵ The alternative-vote portion of the PVS Act would amend the current election rules to enact a simple preference-ordering system for votes. Instead of marking only one

78. *Id.* at 266 tbl.8.1. This is to say that, in 2001 for example, the Labour party won nearly 63% of the seats in the House of Commons despite having won only about 41% of the votes in the election. *Id.*

79. *Id.* at 267.

80. *First-Past-the-Post*, *supra* note 76.

81. *Id.*

82. See JOHNSTON & PATTIE, *supra* note 77, at 296 (describing the Labour party’s policy of promoting tactical voting by Liberal Democrat supporters in constituencies held by the Conservative party).

83. *Id.* at 297.

84. *Id.* at 24–25 (citing claims that third-party voters are now more inclined to switch their vote from their first preference to their second).

85. Parliamentary Voting System and Constituencies Act, 2011, c. 1 (U.K.).

candidate, each voter numerically marks his choices in his preference order, marking 1 for the first choice, 2 for the second, and so on for as many of the candidates as the voter likes.⁸⁶ The PVS Act further amends the vote-counting procedure to redistribute votes according to each voter's preference ranking as is necessary.⁸⁷

Counting of the votes takes place over at least one, and possibly several, rounds.⁸⁸ In the first round of counting, votes are allocated according to each voter's first-preference vote.⁸⁹ For a candidate to be elected, that candidate must receive a majority of all of the first-preference votes cast.⁹⁰ If no candidate receives a majority, the candidate receiving the fewest first-preference votes is eliminated, and each vote that candidate received is reallocated to the remaining candidate the voter ranked highest, if any.⁹¹ Voters who do not specify additional choices do not have their votes reallocated if their choice of candidate is eliminated, and their votes are not considered for subsequent rounds of counting.⁹² The counting process continues by eliminating the candidate with the fewest votes and reallocating votes until one candidate has a majority of the remaining votes.⁹³

3. *The Practical and Constitutional Effects of the Alternative Vote.*—Ackerman argues that the alternative-vote system “will be a boon” to third parties and that “the reformed voting system will give them a good chance of depriving both Conservatives and Labour of the working majorities that have been the norm over the past two centuries.”⁹⁴ Thus, Ackerman accepts the common wisdom that the alternative vote tends to reduce the representation of the majority parties and increase the representation of third parties. To the extent that this common wisdom is correct, the use of the alternative vote may reduce parliamentary sovereignty. This would merely reflect the notion that, to the extent that parliamentary sovereignty is tied to sizable majorities in Parliament, a reduction in the size and strength of those majorities would work against parliamentary sovereignty. In effect, substantially reducing the number of seats won by the majority party would transform Parliament from a government of “front benchers” into a government of “backbenchers” by

86. *Id.* § 9(1).

87. *Id.* § 9(2).

88. *See id.* (calling for repetition of the vote-counting process until one candidate has more votes than all of the other candidates put together).

89. *Id.*

90. *Id.*

91. *Id.*

92. *Id.*

93. *Id.*

94. Ackerman, *supra* note 38, at 32.

increasing the relative power of each member of the majority party vis-à-vis the Prime Minister and the Cabinet.⁹⁵

Studies of recent British elections indicate, however, that the effect of the alternative vote, at least initially, may not be as great as the common wisdom assumes. Analysis of the 2010 British election by the Electoral Reform Society indicates that, if the election had been conducted under the alternative vote instead of the current system, the Liberal Democrats would have won only twenty-two additional seats.⁹⁶ Similarly, the BBC reported a study asserting that the use of the alternative vote would not have affected the outcome of any of the British national elections between 1983 and 2005.⁹⁷ The BBC report does reflect the common wisdom concerning the alternative vote insofar as it indicates that the Liberal Democrats would have gained some seats in each of those elections at the expense of the larger parties, though never enough seats to actually affect the overall outcome of the election.⁹⁸ Of course, the effects of the alternative vote could become more substantial over time as voters alter their first-preference choices away from tactical voting in accordance with the new system. To the extent that a reduction in parliamentary sovereignty would require meaningful reductions in the size of the governing party's majority, however, the constitutional effects of the alternative vote may be quite low if its effects on the seat distribution in the House of Commons are as limited as these studies suggest.

Even if the practical effect of switching to the alternative vote is somewhat limited—though to be sure it could change over time—the change of the electoral system can still have profound constitutional effects because the form of an electoral system reflects the underlying constitutional values of society. In this sense, different electoral systems reflect different choices and value judgments about how government and society ought to operate. At its heart, “the winner-take-all, single-member basis of representation assures that one group, the winning coalition of subgroups constituting the majority, will gain everything and the other group(s) nothing.”⁹⁹ In the most extreme case, a strict first-past-the-post system provides that “[a] well distributed

95. See OLIVER, *supra* note 54, at 139, 146 (describing the lack of influence that backbenchers of the majority party have under the first-past-the-post system and calling for a new system of elections that would more closely reflect voter preferences and increase the power of backbenchers).

96. Alan Travis, *Electoral Reform: Alternative Vote System Would Have Had Minimal Impact on Outcome of General Election*, GUARDIAN.CO.UK (May 10, 2010), <http://www.guardian.co.uk/politics/2010/may/10/alternative-vote-minimal-impact-general-election>.

97. *Would the Alternative Vote Have Changed History?*, BBC NEWS, http://news.bbc.co.uk/2/hi/uk_news/politics/8506306.stm.

98. *Id.* The number of seats the Liberal Democrats would have gained under the BBC's estimates varied from a low of only 6 seats in 2005 to a high of 69 seats in 1997. *Id.* In the case of the 1997 election, the BBC's estimates would have given the Liberal Democrats a total of 115 seats in the House of Commons, which would have made them the main opposition party to Labour, as the Conservatives would have had only 70 seats. *Id.*

99. Sanford Levinson, Commentary, *Gerrymandering and the Brooding Omnipresence of Proportional Representation: Why Won't It Go Away?*, 33 UCLA L. REV. 257, 266–67 (1985).

majority could in theory elect *every* member of the legislature."¹⁰⁰ A first-past-the-post electoral system permits this possibility of disproportionality not because the system is neutral but because the first-past-the-post system embodies a value judgment that is not primarily concerned with "fair and effective representation."¹⁰¹

Instead, under an ideal first-past-the-post system, "politics [is], at best, a process not of bargaining but of creating a collective order with a shared social vision and sense of public interest."¹⁰² Perhaps unsurprisingly, the values of the first-past-the-post system reflect those of Blackstone and Dicey, especially in the sense that legislative politics ought to be directed toward the good of the commonwealth rather than the desires of individual constituents. Under a first-past-the-post system, disproportionality in the allotment of seats is acceptable because the goal of the voting system is not an allotment of seats in the legislature based on voter preference but instead that the legislature put forth a cohesive program directed toward the good of the populace as a whole. In this sense, the first-past-the-post system also comports with the notion concerning the current British system that "[t]he test of the poll will not be mathematical accuracy but whether the result reflects the 'swing' in the country."¹⁰³ This is the case insofar as the first-past-the-post system should indicate which of the two major parties has the cohesive program that the commonwealth as a whole prefers.

In contrast to first-past-the-post, proportional representation would represent a different set of value choices. Proportional representation operates with a "focus on individual rights" and "assumes that people develop their conceptions of themselves and their needs *outside* of the polity."¹⁰⁴ Essentially, this is to say that members of society have their own views and opinions that merit formal recognition in the electoral system and the legislature. Under a proportional-representation system, "[p]olitics is, therefore, reduced to a forum for bargaining and attempted maximization of

100. *Id.* at 267. Levinson illustrates this idea with the example of the 1984 presidential election where "Ronald Reagan's 60% of the popular vote translated into control over 95% of the members of the electoral college." *Id.*

101. *See id.* at 266 n.39 (questioning what theory of elections would allow for a system in which major parties could effectively dilute the strength of minor parties).

102. *Id.* at 274. Here, Levinson is actually describing the idea of classical republicanism held by the framers of the Constitution. *Id.* Levinson describes classical republicanism in contrast to proportional representation. *See id.* at 274-75 (comparing the role of politics in a proportional-representation system with the role of politics in a classical republican system). The framers enacted a first-past-the-post electoral system in this country. *See id.* at 258 ("[W]e have committed ourselves to a geographically based winner-take-all 'system of legislative elections [that by its very nature] looks to a number of separate contests and is thus fundamentally opposed to any concept of adding up national or statewide vote totals.'" (second alteration in original) (quoting Daniel H. Lowenstein & Jonathan Steinberg, *The Quest for Legislative Districting in the Public Interest: Elusive or Illusory?*, 33 UCLA L. REV. 1, 52 (1985))). Given this, Levinson's description should be equally apt when applied to first-past-the-post systems generally.

103. HARVEY & BATHER, *supra* note 23, at 73.

104. Levinson, *supra* note 99, at 274.

individual (or group) preferences.”¹⁰⁵ To be sure, this represents a distinct departure from first-past-the-post’s notion of a shared social vision because proportional representation encourages “partiality in lieu of . . . devotion to the public interest.”¹⁰⁶

In a sense, then, the choice between a first-past-the-post electoral system and a system of proportional representation is a societal enactment of where compromise occurs. Under a system of proportional representation, to the extent that politics is a “forum for bargaining,”¹⁰⁷ it is the legislators who must do that bargaining. Effectively, it is the legislators who must compromise because they are actually beholden to the views of their constituents. This makes the legislators the “trustees” of their constituents. Insofar as a legislator must act to maximize the preferences of his constituents in order to avoid being replaced, it is the constituents who have the power under proportional representation. Of course, this runs exactly counter to the idea of government and parliamentary sovereignty held by Blackstone and Dicey.¹⁰⁸ Under a first-past-the-post system, by contrast, it is the individual constituents who must compromise—each constituent may only select which of the, likely two, shared visions of society held by the major parties is most amenable to him. This system comports, therefore, with the sovereignty of Parliament because the voter has limited power to enact his own personal views, meaning that the legislators hold the power under this system.

While a proportional-representation system would have profound effects on the British Constitution and parliamentary sovereignty,¹⁰⁹ the constitutional effects of the alternative vote are less certain. In a sense, the alternative vote lies somewhere between first-past-the-post and proportional representation—the alternative vote is neither as disproportionate as first-past-the-post nor as proportionate as proportional representation. The alternative vote’s constitutional effects are likely somewhere in the middle as well. The alternative vote marginally increases the power of voters by giving some recognition to their preferred choice of party and, therefore, policy. Parliament, however, also retains significant power, at least in comparison with proportional representation, in that legislators are not directly beholden to the views and opinions of their constituents. Ultimately, the alternative-vote provisions included under the PVS Act likely do not displace parliamentary sovereignty, though they move the British Constitution in the direction of reduced parliamentary sovereignty and increased popular sovereignty.

105. *Id.*

106. *Id.* at 275.

107. *Id.* at 274.

108. *See supra* notes 19–35 and accompanying text.

109. To be sure, a proportional-representation system, such as the single transferable vote, would have an immense practical effect on British elections as well. *See* Travis, *supra* note 96 (modeling the outcome of the 2010 British Election under the single transferable vote).

III. Conclusion

The ultimate constitutional effects of the PVS Act remain uncertain, if only because of the possibility that the British electorate might reject the alternative vote in the referendum. A “no” vote in the referendum, however, would not signal the end of the constitutional reform that the alternative vote represents. To be sure, this is not the first time that Parliament has attempted to change the voting system¹¹⁰—if the current referendum to switch to the alternative vote fails, surely it will not be the last time Parliament ever attempts to move toward a different voting system. Alterations to the electoral system, like all other acts of Parliament, are part of the British Constitution, but these alterations also reflect underlying changes in the values embodied in the British Constitution. The push for the alternative vote or for proportional representation, therefore, represents the change in the British Constitution from government based on parliamentary sovereignty to government based on popular sovereignty.

—*Joshua H. Packman*

110. See *Single Transferable Vote*, ELECTORAL REFORM SOC'Y, <http://www.electoral-reform.org.uk/article.php?id=48> (noting previous attempts in 1917 to institute the single transferable vote that passed the House of Commons but were rejected by the House of Lords).

* * *



JAMAIL CENTER FOR LEGAL RESEARCH
TARLTON LAW LIBRARY
THE UNIVERSITY OF TEXAS SCHOOL OF LAW

The Tarlton Law Library Oral History Series features interviews with outstanding alumni and faculty of The University of Texas School of Law.

Oral History Series

- | | |
|---|--|
| No. 1 - <i>Joseph D. Jamail, Jr.</i> 2005. \$20 | No. 6 - <i>James DeAnda</i> 2006. \$20 |
| No. 2 - <i>Harry M. Reasoner</i> 2005. \$20 | No. 7 - <i>Russell J. Weintraub</i> 2007. \$20 |
| No. 3 - <i>Robert O. Dawson</i> 2006. \$20 | No. 8 - <i>Oscar H. Mauzy</i> 2007. \$20 |
| No. 4 - <i>J. Leon Lebowitz</i> 2006. \$20 | No. 9 - <i>Roy M. Mersky</i> 2008. \$25 |
| No. 5 - <i>Hans W. Baade</i> 2006. \$20 | |

Forthcoming:

Gloria Bradford, Patrick Hazel, James W. McCartney,
Michael Sharlot, Ernest E. Smith, John F. Sutton, Jr.

*Other Oral Histories Published by the
Jamail Center for Legal Research*

- Robert W. Calvert* (Texas Supreme Court Trilogy, Vol. 1). 1998. \$20
Joe R. Greenhill, Sr. (Texas Supreme Court Trilogy, Vol. 2). 1998. \$20
Gus M. Hodges (Tarlton Law Library Legal History Series, No. 3). 2002. \$20
Corwin Johnson (Tarlton Law Library Legal History Series, No. 4). 2003. \$20
W. Page Keeton (Tarlton Legal Bibliography Series, No. 36). 1992. \$25
Jack Pope (Texas Supreme Court Trilogy, Vol. 3). 1998. \$20

Order online at <http://tarlton.law.utexas.edu/> click on Publications

or contact Publications Coordinator,
Tarlton Law Library, UT School of Law,
727 E. Dean Keeton St., Austin, TX 78705

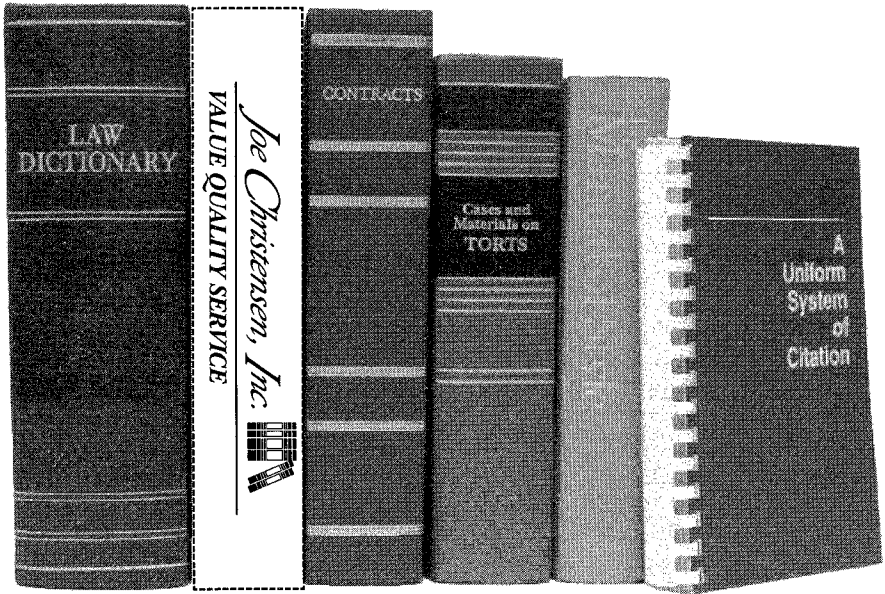
phone (512) 471-6228; fax (512) 471-0243;
email tarltonbooks@law.utexas.edu

THE UNIVERSITY OF TEXAS SCHOOL OF LAW PUBLICATIONS
What the students print here changes the world

<u>Journal</u>	<u>domestic/foreign</u>
Texas Law Review http://www.TexasLRev.com	\$47.00 / \$55.00
Texas International Law Journal http://www.tilj.org	\$45.00 / \$50.00
Texas Environmental Law Journal http://www.texenrls.org/publications_journal.cfm	\$40.00 / \$50.00
American Journal of Criminal Law http://www.ajcl.org	\$30.00 / \$35.00
The Review of Litigation http://www.thereviewoflitigation.org	\$30.00 / \$35.00
Texas Journal of Women and the Law http://www.tjwl.org	\$40.00 / \$45.00
Texas Intellectual Property Law Journal http://www.tiplj.org	\$25.00 / \$30.00
Texas Hispanic Journal of Law & Policy http://www.thjlp.org	\$30.00 / \$40.00
Texas Journal On Civil Liberties & Civil Rights http://www.txjclcr.org	\$40.00 / \$50.00
Texas Review of Law & Politics http://www.trolp.org	\$30.00 / \$35.00
Texas Review of Entertainment & Sports Law http://www.tresl.net	\$40.00 / \$45.00
Texas Journal of Oil, Gas & Energy Law http://www.tjogel.org	\$30.00 / \$40.00
Manuals:	
<i>The Greenbook: Texas Rules of Form</i> 12th ed. ISBN 1-878674-08-0	
<i>Manual on Usage & Style</i> 11th ed. ISBN 1-878674-55-2	

To order, please contact:
The University of Texas School of Law Publications
727 E. Dean Keeton St.
Austin, TX 78705 U.S.A.
Publications@law.utexas.edu

ORDER ONLINE AT:
<http://www.texaslawpublications.com>



We Complete the Picture.

In 1932, Joe Christensen founded a company based on Value, Quality and Service. With 77 years of printing experience, Joe Christensen, Inc. remains the most experienced Law Review printer in the country.

Our printing services bridge the gap between your editorial skills and the production of a high-quality publication. We ease the demands of your assignment by offering you the basis of our business—customer service.

Joe Christensen, Inc. 

1540 Adams Street
Lincoln, Nebraska 68521-1819
Phone: 1-800-228-5030
FAX: 402-476-3094
email: sales@christensen.com

Value

Quality

Service

Your Service Specialists

* * *

* * *

* * *

Texas Law Review

The Greenbook: Texas Rules of Form

Twelfth Edition

A comprehensive guide for Texas citation, newly revised in 2010.

Texas Law Review Manual on Usage & Style

Eleventh Edition

A pocket reference guide on style for all legal writing.

**School of Law Publications
University of Texas at Austin
727 East Dean Keeton Street
Austin, Texas USA 78705
Fax: (512) 471-6988 Tel: (512) 232-1149
Order online: <http://www.texaslawpublications.com>**

