THE LEWIS F. POWELL, JR. DISTINGUISHED LECTURE

From Plyer v. Doe to Trayvon Martin:
Toward Closing the Open Society
Lyle Denniston

ARTICLES

Contingent Capital in Executive Compensation
Wulf A. Kaal

Illuminating Innovation: From Patent Racing to Patent War
Lea Shaver

The Law School Critique in Historical Perspective
A. Benjamin Spencer

NOTES

The Power of a Suggestion: The Use of Forum Selection Clauses by Delaware Corporations
Thomas T. McClendon

Limiting the Legislative Privilege: Analyzing the Scope of the Speech or Debate Clause
Kelly M. McGuire

The Real Error in Citizens United
Joanna M. Meyer

A Case Against the Entire Market Value Rule
Anthony D. Raucci

The Physical Consequences of Emotional Distress: The Need for a New Test to Determine What Amounts Are Excluded from Gross Income Under § 104(a)(2)
C. Anthony Wolfe IV
The LAW ALUMNI ASSOCIATION assists the School of Law by providing the following:

- Tuition Scholarships for students of the School of Law
- Faculty Fellowships to recognize excellence in scholarship and teaching

THE LAW ALUMNI ASSOCIATION

solicits your membership and support of the Law Annual Fund

Your contributions to the Law Annual Fund go toward such specific programs as Law Alumni Association Tuition Scholarships and Faculty Fellowships. Your contributions also help the School maintain its small classes, low faculty-to-student ratio, beautiful campus, and one-on-one attention. The Law Annual Fund enriches the Law School in hundreds of small but significant ways, and makes a critical difference in our ability to fulfill our mission. You can send your contribution to The Law Annual Fund, Washington and Lee University School of Law, Lexington, VA 24450, or you can give online at http://support.wlu.edu/web/page/normal/giftform.html.

Questions? Need more information? Please contact Elizabeth Outland Branner, Director of Law School Advancement, at (540) 458-8191 or brannere@wlu.edu, Sarah Hughes, Assistant Director of Law School Advancement, at (540) 458-8063 or shughes@wlu.edu, or Darlene Moore, Executive Secretary of the Law Alumni Association, at (540) 458-8571 or moored@wlu.edu.
Administration & Faculty—Washington and Lee School of Law 2012–2013

Washington and Lee University

KENNETH P. RUSCIO, B.A., M.P.A., Ph.D.
President of the University

Deans

NORA V. DEMLEITNER, B.A., J.D., LL.M.
Dean & Roy L. Steinheimer, Jr. Professor of Law

JOHANNA E. BOND, B.A., J.D., LL.M.
Associate Dean for Academic Affairs & Associate Professor of Law

JOHN G. KEYSER, B.S., M.S.
Associate Dean for Administration & Technology, Lecturer in Social Sciences & Law

MARY ZANOLLI NATKIN, B.A., J.D.
Assistant Dean for Clinical Education & Public Service, Clinical Professor of Law

STEPHEN TWITTY, B.A., J.D.
Assistant Dean for Student Affairs

SHAWN D. MCSHAY, B.S., M.B.A.
Assistant Dean for Admissions

Emeritus Faculty

DENIS J. BRION, B.S., J.D.
Professor of Law, Emeritus

WILLIAM S. GEIMER, B.S., J.D.
Professor of Law, Emeritus

EDWARD O. HENNEMAN, B.A., J.D.
Associate Professor of Law, Emeritus

FREDERIC L. KIRGIS, B.A., J.D.
Law Alumni Association Professor of Law, Emeritus

LEWIS H. LARUE, A.B., LL.B.
Class of 1968 Law Alumni Professor of Law, Emeritus

ANDREW W. MCTHENIA, JR., A.B., M.A., LL.B.
James P. Morefield Professor of Law, Emeritus

JAMES M. PHEMISTER, B.S., J.D.
Professor of Law, Emeritus

ROY L. STEINHEIMER, JR., A.B., J.D.
Robert E.R. Huntley Professor of Law, Emeritus

JOSEPH E. ULRICH, B.S., J.D.
Professor of Law, Emeritus

Permanent Faculty

C. ELIZABETH BELMONT, B.A., J.D.
Director, Community Legal Practice Center and Associate Clinical Professor of Law

DAVID I. BRUCK, B.A., J.D.
Clinical Professor of Law and Director, Virginia Capital Case Clearinghouse

CHRISTOPHER M. BRUNER, A.B., M.Phil., J.D.
Associate Professor of Law

SAMUEL W. CALHOUN, B.A., J.D.
Professor of Law

ROBERT T. DANFORTH, B.A., J.D.
Professor of Law

KEVIN M. CROTTY, B.A., Ph.D., J.D.
Professor of Classics & Law
Administration & Faculty—Washington and Lee School of Law 2012–2013

MARK A. DRUMBL, B.A., M.A., J.D., LL.M., J.S.D.
Class of 1975 Alumni Professor of Law and Director, Transnational Law Institute

MICHELLE L. DRUMBL, B.A., J.D., LL.M.
Associate Clinical Professor of Law and Director, Tax Clinic

JOSHUA A.T. FAIRFIELD, B.A., J.D.
Associate Professor of Law

JILL M. FRALEY, B.A., J.D., LL.M., J.S.D.
Assistant Professor of Law

SUSAN D. FRANCK, A.A., B.A., J.D., LL.M.
Associate Professor of Law

MARK H. GRUNEWALD, B.A., J.D.
James P. Morefield Professor of Law

BRANT J. HELLWIG, B.S., J.D., LL.M.
Professor of Law

MARGARET HOWARD, A.B., J.D., M.S.W., LL.M.
Law Alumni Association Professor of Law

LYMAN P.Q. JOHNSON, B.A., J.D.
Robert O. Bentley Professor of Law

TIMOTHY S. JOST, B.A., J.D.
Robert L. Willett Family Professor of Law

JOHN D. KING, B.A., J.D., LL.M.
Associate Clinical Professor of Law

ERIK LUNA, B.S., J.D.
Sydney and Frances Lewis Professor of Law

TIMOTHY C. MACDONNELL, B.A., J.D., LL.M., LL.M.
Associate Clinical Professor of Law, Director, Black Lung Clinic

ANN MACLEAN MASSIE, B.A., M.A., J.D.
Professor of Law

RUSSELL A. MILLER, B.A., J.D., M.A., LL.M.
Professor of Law

DAVID MILLON, B.A., M.A., Ph.D., J.D.
J.B. Stombock Professor of Law

JAMES E. MOLITERNO, B.S., J.D.
Vincent Bradford Professor of Law

BRIAN C. MURCHISON, B.A., J.D.
Charles S. Rowe Professor of Law

DOUG RENDLEMAN, B.A., M.A., J.D., LL.M.
Robert E.R. Huntley Professor of Law

CHRISTOPHER B. SEAMAN, B.A., J.D.
Assistant Professor of Law

JOAN M. SHAUGHNESSY, B.A., J.D.
Roger D. Groot Professor of Law

A. BENJAMIN SPENCER, B.A., M.Sc., J.D.
Professor of Law and Director, Frances Lewis Law Center

SARAH K. WIANT, B.A., M.L.S., J.D.
Professor of Law

ROBIN FRETWELL WILSON, B.A., J.D.
Class of 1958 Law Alumni Professor of Law and Law Alumni Faculty Fellow

Visiting Faculty

PAUL BELONICK, B.A., M.A., J.D.
Visiting Professor of Law
Administration & Faculty—Washington and Lee School of Law 2012–2013

MICHAL BUCHHANDLER-RAPHAEL, L.L.B., LL.M., S.J.D.
Visiting Assistant Professor of Law

DONALD EARL CHILDERESS III, B.A., M.A., J.D., LL.M.
Visiting Associate Professor of Law

ROBERT CREIGH DEEDS, B.A., J.D.
Visiting Professor of Law

J. GORDON HYLTON, A.B., J.D., M.A., Ph.D.
Visiting Professor of Law

JUSTICE DONALD W. LEMONS, B.A., J.D.
Distinguished Professor of Judicial Studies

PATRICK LUFF, B.A., J.D.
Distinguished Professor of Judicial Studies

TODD C. PEPPERS, B.A., J.D., Ph.D.
Visiting Professor of Law

JONATHAN SHAPIRO, B.A., J.D.
Visiting Professor of Law

JAMES E. VIATOR, B.A., J.D.
Visiting Professor of Law

CHRISTOPHER J. WHELAN, LL.B., M.A., Ph.D.
Visiting Professor of Law

Adjunct & Extended Faculty

HARLAN RAY BECKLEY, B.S., M.Div., M.A., Ph.D.
Lecturer in Religion & Law, Director, Shepherd Program, Fletcher Otey Professor of Religion

NEIL V. BIRKHOFF, B.A., J.D., LL.M.
Adjunct Professor of Law

MICHELE K. BURKE, B.S., J.D.
Professor of Practice

R. EDWIN BURNETTE, JR., B.A., J.D.
Adjunct Professor of Law

ALBERT V. CARR, B.A., J.D.
Professor of Practice

JUDY CLARKE, B.A., J.D.
Professor of Practice

GREGORY J. COOPER, B.A., Ph.D.
Lecturer in Philosophy & Law, Associate Professor of Philosophy

JOHN H. CRADDOCK, JR., B.S.E., J.D., LL.M.
Professor of Practice

DAVID DEJONG, B.A., J.D., LL.M.
Adjunct Professor of Practice

DANIEL EVANS, B.A. B.S., J.D.
Professor of Practice

KELLY L. FAGLIONI, B.A., J.D.
Professor of Practice

JUDGE JOHN F. FORSTER, JR., B.A., J.D.
Professor of Practice and Director, Judicial Externships

AARON SAMUEL HAAS, B.A., J.D.
Oliver Hill Fellow (Adjunct)

DAVID L. HEILBERG, B.A., J.D.
Adjunct Professor of Practice

JAMES W. JENNINGS, JR., B.A., J.D.
Professor of Practice
Administration & Faculty—Washington and Lee School of Law 2012–2013

TIMOTHY JOSEPH KEEFER, B.A., J.D.
Adjunct Professor of Law

REX LAMB, B.A., J.D.
Adjunct Professor of Law

CHARLES LOWNEY, B.A., M.A., Ph.D.
Lecturer in Philosophy & Law, Visiting Professor of Philosophy

TIMOTHY LUBIN, B.A., M.T.S., Ph.D.
Lecturer in Law and Religion, Professor of Religion

A. CARTER MAGEE, JR., B.A., J.D.
Professor of Practice

JAMES E. MAHON, B.A., M. Phil., Ph.D.
Lecturer in Philosophy and the Law, Professor of Philosophy

JOHN LAWRENCE MANNING, B.S., LL.B., J.D.
Adjunct Professor of Law

JOHN J. MILES, B.S., M.A., J.D.
Adjunct Professor of Law

FRANK W. MORRISON, B.A., J.D.
Adjunct Professor of Law

LAWRENCE LLOYD MUIR, JR., B.A., J.D.
Adjunct Professor of Practice

STEPHEN WILLS MURPHY, B.A., M.Phil., Ph.D., J.D.
Adjunct Professor of Law

CAROL D. NEWMAN, B.A., Ph.D., J.D.
Professor of Practice

J. LEE E. OSBORNE, B.A., J.D., LL.M.
Adjunct Professor of Law

G. MICHAEL PACE, JR., B.A., J.D.
Professor of Practice

JAMES H. PANNABECKER, B.A., J.D., M.R.P.
Professor of Practice

HOWARD PICKETT
Adjunct Instructor of Shepard Poverty

THOMAS H. SPEEDY RICE, B.A., J.D.
Professor of Practice, Transnational Law Institute

CHRISTOPHER B. RUSSELL, A.B., J.D.
Adjunct Professor of Law

GILBERT E. SCHILL, JR., B.A., J.D.
Professor of Practice

JAMES S. SEEVERS, JR., B.A., J.D.
Professor of Practice

GUY BRYAN SEREFF II, B.S., J.D.
Professor of Practice

JONNIE SPEIGHT, B.A., J.D.
Adjunct Professor of Law

GREGORY N. STILLMAN, B.A., J.D.
Professor of Practice

ADAM ROBERT TRUSNER, B.A., J.D.
Adjunct Professor of Law

WILSON F. VELLINES, Jr., B.A., J.D.
Adjunct Professor of Law

DANIEL J. VICTOR, B.A., J.D.
Professor of Practice
Administration & Faculty—Washington and Lee School of Law 2012–2013

WILLIAM P. WALLACE, JR., B.A., J.D.
Adjunct Professor of Law

MARK ALLAN WILLIAMS, B.A., J.D., MBA.
Professor of Practice

R. CRAIG WOOD, B.A., M.Ed., J.D.
Professor of Practice

ROBERT C. WOOD, III, B.A., LL.B.
Adjunct Professor of Law

HENRY L. WOODWARD, A.B., LL.B.
Professor of Practice

CHARLES L. WOODY, B.A., J.D.
Adjunct Professor of Law
Each year, the Frances Lewis Law Center, the Washington and Lee Law Review, and other Washington and Lee University School of Law organizations co-sponsor at least one major symposium examining prospects for change in an important area of the law.

**Previous Major Symposia**

- *Gideon at 50: Reassessing the Right to Council*  
  (November 8–9, 2012)  
  Volume 70:2 (2013)

- *Regulation in the Fringe Economy*  
  (November 10–11, 2011)  
  Volume 69:2 (2012)

- *Restitution Rollout: The Restatement (Third) of Restitution & Unjust Enrichment*  
  (February 25, 2011)  
  Volume 68:3 (2011)

- *Prosecutorial Power: A Transnational Symposium*  
  (April 1–2, 2010)  

- *Protecting Virtual Playgrounds: Children, Law, and Play Online*  
  (October 3, 2008)  
  Volume 66:3 (2009)

- *The Restitution Roundtable*  
  (December 14, 2007)  
  Volume 65:3 (2008)

- *Gender-Relevant Legislative Change in Muslim and Non-Muslim Countries*  
  (March 30, 2007)  
  Volume 64:4 (2007)

- *Understanding Corporate Law Through History*  
  (March 24, 2006)  

For information about upcoming symposia, please contact:  
Wendy Rice, Frances Lewis Law Center, Washington and Lee University School of Law, 1 Denny Circle, Lexington, Virginia 24450; (540) 458-8962.

To obtain Washington and Lee Law Review symposia issues, please contact:  
WASHINGTON AND LEE LAW REVIEW

Statement of Ownership, Management, and Circulation: The Washington and Lee Law Review (ISSN 0043-0463) is owned by Washington and Lee University and is published quarterly by students of Washington and Lee University School of Law. The Law Review is printed by Western Newspaper Publishing Co., Inc., 537 East Ohio Street, Indianapolis, Indiana 46204, and is entered as a periodical with postage paid at post offices in Lexington, Virginia, and Indianapolis, Indiana. POSTMASTER: Send address corrections to Washington and Lee Law Review, Room 408, Washington and Lee University School of Law, 1 Denny Circle, Lexington, Virginia 24450-2116. Alexander M. Sugzda is the 2012–2013 Editor in Chief. The phone number is (540) 458-8566; the fax number is (540) 458-4679; and the e-mail address is: lawreview@wlu.edu. The average number of copies printed of each issue during the preceding twelve months was 520.

Subscription Information: The subscription price for Volume 70 (four issues) of the Washington and Lee Law Review is $35.00 ($40.00 for foreign subscriptions). Please e-mail all subscription requests to the Washington and Lee Law Review at lawreview@wlu.edu. Subscriptions will be renewed automatically unless the Law Review receives notice of cancellation by December 31.

Back Issues: Volumes 68 through 69, issues 1–2 may each be purchased directly from the Washington and Lee Law Review at the current subscription price of $35.00 ($40.00 for foreign subscriptions) per volume. Individual issues cost $25.00 (includes postage). Volumes 1 through 63 may be obtained from:

William S. Hein & Co., Inc.
1285 Main Street
Buffalo, NY 14209-1987
(800) 828-7571; order@wshein.com.

Copies of Articles: Except as otherwise provided, the author of each article in this issue and the Washington and Lee Law Review grant permission for copies to be made available for classroom use, provided that: (1) the copies are distributed at or below cost; (2) the author and the Law Review are identified on each copy; (3) a proper notice of copyright is affixed to each copy; and (4) the user notifies the Law Review that copies have been made. Articles are also available through Westlaw, LEXIS, and Hein Online.

World Wide Web Address: The Washington and Lee Law Review maintains a website at http://law.wlu.edu/lawreview. The website provides general information regarding the Law Review and may be used to order subscriptions, notify the Law Review that copies have been made for classroom use, and contact individual editors and staffwriters.

____________________________________________________________

AN INVITATION TO AUTHORS

The Editorial Board of the Washington and Lee Law Review invites authors to submit articles and essays on areas of the law that are currently of interest to the legal profession. Student-written work is not accepted.


Submission of Manuscripts: Please submit manuscripts to Joanna Heiberg or Anthony Raucci, Senior Articles Editors, Washington and Lee Law Review, Room 408, Washington and Lee University School of Law, 1 Denny Circle, Lexington, Virginia 24450-2116. Electronic submissions may only be made through our ExpressO webpage at http://law.bepress.com/cgi/submit_lawreview.cgi?a=Washingtonallr. Alternatively, hard copy submissions are accepted. Please do not duplicate your submission by sending a hard copy as well as an electronic copy.

All manuscripts must be typed in Microsoft Word, not converted from any other software. The font of both the text and the footnotes should be 11 pt., Century Schoolbook, and double-spaced on 8½ x 11 inch paper. Footnotes should appear at the bottom of each page of text. Text and footnotes should conform to The Bluebook: A Uniform System of Citation (19th ed. 2010). Article submissions should be accompanied by an abstract of not more than 300 words.
THE LEWIS F. POWELL, JR. DISTINGUISHED LECTURE

From *Plyer v. Doe* to Trayvon Martin: 1799
Toward Closing the Open Society
*Lyle Denniston*

ARTICLES

Contingent Capital in Executive Compensation 1821
*Wulf A. Kaal*

Illuminating Innovation: From Patent Racing to Patent War 1891
*Lea Shaver*

The Law School Critique in Historical Perspective 1949
*A. Benjamin Spencer*

NOTES

The Power of a Suggestion: The Use of Forum Selection Clauses by Delaware Corporations 2067
*Thomas T. McClendon*

Limiting the Legislative Privilege: Analyzing the Scope of the Speech or Debate Clause 2125
*Kelly M. McGuire*

The Real Error in *Citizens United* 2171
*Joanna M. Meyer*
THE LEWIS F. POWELL, JR. DISTINGUISHED LECTURE
From *Plyler v. Doe* to Trayvon Martin: Toward Closing the Open Society

Lyle Denniston*

Abstract

Lyle Denniston, the longest serving and most experienced journalist covering the United States Supreme Court, takes his theme of an inclusive and open society from the constitutional and cultural vision of the late Justice Lewis F. Powell, Jr. and then offers a detailed argument that America is forfeiting—or at least compromising—that vision in favor of a safer, more secure and more cramped society, at home and abroad. The Article, taken from a memorial lecture in Justice Powell’s honor at Washington and Lee University in April 2012, draws upon a variety of very different societal and legal developments that are found to have in common the movement of people and ideas to the margins or even out of sight. The analysis ranges from the treatment of individuals captured by the American military during the perceived “war on terrorism,” to the new xenophobic reaction to undocumented immigrants, to the expanded militarization of the American intelligence apparatus, to the narrowing of the supposedly open digital culture, to the new cultural ghettoization of the white-dominated American suburb, to the still rampant exclusion and discrimination against people of color and people of different sexual identities. Against the possible perception that Denniston has extolled only liberal or progressive values, he argues that people of all ideologies have much to lose if American society becomes further afraid of diversity, change, and openness.

It is a distinct privilege and a high honor to be invited to deliver a lecture on your campus, in honor of Justice Powell—an

* The Lewis F. Powell, Jr. Distinguished Lecture, presented April 5, 2012 at the Washington and Lee University School of Law. Presently, Lyle Denniston is the Supreme Court reporter for scotusblog.com, an online clearinghouse of information about the Court.
esteemed graduate of your university and your law school, a 
Virginia gentleman capable of deep kindness and great tolerance, 
and a most distinguished (if initially reluctant) member of the 
Supreme Court of the United States, where he served for fifteen 
Powell's papers,\footnote{See \textit{id.} (“Throughout his life, he kept strong ties to Washington and Lee, to 
which he gave his personal and Supreme Court papers, now housed in the Lewis F. 
Powell Jr. Archives on the campus in Lexington.”); \textit{About the Powell Archives, 
(“The Lewis F. Powell, Jr. Archives comprises the Papers of Lewis F. Powell, Jr., 
other manuscript collections held by the law library, the rare book collection, and 
the archives of the School of Law.”) (on file with the Washington and Lee Law 
Review).} which well document a judicial career in which 
he was a dominant figure, making majorities in an often-
fragmented Court, and standing courageously alone when he felt 
the need to do so.

I am much indebted to your colleague and classmate, Parker 
Kasmer, who has made it easy for me to be here, and I am 
grateful to the students for their civic act in inaugurating the 
Powell Lecture Series a decade ago.\footnote{See \textit{About the Powell Lecture Series, WASHINGTON AND LEE UNIVERSITY SCHOOL OF LAW}, http://law.wlu.edu/powelllecture/page.asp?pageid=856 (last visited Nov. 14, 2012) (“In 2002, the students of Washington and Lee University School of 
Law founded The Lewis F. Powell, Jr. Distinguished Lecture Series in honor of 
former Supreme Court Justice Lewis F. Powell, Jr.”) (on file with the Washington 
and Lee Law Review).} I know I am in good 
company when I take this place.

My topic tonight is a sober one, but I would like to begin on a 
more pleasant note: a story about Justice Powell that, in fact, I 
heard from a student when I visited this campus some years ago. 

A student who sat next to me at dinner said his father was a 
friend of Justice Powell’s, and during a visit to Powell’s chambers, 
was strongly urged to send his son to Washington and Lee. The 
Justice was very enthusiastic about the college during that 
conversation. After the discussion moved on to other topics and 
then ended, the father left Powell’s chambers. As he was walking
down the corridor, Powell came swiftly after him, caught up with
him, and again urged the father to make sure that the boy went
to W&L. This time, though, he offered an argument that he must
have felt would be most convincing: he ticked off the names of
women’s colleges and the exact number of miles each was from
Lexington. The lad did not tell me whether that was the clincher.

May I turn, then, to my subject and start with an opinion
written by Justice Powell almost exactly thirty years ago. It was
a separate opinion in the case of Plyler v. Doe. The case had a
secondary title, and it is far more descriptive: Texas v. Certain
Named and Unnamed Undocumented Alien Children.

Justice Powell’s opinion speaks of an open society that, to
my mind, no longer exists in this country for undocumented
immigrants. You will know my sensitivity on this subject when I
tell you that, as I write stories today about it, I refuse to describe
such human beings as “aliens.”

Still, the actual or threatened closing of an open society goes
well beyond the fate of those who live in America without legal
permission to be here. The trend is, for me, apparent in a variety
of current policies and legal developments, perhaps very different
from each other in fundamental ways but, in combination,
moving people or ideas beyond the pale or, indeed, out of sight.
We find it much easier to define something of which we are
ignorant, or something that stirs fear in us, as necessarily to be
excluded from our circle of acceptance.

Part of it has to do with the creation of pariah classes,
pushed into the category of The Other. Another part of it has to

5. See id. (noting the secondary title).
6. See id. at 236–41 (Powell, J., concurring).
7. See, e.g., id. at 237 (“The classification in question severely disadvantages
children who are the victims of a combination of circumstances. Access from Mexico
to this country across our 2,000-mile border, is readily available and virtually
uncontrollable. Illegal aliens are attracted by our employment opportunities, and
perhaps by other benefits as well.”); id. at 238 (“The classification at issue deprives
a group of children of the opportunity for education afforded all other children
simply because they have been assigned a legal status due to a violation of law by
their parents.”).
8. See generally Adriaan Peperzak, To the Other: An Introduction to the
purdue.edu/purduepress_ebooks/20/ (discussing the philosophical concept of The
Other, a concept first developed by French philosopher Emmanuel Levinas).
do with the refusal to confront—openly and candidly—very deep wrongs. And part of it has to do with intentionally clogging or closing the channels of public discourse.

The phenomenon extends to the creation of a military garrison with wired enclosures at Guantanamo Bay in Cuba, in order to detain—perhaps for the remainder of their lives for some of them—a class of people we want nowhere near our shores and are unwilling to bring before our courts so that their fate might be decided justly.9

And it reaches to a fearsome prison outside of Kabul, Afghanistan, where we have chosen to place another group of people detained as terrorist suspects, precisely to put them beyond the reach of our nation’s courts of law.10

Further, it reaches the impenetrable wall of secrecy that surrounds a national policy that is, slowly but perceptibly, turning the entire American intelligence apparatus into a lethal military machine that does not even exempt United States citizens from its retribution.11

---


Some of those developments may be known to you.

But there is more, much more, in areas of our culture where free exploration had seemed the norm. There is the determined effort—stalled for the time being by a true internet uprising—to put severe new restraints on the openness of digital culture. The Supreme Court has recently declared that there is no guarantee that a public domain in literature and music has any constitutional claim to remain open when the juggernaut of Hollywood-driven copyright bears down upon it.

And there is the steady march of intolerance against people because of the color of their skin or the nature of their sexual identity. We have come to the place, it seems, where a black teenager can be slain in the street by a neighborhood sentinel, and we are left to debate not the inexplicable reason for such an act, but whether it was legally justified by a barrel-chested law that tells one to “stand his or her ground.” And we have come to the place, it is obvious, where three distinguished judges of a state supreme court can be denied reelection solely because they interpreted an equality guarantee in their state’s constitution to allow for same-sex marriage.


13. See Golan v. Holder, 132 S. Ct. 873, 878 (2012) (determining that a statute allowing works to be removed from the public domain “does not transgress constitutional limitations on Congress’ authority”).


15. See Fla. Stat. § 776.013(3) (2012) (“A person who is not engaged in an unlawful activity and who is attacked in any other place where he or she has a right to be has no duty to retreat and has the right to stand his or her ground and meet force with force . . . .”).

16. See Varnum v. Brien, 763 N.W.2d 862, 872 (Iowa 2009) (holding that the
In the face of such unsettling times, let me start over again with Lewis Powell and those “Undocumented Alien Children”\(^\text{17}\) in Texas.

The Texas legislature had passed a law barring local school districts from using any state funds to finance the education of children who were not legal residents of this country and denying such children even entry into the public schools.\(^\text{18}\) By a vote of five to four, in *Plyler v. Doe*, the Supreme Court found that statute to be a violation of the equal protection guarantee of the Fourteenth Amendment.\(^\text{19}\)

Justice Powell wrote separately,\(^\text{20}\) even while joining the majority opinion written by Justice Brennan.\(^\text{21}\) Recalling that the Court had previously made clear that children are not to be condemned for the misdeeds of their parents,\(^\text{22}\) Justice Powell wrote:

> The classification at issue deprives a group of children of the opportunity for education afforded all other children simply because they have been assigned a legal status due to a violation of law by their parents. These children thus have been singled out for a lifelong penalty and stigma. A legislative classification that threatens the creation of an underclass of

---


\(^{19}\) See *Plyler*, 457 U.S. at 230. (striking down the Texas statute because it violated the Equal Protection Clause of the Fourteenth Amendment).

\(^{20}\) See id. at 237–41 (Powell, J., concurring).

\(^{21}\) See id. at 205–30 (majority opinion).

\(^{22}\) See id. at 238 (Powell, J., concurring) (“[V]isiting . . . condemnation on the head of an infant’ for the misdeeds of the parents is illogical, unjust, and ‘contrary to the basic concept of our system that legal burdens should bear some relationship to individual responsibility or wrongdoing.’” (quoting *Weber v. Aetna Cas. & Sur. Co.*, 406 U.S. 164, 175 (1972))).
future citizens and residents cannot be reconciled with one of the fundamental purposes of the Fourteenth Amendment.\textsuperscript{23}

Those words are phrased as law, but they are infused with a sense of the decency that was so much a part of the work of Lewis Powell. There is in those phrases not the remotest hint of intolerance. He laments the failure of the federal government to police the border, but he refuses to make the children pay for that governmental lapse.\textsuperscript{24}

Viewed from today’s perspective, Lewis Powell’s vision of an inclusive society, one generous in its acceptance of people who look and are different from the majority, seems somewhat dated, somewhat unreal, perhaps somewhat naive.

Move forward from \textit{Plyler v. Doe}, leap across thirty years, and one encounters a seemingly different world. One encounters, for example, a state law like Alabama’s House Bill 56.\textsuperscript{25} One of the provisions of that new state law turns public school officials into immigration enforcers, requiring all public schools in the state to determine if each child entering any grade was born outside the United States or is a child of an illegal alien parent.\textsuperscript{26} Because of \textit{Plyler v. Doe}, they cannot be barred from school,\textsuperscript{27} but the information gathered about them can provide the basis for

\textsuperscript{23}Id. at 238–39.

\textsuperscript{24}See id. at 240–41 (“So long as the ease of entry remains inviting, and the power to deport is exercised infrequently by the Federal Government, the additional expense of admitting these children to public schools might fairly be shared by the Federal and State governments.”).


\textsuperscript{26}See id. § 31-13-27(a)(1) (“Every public . . . school in this state, at the time of enrollment . . . shall determine whether the student enrolling . . . was born outside the jurisdiction of the United States or is the child of an alien not lawfully present in the United States . . . .”), invalidated by Hispanic Interest Coal. of Ala. v. Governor of Ala., Nos. 11-14535, 11-14675, 2012 WL 3553613 (11th Cir. Aug. 20, 2012) (finding the provision in violation of the Equal Protection Clause).

\textsuperscript{27}See Plyler v. Doe, 457 U.S. 202, 230 (1982) (determining that Texas cannot “deny a discrete group of innocent children the free public education that it offers to other children residing within its borders” without showing that the denial would further a substantial state interest and finding that Texas did not make the necessary showing).
their, or at least their parents’, deportation in what would surely be a crisis for many families.

Then cross over into Louisiana. Officials of that state recently asked the Supreme Court of the United States to forbid the federal government from counting any undocumented immigrants living anywhere in the United States, and to order the government to recalculate the 2010 Census without them.\(^{28}\) It was an undisguised effort to declare such people outside of the American political community because the Census data is used to calculate the distribution of seats in the United States House of Representatives.\(^{29}\) It is a fact that, since the first census in 1790, all such undocumented people have been counted.\(^{30}\)

The Court did turn down that request, without comment, in March 2012,\(^{31}\) but that did not take away the message that Louisiana’s legal maneuver sent, and, indeed, it does not prevent Louisiana from beginning its challenge anew in a lower court.

In addition, in April 2012, the Supreme Court held a hearing on Senate Bill 1070,\(^{32}\) the Arizona law seeking to control the lives


\(^{30}\) See Campbell J. Gibson & Emily Lennon, Historical Census Statistics on the Foreign-born Population of the United States: 1850–1990 (U.S. Bureau of the Census Population Div., Working Paper No. 29, 1999), available at http://www.census.gov/population/www/documentation/twps0029/twps0029.html (stating that although “[t]he first decennial census of the U.S. population was taken in 1790 . . . in order to obtain the population counts needed for Congressional apportionment . . . [a] question on place of birth, which is the source of data on the foreign-born population, was not added until the 1850 census” (citation omitted)) (on file with the Washington and Lee Law Review).


of undocumented immigrants living in that state—controls that are potentially so onerous that they could only be expected to drive many of those people out of the state altogether, and that is precisely its goal.33

In a nation where, by the year 2050, according to the Census Bureau, 54% of the total population will be made up of minorities,34 these gestures among the states appear to be remarkable efforts to stand against history. And driven as they often are by claims of rampant crime by such immigrants,35 they add to a pervasive complex of fear.

It is that very fear that took the life of seventeen-year-old Trayvon Benjamin Martin on a pleasant street—Retreat View Circle in Sanford, Florida—on February 26.36 The story we know so far is that a neighborhood watch volunteer in a gated community, George Zimmerman, felt the need to pursue that young black man wearing a hooded sweatshirt.37 Zimmerman was diligent about it; he apparently kept up the pursuit even though a


33. See 2010 Ariz. Sess. Laws 113, § 1 (“The provisions of this act are intended to work together to discourage and deter the unlawful entry and presence of aliens and economic activity by persons unlawfully present in the United States.”).


37. See Botelho, supra note 14 (describing what happened on the night Martin was killed).
police dispatcher told him when he called in, “OK. We don’t need you to do that.”

There apparently was a scuffle between the two, and, in a moment, the youth lay in the grass, bleeding to death from a gunshot wound.

Whatever we may assume about racial tension and its possible role in this tragic incident, we do know that it occurred in a community that wanted to be insulated from the threats that lay beyond its gate and its fences. That subdivision’s very name, the Retreat at Twin Lakes, tells of a desire to be left alone.

But if we are to believe the findings in a new book by Rich Benjamin, such gated communities are creating a new cultural reality in America, and it is one in which the incident in Sanford may have been predictable.

Benjamin is a black author whose book is titled Searching for Whitopia: An Improbable Journey to the Heart of White America. He spent two years traveling to see what life would be like in predominantly white, gated communities across the country.

The book, I suspect, will take its place in the annals of America’s cultural studies alongside such classics as Robert and Helen Lynd’s Middletown, their social science reflections on Muncie, Indiana; Thornton Wilder’s play Our Town, about the fictional village of Grover’s Corners, New Hampshire, and Sinclair Lewis’s satirical novel, Babbitt, about a businessman busily climbing the social ladder in the fictional town of Zenith, somewhere in the Midwest.

38. Id. (internal quotation marks omitted).
39. See id. (stating that police “found Martin ‘face down in the grass’”).
40. See DeGregory, supra note 36 (describing the community as a place where “[e]verything was walled in, to keep out the unknown”).
41. See generally Rich Benjamin, Searching for Whitopia: An Improbable Journey to the Heart of White America (2009) (discussing the growth of predominantly white communities, which Benjamin calls “whitopias,” throughout the United States).
42. Id.
43. See id. at 8 (stating that he “lived in three such communities”).
44. Robert S. Lynd & Helen Merrell Lynd, Middletown: A Study in Contemporary American Culture (1929).
Rich Benjamin’s judgment about a typical gated community is harsh.47 As he described his findings in The New York Times in March 2012, he said such a community is “self-contained, conservative and overzealous in its demands for ‘safety.’”48 “Gated communities,” he said, “churn a vicious cycle by attracting like-minded residents who seek shelter from outsiders and whose physical seclusion then worsens paranoid groupthink against outsiders.”49 They have, he said, “a bunker mentality.”50

If Benjamin is right, the presence of a hooded black teenager on a residential street at nighttime would be the personification of The Other,51 a disturber of domestic tranquility. And in the context of my remarks here, what happened to Trayvon Martin may be yet another example of the closing down of an open society.

The phenomenon does not stop there.

Reflect, for a moment, on the “bunker mentality”52 that has developed as America confronted global terrorism in the wake of the attacks of September 11, 2001. Former Vice President Cheney’s “one percent doctrine” exhibited that attitude: if there is a one percent chance that an act of terrorism will occur, it will occur, and preparation must be made to head it off at all cost.53

We have 166 individuals still housed in the military prison at Guantanamo Bay, Cuba, and many of them have been cleared for release, but they have nowhere to go.54 The Congress of the

47. See generally Benjamin, supra note 41 (criticizing “whitopias,” which are becoming increasingly common in the United States).
49. Id.
50. Id.
51. See generally Peperzak, supra note 8.
52. See Benjamin, supra note 48 (discussing the “bunker mentality”).
53. See Ron Suskind, The One Percent Doctrine: Deep Inside America’s Pursuit of Its Enemies Since 9/11, at 62 (2006) (analyzing former Vice President Cheney’s view that “[i]f there’s a one percent chance that Pakistani scientists are helping al Qaeda build or develop a nuclear weapon, we have to treat it as certainty in terms of our response”).
United States has decreed that they cannot be brought to this country. In effect, the lawmakers have declared that the entirety of America is a gated community that these people dare not enter. Congress also has mandated that, if a Guantanamo prisoner should be sent to any other country, his or her life thereafter must be controlled by conditions that no self-respecting sovereign nation would agree to accept within its own borders. Hence, many simply stay at Guantanamo.

Four years ago, the Supreme Court declared that those at Guantanamo had a constitutional right to go to an American court to challenge their continued confinement, and a chance to obtain actual release. But the federal appeals court in Washington, D.C., in a string of decisions, has hollowed out that

10th-anniversary/ (last visited Nov. 14, 2012) ("Today, 171 detainees remain in the detention center, some of them deemed too dangerous to release and others with nowhere to go.") (on file with the Washington and Lee Law Review). Since then, four prisoners have been released or transferred, and one detainees has died. See Rod Nickel, Guantanamo's Last Western Detainee Returned to Canada, REUTERS (Sept. 29, 2012, 3:37 PM), http://www.reuters.com/article/2012/09/29/us-guantanamo-khadr-idUSBRE88S0AN20120929 (last visited Nov. 14, 2012) (stating that 166 detainees now remain at Guantanamo) (on file with the Washington and Lee Law Review).


58. See, e.g., Latif v. Obama, 677 F.3d 1175, 1176 (D.C. Cir. 2011) (vacating the grant of a Guantanamo detainee's petition for writ of habeas corpus and remanding the case for further consideration); Almerfedi v. Obama, 654 F.3d 1, 2 (D.C. Cir. 2011) (reversing the grant of a Guantanamo detainee's petition for writ of habeas corpus), cert. denied, 132 S. Ct. 2739 (2012); Uthman v. Obama, 637 F.3d 400, 402 (D.C. Cir. 2011) (reversing the grant of a Guantanamo detainee's petition for writ of habeas corpus and ordering the petition to be denied on remand), cert. denied, 132 S. Ct. 2739 (2012); Al-Bihani v. Obama, 590 F.3d 866, 869 (D.C. Cir. 2010) (affirming the denial of a Guantanamo detainee's petition for writ of habeas
legal remedy to the point where a judge on that court has said publicly that he does not believe any of his colleagues would uphold any federal judge’s order to release anyone from Guantanamo if there is even the slightest chance that they are, in fact, inclined toward terrorism\(^{59}\)—when terrorism is defined so loosely that no one can know precisely what it means.

And it is not even clear that habeas corpus, an increasingly limited legal option open to those held at Guantanamo Bay, is available for hundreds of other prisoners held by the United States military at a prison in Afghanistan.\(^{60}\) Those prisoners’ fates may be determined entirely by the Afghan government now that control of that prison has been handed over to local authorities.\(^{61}\)

Certainly, the government faces fierce and awesomely complicated challenges on what to do about terrorism, but surely we do not need to send the Constitution into wartime exile as part of that response.

We seem also to have put the Constitution’s core concept of checks and balances on hold as we seek to know how the government is actually waging a “war on terrorism.”

A doctrine of “state secrets” that began in 1950 as a limited means of excluding specific items of evidence bearing on national security from being explored in court cases\(^{62}\) has now mushroomed into a doctrine that shuts down judicial inquiry altogether, merely upon a classified declaration by a government

---

59. See Esmail v. Obama, 639 F.3d 1075, 1078 (D.C. Cir. 2011) (Silberman, J., concurring) (“I doubt any of my colleagues will vote to grant a petition if he or she believes that it is somewhat likely that the petitioner is an al Qaeda adherent or an active supporter.”).

60. See Al Maqaleh v. Gates, 605 F.3d 84, 99 (D.C. Cir. 2010) (holding “that the jurisdiction of the courts to afford the right to habeas relief and the protection of the Suspension Clause does not extend to aliens held in Executive detention in the Bagram detention facility in the Afghan theater of war”).


security official.63 And those who have had a peek at the reasons claimed for such invocations of national security, while not revealing what they had learned, have said that the claims of national security are frequently exaggerated and often are based on no more than unfiltered raw intelligence of an extremely dubious kind.64

There is a world of difference, of course, between an open society and a national security state, but one must ask whether there is no middle ground. Is there no way by which the Judiciary can inquire into the use of torture as a means of intelligence gathering? Is there no means by which the Judiciary can examine the policy of targeted killing even of United States citizens abroad who are deemed to be terrorist threats? And is there nowhere within the structure of government where the secret world of a thoroughly militarized Central Intelligence Agency can be meaningfully examined?

Put another way, an abiding question for Americans living in this frightened age is whether accountability has become an unaffordable luxury, whether a faith in democratic principle is naively out of date.

Are there other threats to the open society that we thought we had? Of course there are, and they have nothing to do with national security.

America has never had an opportunity for openness to match what came with the arrival of the Digital Age. We know so much more about each other, including, perhaps, some things we would just as soon not know. But leaving that aside, has any one of you in this audience never marveled about what could be reached with a click or two of a mouse? What a universe of ideas and


64. See Latif v. Obama, 677 F.3d 1175, 1208 (D.C. Cir. 2011) (Tatel, J., dissenting) (“[T]he Report at issue here was produced in the fog of war by a clandestine method that we know almost nothing about. It is not familiar, transparent, generally understood as reliable, or accessible; nor is it mundane, quotidian data entry akin to state court dockets or tax receipts.”); see also Thomas I. Emerson, National Security and Civil Liberties, 9 YALE J. WORLD PUB. ORD. 78, 80 (1982) (stating that “[s]ubsequent events almost always demonstrate that the asserted dangers to national security have been grossly exaggerated”).
experiences is out there, and so very much of it is available free—free of cost, free of predigestion, free of censorship, free of bias, free of lunacy. The selection of what to examine and to absorb or reject is solely by unmediated private choice.

And I am not talking about the number of apps that you have on your phone that enable you to live your daily life so much more conveniently. I would imagine you have seen that television commercial in which one person has a great idea about something, only to be told by another person that it has already happened so many seconds ago. It almost seems that we have created apps that allow time to run backwards!

But the digital world is not as open as you might think, and there are very real signs that it may lose more of its openness. From early on in this age, we have learned that electronic exploration and the concept of copyright were on a collision course or at least were in considerable tension. You may remember the Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.65 episode, in which the Supreme Court informed us that the free trade in creative expression was not really very free after all.66 We have heard a lot since then about piracy.

That was nothing, though, in comparison to the restraints on digital communication that proposed new federal legislation would bring about. The so-called Stop Online Piracy Act,67 and variations of it, appeared to be absolutely draconian to major players in the digital community.68 And just when it seemed that the major entertainment and recording companies had Congress in their control on this issue, the Internet rose up in angry protest in January and drove that legislation off of the

66. See id. at 919 (“We hold that one who distributes a device with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement, is liable for the resulting acts of infringement by third parties.”).
calendars in both the House and Senate. So much for a passive citizenry. But, we are told, the antipiracy movement is far from dead; it is just lying in wait for another opportunity.

One very unsettling facet about that entire incident is that among the loudest cheerleaders for the legislation was none other than the nation’s most fervent defender of First Amendment freedoms, renowned New York attorney Floyd Abrams.

One can be reasonably certain, I would suggest, that if the Stop Online Piracy Act or some version of it does emerge from Congress, its validity will be upheld by the Supreme Court—as long as the Court’s current prevailing view of copyright holds.

While the Constitution speaks of copyright as protected only for “limited times,” the Court in the 

Eldred v. Ashcroft decision in 2003 upheld the concept that limited does not really mean very limited. That decision validated the Hollywood-initiated Sonny Bono Copyright Term Extension Act that


71. See Letter from Floyd Abrams, Partner, Cahill Gordon & Reindel LLP, to Chairman Lamar Smith & Ranking Member John Conyers of the House Comm. on the Judiciary 13 (Nov. 7, 2011), http://www.mpaa.org/Re sources/1227ef12-e209-4edf-bb8b-8bab4a768430c.pdf (“This proposed legislation is not inconsistent with the First Amendment; it would protect creators of speech, as Congress has done since this Nation was founded, by combating its theft.”).

72. See U.S. CONST. art. I, § 8, cl. 8 (giving Congress the power “[t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries . . .” (emphasis added)).


74. See id. at 204 (allowing Congress to extend the duration of existing copyrights).

extended existing copyright protection on many creative works for very long additional periods.\textsuperscript{76}

And \textit{Eldred}, it seems, did not reach the outermost expansion of copyright. Just this past January, in the decision in \textit{Golan v. Holder},\textsuperscript{77} the Supreme Court essentially destroyed—at least in constitutional terms—the concept of a “public domain.”\textsuperscript{78}

For at least the past thirty years, scholars have been developing the notion that, when a once-copyrighted work reaches the end of that monopoly protection, it enters into a “public domain” that has its own claim to constitutional protection, under the First Amendment.\textsuperscript{79} Works that had never been put under copyright in the United States were already in the “public domain” and, so it was argued, could not be withdrawn into a copyrighted status.\textsuperscript{80}

The \textit{Golan} decision explicitly rejected that idea as a constitutionally grounded principle.\textsuperscript{81} Those who have been using for their own purposes a literary or musical work that has been in the public domain do not thereby acquire First Amendment protection against having such a work newly copyrighted, the Court said.\textsuperscript{82}

No one, the Court declared, obtains any personal right under the Constitution to copy or perform a work just because it has

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{76} See id. (extending copyright terms in the United States by twenty years).
\item \textsuperscript{77} \textit{Golan v. Holder}, 132 S. Ct. 873 (2012).
\item \textsuperscript{78} See id. at 878 (“Neither the Copyright and Patent Clause nor the First Amendment, we hold, makes the public domain, in any and all cases, a territory that works may never exit.”).
\item \textsuperscript{80} See, e.g., Tyler T. Ochoa, \textit{Origins and Meanings of the Public Domain}, 28 U. DAYTON L. REV. 215, 259 (2002) (“[T]he defining characteristic of the public domain in intellectual property is that any individual is free to use the material as he or she sees fit; and once matter enters the public domain, the government cannot alienate that ‘property’ by removing it from the public domain.”) (citations omitted).
\item \textsuperscript{81} See \textit{Golan}, 132 S. Ct. at 889–91 (explaining why a statute granting copyright protection to works already in the public domain did not violate the First Amendment).
\item \textsuperscript{82} See id. (finding that “nothing in the historical record, congressional practice, or [the Court’s] own jurisprudence warrants exceptional First Amendment solicitude for copyrighted works that were once in the public domain”).
\end{enumerate}
\end{footnotesize}
come out from under earlier copyright protection or has never been copyrighted in the United States, so no one can object if copyright is later conferred. Any legal rights that exist belong only to the author or composer, it added.

And to add to the indignity of this insult to the public domain, the Court’s opinion disposed of the claim of constitutional significance for a public domain in a footnote—Footnote 26.

The greatest casualty of the *Golan* and *Eldred* decisions, of course, is the global digital library that has been developing in recent years, and has been profoundly dependent upon the existence of a viable public domain.

The risk to an open society of creative expression is now beyond dispute, even granting that some measure of copyright protection does contribute to the creative urge in artists, writers, and composers. How long, we may ask, must it exist before it stifles further creative expression that is derived from an original work?

I would like to turn then to a final—and probably more controversial—threat to an open society. I refer, of course, to the effort, pursued in so many sectors of our public policy and our civic discourse, to exclude gays and lesbians from an equal role in society and its institutions—especially in those central institutions in mainstream culture, marriage and the military.

83. See id. at 892 (“Anyone has free access to the public domain, but no one, after the copyright term has expired, acquires ownership rights in the once-protected works.”).

84. See id. (“Rights typically vest at the outset of copyright protection, in an author or rightholder.” (citation omitted)).

85. Id. at 888 n.26 (disagreeing with the dissent’s suggestion “that at least where copyright legislation extends protection to works previously in the public domain, Congress must counterbalance that restriction with new incentives to create”).

86. See Defense of Marriage Act, Pub. L. No. 104-199, § 3, 110 Stat. 2419, 2419 (1996) (codified at 1 U.S.C. § 7) (“[T]he word ‘marriage’ means only a legal union between one man and one woman as husband and wife, and the word ‘spouse’ refers only to a person of the opposite sex who is a husband or a wife.”).

And this, of course, gives me an opportunity to return to Lewis Powell and his approach to life and the law.

The gay rights issue, overall, is among the most agonizing public questions confronting America, and Lewis Powell’s own struggle with it was itself agonizing, as is so painfully examined in John Jeffries’s great biography of the Justice.\footnote{See John C. Jeffries, Jr., Justice Lewis F. Powell, Jr. 511–30 (1994) (describing Justice Powell’s struggle with the gay rights issue).} Jeffries’s ultimate conclusion is that, “[o]n this issue, Powell never found the middle ground.”\footnote{Id. at 530.}

Jeffries’s chapter on \textit{Bowers v. Hardwick},\footnote{See id. at 511–30; see also Bowers v. Hardwick, 478 U.S. 186, 196 (1986) (finding a Georgia law criminalizing consensual homosexual sodomy constitutional), overruled by Lawrence v. Texas, 539 U.S. 558, 578 (2003).} the 1986 decision refusing to recognize constitutional equality for sexual intimacy for homosexuals, spells out in detail Justice Powell’s role in that case, switching sides but ultimately voting with the majority.\footnote{See Jeffries, supra note 88, at 513–30 (discussing Justice Powell’s role in the \textit{Bowers} decision, including his indecisiveness and changes of mind).} In a real sense, Powell even then recognized the long-range constitutional implications of the controversy—the effect that could be expected on marriage and the military.\footnote{See id. at 518 (describing Powell’s anticipation of the effect that \textit{Bowers} would have on marriage and the military).} And he fretted deeply about that.\footnote{Id. at 530.}

Although he would say, four years later, that he had made a mistake in \textit{Bowers}, and that the dissenters there “had the better of the arguments,”\footnote{See id. (acknowledging Powell’s concerns over the implications of \textit{Bowers}).} it is far from clear that his mind and heart were ever settled on the ultimate question of gay equality. He died five years before \textit{Bowers} was explicitly overruled in \textit{Lawrence v. Texas},\footnote{See Lawrence v. Texas, 539 U.S. 558, 578 (2003) ("\textit{Bowers v. Hardwick} should be and now is overruled."); Greenhouse, supra note 1 (announcing Justice Powell’s death in 1998).} which, of course, set the constitutional stage for challenges to the exclusion of gays both from the military and from the civil institution of marriage.

America is probably more open politically to homosexual equality than it was in Lewis Powell’s day, and that ultimately
may make a constitutional difference. The resolution of the issue, though, could certainly use—as part of the discussion—the sense of decency and humility that was so much a part of Justice Powell’s life.

When one delivers a memorial lecture like this one, in honor of a person no longer with us, there may be a strong temptation to enlist that celebrated person as an ally on one side or the other of contemporary controversy. One might ask a variation of that popular question of today: What would Lewis Powell say? And one might pretend to know the answer.

I have no idea how he would have reacted to each or any of the specific examples I have put forth here, or on the general question of what I see as the accelerating trend toward closing an open society. I have to believe he would have counseled caution and prudence, the exercise of measured judgment, and the patience to hear what the other side might have to say.

Lewis Powell was a conservative, as a man and as a Justice, but I do not mean by that to place him in a static category along a fixed ideological spectrum. His conservatism seemed to me always to bespeak his belief that even society’s most intractable problems could be resolved if calm judgment were brought to bear in a spirit of mutual respect and accommodation.

Many of the views that I have stated here today are, I recognize, at least tinged with values that most would recognize—in today’s political parlance—as liberal in nature. I would feel that I had failed if they were received here only in that way.

The openness of American society is not a liberal or a conservative cause; it is the essence of a society unafraid of change or of diversity, the character of a society that appreciates that time does not stand still and that civic reform is never fully done.

And it is a society that looks forward to each new journey along an open road, remembering the advice that John Milton gave to Parliament nearly four centuries ago: “[H]e who thinks we are to pitch our tent here, and have attain[e]d the utmost prospect of reformation... that man by this very opinion declares that he is yet far[] short of [t]ruth.”96

Thank you.

ARTICLES
Contingent Capital in Executive Compensation

Wulf A. Kaal*

Abstract

Contingent capital has great potential to improve corporate governance in Systemically Important Financial Institutions (SIFIs). Early initiatives by European SIFIs to include contingent convertible bonds in executive compensation packages lack governance-improving designs. This Article suggests the use of contingent convertible bonds with an early conversion trigger in executive compensation. The proposal adds an important element to the literature on inside debt and the creditor-centered approach to executive compensation. Contingent convertible bonds with early triggers could be preferable to other debt instruments because, in addition to lowering income inequality and increasing sustainability, the early trigger design can improve incentives for executives to lower risk-taking, improve signaling of default risk, and increase incentives for monitoring by creditors and shareholders. The recognition of ownership characteristics in design features adds an important element to the literature on contingent capital trigger designs. The methodological assumptions of incomplete contract theory can improve the analysis of executive compensation arrangements.

Table of Contents

I. Introduction ................................................................. 1822

* Associate Professor, University of Saint Thomas School of Law (Minneapolis). The author would like to thank his colleagues for comments and suggestions on earlier drafts, especially Rudolf Richter, Geoffrey Miller, Mark Ramseyer, Claire Hill, Lisa Casey, Richard Painter, Lyman Johnson, Alan Palmiter, Brett McDonnell, Michelle Harner, Tom Joyce, and James Gillespie. The author is also grateful for outstanding research assistance from librarians Valerie Aggerbeck, Debby Hackerson, and Mary Wells.
Since the financial crisis of 2008–2009 (Financial Crisis), existing executive compensation policies and the level of executive compensation have been increasingly scrutinized.¹


2. See generally Lucian A. Bebchuk, Alma Cohen & Holger Spamann, The Wages of Failure: Executive Compensation at Bear Stearns and Lehman 2000–2008, 27 YALE J. ON REG. 257 (2010); see also Bebchuk & Spamann, supra note 1, at 274–75 (arguing that because shareholders are incentivized to encourage management to take risks beyond the socially optimal level, corporate governance reforms should include reforms of executive compensation policies). Because of implicit government guarantees for bank debt, existing executive compensation policies do not incentivize bondholders and other creditors to monitor risk-taking by executives. Id.

compensation in executive pay packages may have resulted in inappropriate executive risk-taking,\textsuperscript{4} short-termism,\textsuperscript{5} a lack of sustainability, income inequality, and a classic moral hazard problem.\textsuperscript{6}

For more than three decades, theoretical research on executive compensation has focused almost exclusively on adjusting executive compensation with equity-based products alone, such as stocks and stock options.\textsuperscript{7} While equity-based compensation policies may increase risk-taking, some empirical studies have shown that risk-taking can decline if executives hold more debt relative to their equity holdings.\textsuperscript{8} An increasing part of

\begin{thebibliography}{9}
\item 5. \textit{Id.}
\item 6. Executive remuneration via stock options resulted in executives sharing in shareholders' gain but insulated them from shareholders' losses. This may have led executives to use excessively risky strategies because there was no penalty for management. See Bebchuk & Spamann, supra note 1, at 248–50 (discussing the role of excessive risk-taking in the financial sector and its impact on the Financial Crisis); see also Bebchuk, Cohen & Spamann, supra note 2, at 259 (analyzing executive compensation at Bear Stearns and Lehman from 2000–2008); Rüdiger Fahlenbrach & Rene M. Stulz, \textit{Bank CEO Incentives and the Credit Crisis}, 99 J. Fin. Econ. 11, 12 (2011) (arguing that the most plausible explanation for these findings is that CEOs “took actions that they believed the market would welcome,” but “[e]x post, these actions were costly to their banks”); Andrea Beltratti & Rene M. Stulz, \textit{Why Did Some Banks Perform Better During the Credit Crisis? A Cross-Country Study of the Impact of Governance and Regulation} (Nat'I Bureau of Econ. Research, Working Paper No. 15180, 2009), http://www.nber.org/papers/w15180.pdf.
the literature is now considering the role of debt for manager incentives (the creditor-centered approach).\(^9\) Debt in executive compensation packages can help lower risk incentives, lower income inequality, address short-termism, and create sustainability.\(^10\)

The benefits of debt in executive compensation packages can be enhanced by using contingent convertible bonds (CCBs),\(^11\) which can either be written down or converted into equity upon a triggering event.\(^12\) Unlike traditional contingent convertible

---


10. See generally Bebchuk & Spamann, supra note 1.

11. Other terms for contingent capital securities (CCSs) include contingent convertible bonds (CCBs), or CoCos. This Article will predominantly refer to these hybrid instruments as contingent convertible bonds or CCBs.

12. See CHARLES HIMMELBERG, AMANDA HINDLION, SANDRA LAWSON &
bonds issued to investors, the emphasis for contingent convertible bonds in executive compensation is not on a capital infusion when the Systemically Important Financial Institution (SIFI) is in a crisis,\(^\text{13}\) but rather on governance-improving designs to help optimize management’s incentives.\(^\text{14}\) Adding contingent convertible bonds with an early trigger to the calibration of executive compensation packages could improve the corporate governance of SIFIs.\(^\text{15}\) More specifically, contingent convertible bonds with early triggers in executive compensation packages can help to improve incentives for risk-taking by executives, facilitate monitoring by creditors and shareholders, align executives’ interests with those of different constituents, promote sustainability, and reduce income inequality.\(^\text{16}\)

Using contingent convertible bonds with early triggers in executive compensation is not a mere theoretical proposal.


\(^\text{14}\) See generally Wulf A. Kaal, Initial Reflections on the Possible Application of Contingent Capital in Corporate Governance, 26 NOTRE DAME J.L. ETHICS & PUB. POLY 101 (2012) (summarizing the possible applications of contingent convertible bonds in corporate governance); see also Wulf A. Kaal & Christoph K. Henkel, Contingent Capital with Sequential Triggers, 49 SAN DIEGO L. REV. 221 (2012). The conversion feature of CCBs and the threat of dilution for equity holders could change the power structure, control dynamic, and dependencies within SIFIs. The market in contingent convertible bonds is slowly evolving. See infra Part VI.B. With increasing issuances, contingent capital design features will continue to develop. The efficient functioning of contingent capital designs could benefit from experimentation and a learning experience that takes into account corporate governance considerations. Combined with other corporate governance mechanisms, CCBs, as an internal institution-specific mechanism, could help fill the void left by regulators’ seeming inability to supervise financial institutions effectively.

\(^\text{15}\) See Coffee, supra note 9, at 807 (discussing how implementing an early trigger for conversion may deter excessive risk-taking).

\(^\text{16}\) See infra Part V.B.2.
Barclays, Inc. (Barclays) has already issued contingent capital securities to its executives under its Contingent Capital Plan (CCP).\textsuperscript{17} Barclays's CCP, however, does not allow for conversion from debt into equity, much less with an early trigger before any other contingent convertible bonds can be triggered. Barclays's issuance seems to serve a mere signaling function, only marginally improves corporate governance, and leaves the incentives for executives untouched. The design adjustment proposed in this Article helps optimize the effectiveness and corporate governance improvements of contingent convertible bonds in executive compensation.

This Article includes five Parts. Part II evaluates reform proposals for executive compensation policies before and after the enactment of the Dodd–Frank Act and demonstrates that the creditor-centered approach to executive compensation adds important elements to the debate on reform proposals. Part III introduces the concept of contingent capital securities, contingent capital's quasi-public-good characteristics, and the possible application of contingent capital bonds for corporate governance improvements in SIFIs. Part IV shows that the relational elements in executive compensation contracts are inadequately acknowledged by classical contract theory and spot contract theory. The shortcomings in the analysis of executive compensation contracts under the classical contract model and spot contract model can be overcome with the methodological assumptions of the relational or incomplete contract model in New Institutional Economics. Part V introduces the idea of contingent capital bonds in executive compensation and highlights the design and governance shortcomings in Barclays's Contingent Capital Plan. Part V shows that contingent convertible bonds with early triggers can add important elements to the literature on inside debt and the creditor-centered approach to executive compensation. The recognition of

\textsuperscript{17} See Barclays, Barclays PLC Annual Report 2010 172 (2010) [hereinafter Barclays Annual Report], http://reports.barclays.com/ar10/files/Annual_Report_2010.pdf (stating that 50% of the deferred incentive rewards will be contingent capital awards under the CCP); Megan Murphy & Jennifer Hughes, Barclays Causes a Stir with Cocos Plan, Fin. Times, Jan. 25, 2011, at 22 (discussing the issuance of contingent convertible bonds as part of employees’ bonuses) (on file with the Washington and Lee Law Review).
ownership characteristics in design features also adds an important element to the literature on contingent capital trigger designs.

II. Reform Proposals for Executive Compensation

Corporate governance reform proposals after the Financial Crisis have recognized the importance of executive compensation.18 Although the effects of equity-based compensation are unclear and are the subject of a long academic debate,19 equity-based compensation predominates

18. The Financial Crisis Inquiry Commission lists executive compensation as one of the primary factors contributing to the crisis (among other factors such as lack of transparency, excessive borrowing, and high risk investments). FINAL REPORT, supra note 1, at xix, xxvi. Executive compensation takes a prominent role among other important factors (such as accounting, liquidity, and capital regulation) in the Financial Services Authority's Turner Review in the United Kingdom. FIN. SERVS. AUTH., THE TURNER REVIEW: A REGULATORY RESPONSE TO THE GLOBAL BANKING CRISIS 80 (2009), http://www.fsa.gov.uk/pubs/other/turner_review.pdf (highlighting contributing factors and international approaches to banking reform).

executive compensation in the United States. The focus on equity-based compensation in executive compensation packages may have resulted in short-termism, suboptimal incentives for managers, a lack of sustainability, income inequality, and a classic moral hazard problem. The appropriate response to the shortcomings in equity-based executive compensation is debated among academics and


20. See DANIEL A. COHEN, AIYESHA DEY & THOMAS Z. LYS, THE SARBANES–OXLEY ACT OF 2002: IMPLICATIONS FOR COMPENSATION STRUCTURE AND RISK-TAKING INCENTIVES OF CEOs 29 (2004), http://leeds-faculty.colorado.edu/Bhagat/SOX-CEO-Compensation-Investment.pdf (covering the period from 1992 to 2003 and dividing compensation into fixed salary, bonuses, and options); Cassell et. al., supra note 8, at 597 (depicting CEO-to-firm debt–equity ratio); Sundaram & Yermack, supra note 8, at 1553 (defining the ratio of equity-to-inside-debt as benefit pensions and deferred compensation); Tung & Wang, supra note 8, at 13 (measuring debt as defined benefit pension and deferred compensation).

21. See Bebchuk & Spamann, supra note 1, at 249–50 (discussing problems in bank executives’ pay, possible remedies, and government intervention); Bebchuk, Cohen & Spamann, supra note 2, at 259 (analyzing executive compensation at Bear Stearns and Lehman from 2000–2008); Fahlenbrach & Stulz, supra note 6, at 12 (arguing that the most plausible explanation for these findings is that CEOs “took actions that they believed the market would welcome,” but “[e]x post, these actions were costly to their banks”); Beltratti & Stulz, supra note 6, at 1–2 (investigating limitations for bank performance during the Financial Crisis).

22. See, e.g., Jennifer G. Hill, Regulating Executive Remuneration after the Global Financial Crisis: Common Law Perspectives, in RESEARCH HANDBOOK ON EXECUTIVE PAY (Jennifer G. Hill & Randall S. Thomas eds., 2012) (discussing the nexus between executive compensation and the Financial Crisis); Bebchuk & Spamann, supra note 1 (analyzing how banks’ compensation structures produced incentives for excessive risk-taking); see also Posner, supra note 1, at 1040–41 (arguing that, while the Financial Crisis cannot be attributed directly to executive overcompensation, CEOs have an incentive to increase leverage because of compensation tied by stock options to share value, generous severance packages, etc.); FINAL REPORT, supra note 1, at xix (stating that “compensation systems—designed in an environment of cheap money, intense
policy makers.\textsuperscript{23} The debate pre- and post- Dodd–Frank Act has been shaped by the apparent shortcomings in equity-based competition, and light regulation—too often rewarded the quick deal, the short-term gain—without proper consideration of long-term consequences). “Often, those systems encouraged the big bet—where the payoff on the upside could be huge and the downside limited. This was the case up and down the line—from the corporate boardroom to the mortgage broker on the street.” \textit{Id. But see} Fahlenbrach & Stulz, \textit{supra} note 6, at 11–12 (investigating whether bank performance during the recent credit crisis is related to CEO incentives before the crisis and finding some evidence that banks with CEOs whose incentives were better aligned with the interests of shareholders performed worse and no evidence that they performed better); Beltratti & Stulz, \textit{supra} note 6, at 2–5 (examining whether bank performance is related to bank-level governance, country-level governance, country-level regulation, and bank balance sheet and profitability characteristics before the crisis).

executive compensation, by experiences with Enron and other financial scandals, and by the Sarbanes–Oxley Act, as well as by the Financial Crisis. Recent efforts to include debt-based instruments in executive compensation could help address many of the shortcomings in equity-based compensation.

A. Pre Dodd–Frank Act

There is some evidence that executive compensation played a role in governance shortcomings before the Enron scandal.24 Despite this evidence, the regulatory response after the collapse of Enron focused predominantly on audit failure.25 Executive compensation was only marginally addressed in the Sarbanes–Oxley Act (SOX).26 Increased shareholder involvement in

remuneration, requirements on risk alignment and disclosure in the context of remuneration policies and practices); FIN. STABILITY FORUM, FSF PRINCIPLES FOR SOUND COMPENSATION PRACTICES (2009), http://www.financialstabilityboard.org/publications/r_0904b.pdf (describing the FSF Principles for Sound Compensation Practices and how they apply to major financial institutions).


executive compensation, director elections, and other corporate governance matters were notably absent from the Sarbanes–Oxley Act. Reform proposals after the Enron scandal but before the Financial Crisis and the eventual enactment of the Dodd–Frank Act focused in part on increasing shareholder involvement, linking compensation to executive performance, and optimizing transparency.

Before the Financial Crisis and the enactment of the Dodd–Frank Act that would eventually include provisions to enhance shareholder involvement, proposals for reform included attempts to establish say-on-pay plans requiring a stockholder vote on manager compensation. 27 The United Kingdom implemented a say-on-pay scheme in 2002 28 and Germany began allowing

("The majority of Sarbanes-Oxley is not dedicated to executive compensation issues . . . ."); Roberta Romano, The Sarbanes-Oxley Act and the Making of Quack Corporate Governance, 114 YALE L.J. 1521, 1529, 1538 (2005) (criticizing SOX § 402, which prohibits personal loans to directors and officers).


Contingent Capital

1833

Nonbinding votes on management compensation by shareholders during the annual meeting in 2010.29

Other reform proposals before the enactment of the Dodd–Frank Act focused on linking pay with executive performance. Proposals in this context included tying executive compensation to the entity’s performance by granting stock options,30 back loading executive compensation and tying it to the future performance of the company,31 prohibiting severance pay,32 and granting restricted stock with a mandatory holding period.33 Other proposals to increase executive performance suggest a reduction in equity compensation and bonuses caused by industry-based movements and changes in the economy,34 awarding bonuses only for accounting improvements that are sustained over time,35 and curtailing “soft-landing” arrangements.36

Another major focus of reform proposals before the enactment of the Dodd–Frank Act had been transparency in the disclosure of executives’ compensation packages. Proposals to improve transparency included placing a dollar value on all


32. Id.

33. Bhagat & Romano, supra note 1, at 363; Samuelson & Stout, supra note 1, at A13.

34. Bebchuk & Fried, supra note 1, at 669.

35. Id. at 670.

36. See id. at 671–72 (stating that these arrangements “provide generous compensation for executives being pushed out due to failure” and narrow “the payoff gap between good and poor performance”).
elements of executive compensation and disclosing the value in SEC filings, disclosing sales of equity instruments, disclosing nondeductible compensation, appointing a compensation representative to represent shareholder interests in setting executive pay, and appointing a “high-quality” compensation committee consisting of experienced, independent members.

B. Post Dodd–Frank Act

Congress passed the Dodd–Frank Act as a response to the Financial Crisis. The Act adopted many of the suggestions from the precrisis literature on executive compensation and added other safeguards. The Dodd–Frank Act’s provisions pertaining to executive compensation provide for a nonbinding shareholder vote to approve executive compensation, disclosure of relationship between executive compensation and the financial

37. Posner, supra note 1, at 1045.
38. Bebchuk & Fried, supra note 1, at 668.
39. Id. at 668–69.
40. Id. at 668.
44. Id. § 951.
performance of the entity, disclosure of the annual total compensation of the CEO and the relationship to the median annual compensation of employees, disclosure of hedging behavior, claw-back provisions, an independent compensation committee, and prohibition of compensation arrangements that encourage inappropriate risks or could lead to material financial loss to the financial institution.

The post-Dodd–Frank Act-reform debate is predominated by dissension over say-on-pay provisions. Critics argue that shareholder voting on executive compensation could actually hurt shareholders because it diffuses responsibility regarding compensation and insulates directors’ reputations. With say-on-pay provisions in place, directors may be incentivized to authorize larger compensation packages that are less sensitive to performance. Moreover, shareholders may lack the incentives and resources to evaluate the information and may not be able to determine whether executive pay is reasonable. Say-on-pay provisions could give proxy advisory firms more power.

45. Id. § 953(a)(i).
46. Id. § 953(b).
47. Id. § 955.
48. Id. § 954(b).
49. Id. § 952.
50. Id.
51. Id. § 956(a)(1)(B).
54. See id. (proposing an opt-out of the say-on-pay regime by shareholder vote).
56. Gordon, supra note 26, at 325.
To remedy the shortcomings, some suggest that Congress give shareholders a right to opt into a binding vote on the board’s pay scheme, while others suggest that shareholders should have a right to decide whether a public firm should schedule a vote on executive compensation. Information disclosed should include details on the pay of executives at competitor companies, and shareholders should have a right to decide whether a public firm should schedule a vote on executive compensation.

Despite its many critics, say-on-pay may have led to improvements. As a result of say-on-pay requirements, some firms may have reduced compensation and increased performance measures for executive compensation. Because poorly performing companies with high pay levels can expect shareholder dissent, say-on-pay may attract strong shareholder support. Shareholders may perceive compensation procedures as fairer under say-on-pay, which could increase shareholder confidence in an entity’s board of directors and increased investor interest in the entity.


58. Gordon, supra note 26, at 326 (arguing that this regime would focus attention on firms with the most questionable practices).

59. Roddenberry, supra note 55, at 934, 952.

60. Gordon, supra note 26, at 326 (arguing that this regime would focus attention on firms with the most questionable practices, enabling successful implementation to be observed by similar firms and possibly causing them to change their behavior).

61. For a summary of criticism on say-on-pay provisions in the Dodd–Frank Act, see Roddenberry, supra note 55, at 940; Bethune, supra note 57, at 610–14.


63. See id. ("In this paper we find affected firms reduce their compensation, with that decrease being greater for firms that overpaid their CEOs in prior periods. We also find evidence that they increased their use of performance-based compensation.").

Executive compensation in the United States mostly takes the form of equity-based instruments. Changes to the tax code and other regulation encourage equity-based compensation for executives.\textsuperscript{65} The justification of executive compensation with equity-based products has been a major focus of theoretical research on executive compensation in the past three decades.\textsuperscript{66} However, an increasing part of the literature is now considering the role of debt for manager incentives,\textsuperscript{67} and empirical studies show that risk-taking can decline if executives hold more debt relative to their equity holdings.\textsuperscript{68}

Bebchuk and Spaman show that a compensation package that includes a basket of securities representing a predefined percentage of the aggregate value of all outstanding bonds, preferred shares, and common shares could help address the shortcomings of existing executive compensation practices.\textsuperscript{69} Tying executive compensation to this basket of securities issued by either the bank holding company or the bank\textsuperscript{70} can improve incentives for executives to

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{66} See generally Jensen & Murphy, supra note 7 (showing that it is more important to focus on the form of executive compensation than the amount of compensation).
\item \textsuperscript{67} See generally Bebchuk & Spaman, supra note 1; Tung, supra note 9; Hill & Painter, supra note 9.
\item \textsuperscript{68} See generally Tung & Wang, supra note 8; Wei & Yermack, supra note 8; Sundaram & Yermack, supra note 8, at 1553; see also Gerakos, supra note 8, at 23 (finding that pension benefits may reduce risk-taking).
\item \textsuperscript{69} See Bebchuk & Spaman, supra note 1, at 253, 283–84. How the securities in the basket should be weighted is unclear. Weighting debt securities, including contingent convertible bonds, heavily in the calibration of executive compensation may increase the positive effects of debt in executive compensation packages. At the same time, debt may not be the preferred form of compensation for executives. Calibrating executive compensation packages to account for desired incentives and governance improvements while giving sufficient incentives for executives to perform within expected parameters could require an institution specific relational approach and a learning process for institutions. See discussion supra Part IV (reviewing the benefits of the incomplete contract theory of New Institutional Economics (NIE)).
\item \textsuperscript{70} Id.
\end{itemize}
\end{footnotesize}
consider the losses that risk-taking could impose on shareholders, bondholders, depositors, and taxpayers. Regulators could place constraints on the pay schemes to shape how executives choose the actions that are allowed by direct regulation. Moreover, bonuses could be based not only on earnings per share, but rather on broader metrics that also reflect the interests of preferred shareholders, bondholders, and the government as guarantor of deposits. Hill and Painter suggest mandatory partnership or joint venture agreements and assessable stock to ensure that bankers have some personal liability. This personal liability could improve creditor protection because executives would be exposed to some downside risk and would be disincentivized from taking excessive risks.

The literature on inside debt similarly emphasizes creditor protection. Inside debt in the form of deferred compensation and pension plans, among other instruments, can help optimize managers’ incentives and serves an important function in the calibration of executive compensation packages. Fred Tung suggests using public subordinated debt securities for part of the compensation of bank executives. Debt and equity hold different risk preferences, and creditors’ preferences for more conservative management strategies can help curb managers’ risk-taking.

71. Id. at 247, 253, 283–84.
72. Id. at 253.
73. Id.
74. Hill & Painter, supra note 9, at 1174–75.
75. Id.
76. See Sundaram & Yermack, supra note 8, at 1552 (arguing that inside debt alters managerial incentives). This in turn alters the size of the firm’s payouts, the composition of these payouts (dividends versus share repurchases), the firm’s cost of debt and its capital structure, the choice of new securities to be issued (debt versus equity), project choice, capital expenditure choice, and the incentive to pursue diversifying mergers, among many other things; and discussing whether and under what conditions such debt holdings could be part of an optimal compensation package. Id.
77. Tung, supra note 9.
Inside debt in the form of defined benefit plans can make executives sensitive to firm value in bankruptcy, which is desired by creditors. Measured by an entity’s distance from default, aligning managers’ interests with creditors’ can reduce a firm’s risk of defaulting and improve its credit rating. Debt should be part of executive compensation because it is an efficient deterrent against risk-shifting.


80. Sundaram & Yermack, supra note 8.

81. Gerakos, supra note 8, at 11–12.

82. Alex Edmans & Qi Liu, Inside Debt, 15 REV. FIN. 75, 77–78, 91 (2011) (justifying the use of debt as efficient compensation and arguing that a debt bias can improve effort as well as deter risk-shifting); see also Riccardo Calcagno & Luc Renneboog, The Incentive to Give Incentives: On the Relative Seniority of Debt Claims and Managerial Compensation, 31 J. BANKING & FIN. 1795, 1795, 1809 (2007) (arguing that the increase in the leverage of Anglo-American corporations has stimulated the interest in the role of debt as a direct incentive device for management to generate stronger corporate performance); Tung & Wang, supra note 8, at 5 (“Our empirical evidence provides a rationale for the use of inside debt compensation in structuring executive compensation in the banking context.”); Cassell et al., supra note 8, at 588 (stating that “CEO inside debt holdings . . . are generally unsecured and unfunded liabilities of the firm,” and therefore expose “the CEO to default risk similar to that faced by outside creditors,” and arguing that CEOs with large inside debt holdings will display lower levels of risk-seeking behavior); Sallie Krawcheck, Four Ways to Fix Banks, 90 HARV. BUS. REV. 106, 109 (2012) (suggesting paying top executives with debt instead of equity-based compensation to give them more incentive to worry about risk); Hernan Ortiz-Molina, Executive Compensation and Capital Structure: The Effects of Convertible Debt and Straight Debt on CEO Pay, 43 J. ACCT. & ECON. 69, 71 (2007) (arguing that the hypothesis that debt reduces manager–shareholder conflicts can explain some but not all of the results); Alex Edmans, How to Fix Executive Compensation: For Starters, Don’t Link Pay Packages Just to Stock; Tie Them to Debt as Well, WALL ST. J., Feb. 27, 2012, at R.1 (advocating for the inclusion of debt in executive compensation packages in order to reduce risk-taking); Bolton, Mehran & Shapiro, supra note 19, at 34 (presenting evidence that the market believes that “including debtlike instruments in CEO compensation packages will reduce risk” for financial institutions); see generally Yair Listokin, Paying for Performance in Bankruptcy: Why CEOs Should Be Compensated with Debt, 155 U. PA. L. REV. 777 (2007) (proposing a novel bankruptcy compensation plan, otherwise known as debt
The creditor-centered approach to executive compensation has encountered some critics who argue that inside debt can be inefficient and may not influence executives’ conduct sufficiently while creating complicated incentive structures.83 Other criticisms include allegations of too strong a focus on the banking sector and managers’ ability to manipulate book value as the only available measure of asset value.84 Critics have also attacked the creditor-centered approach on methodological grounds because they assume that the principal–agent relationship in that approach is a one-shot transaction rather than a relational contract.85 Given the shortcomings pointed out by the debate, some call for a paradigm shift to overcome the unnecessary creation of new remuneration narratives.86

III. Contingent Capital

Contingent convertible bonds (CCBs) are debt securities that can either be written down or converted into equity upon a triggering event.87 The many applications and benefits of contingent convertible bonds88 include their ability to stabilize and prepare

compensation, that is expected to provide better incentives for CEOs to perform efficiently).


84. Sepe, supra note 1, at 210–11.

85. Id. at 193, 211–12.


88. See Kaal, supra note 14 (summarizing the benefits of contingent convertible bonds).
SIFIs for future financial crises, signal default risk prevent bailouts, decrease risk-taking, minimize moral hazard, incentivize the increase in capital, internalize bank failure


92. Pennacchi et al., supra note 13, at 36; Dudley, supra note 90 (asserting that because bank difficulties would trigger conversion, the dilution of shareholders creates an incentive for bank managers to “manage not only for good outcomes on the upside of the boom, but also against bad outcomes on the downside”).

93. See Flannery, No Pain, supra note 89, at 15 (“Frequent trigger evaluations eliminate moral hazard incentives and . . . [result in] surprisingly low default risk.”).

94. See Squam Lake Working Grp., supra note 91, at 3–4 (arguing that conversion of contingent capital bonds would quickly recapitalize banks); Calomiris & Herring supra note 91, at 39 (“A proper CoCos requirement can provide strong incentives for the prompt recapitalization of banks after significant losses of equity.”).
cost,95 avoid financial contagion,96 and “limit systemic risk.”97 As a hybrid instrument, contingent convertible bonds combine the limited upside of debt in the form of the coupon rate with the unlimited downside risk of equity, i.e., the total loss of the investment. Several successful contingent convertible bond issuances in Europe with coupon rates between 7% and 9.5% show that these securities can display a combination of features that investors and issuers find attractive.98 The market in contingent convertible bonds is slowly evolving.99 Because contingent convertible debt has many applications and could help reform policy in many areas, the concept finds increasing support among academics100 and policy


96. See GOLDMAN SACHS TBTF, supra note 12, at 6 (noting that if the appropriate triggers are in place, it could prevent bank runs—though if the trigger is based on market prices, it could worsen bank runs); see also Darrell Duffie, Contractual Methods for Out-Of-Court Restructuring of Systemically Important Financial Institutions 5 (Hoover Inst., Working Papers on Econ. Pol’y, 2009), available at http://media.hoover.org/sites/default/files/documents/06EndingGovernmentBailoutsAsWeKnowThemDuffie.pdf (describing how conversion of contingent capital debt could “forestall . . . a liquidity crisis”).

97. Kaal, supra note 14, at 106; see also Coffee, supra note 9, at 806 (proposing use of contingent capital securities to reduce “pressure on corporate managers to accept greater risk and leverage”).


99. Id. at 136 (acknowledging a “lack of regulatory guidance”).

100. See id.; DAVID SKEEL, THE NEW FINANCIAL DEAL: UNDERSTANDING THE DODD–FRANK ACT AND ITS (UNINTENDED) CONSEQUENCES 84–85 (2011) (noting that the Dodd–Frank Act instructs the General Accountability Office to conduct a study on contingent capital and to begin using it when the study is completed); Coffee, supra note 9, at 839; Kaal & Henkel, supra note 14; Pennacchi et al., supra note 13, at 36; Douglas W. Diamond & Raghuram G. Rajan, Fear of Fire Sales and the Credit Freeze 28 (Bank for Int’l Settlements, Working Paper No. 305, 2010), http://www.bis.org/publ/work305.pdf (“[C]ontingent capital is like installing sprinklers . . . . [W]hen the fire threatens, the sprinklers will turn on.”); Flannery, No Pain, supra note 89, at 15. But see Christian Koziol & Jochen Lawrenz, Contingent Convertibles: Solving or Seeding the Next Banking Crisis?, 36 J. BANKING & FIN. 90, 91 (2012) (explaining that in situations involving incomplete contracts, contingent convertible bonds may subject banks to greater financial distress); Duffie, supra note 96, at 5 (stating that contingent convertible bonds are “unlikely to stop a [bank’s] liquidity crisis once it begins”); McDonald, supra note 95, at 20–21 (describing situations where contingent
Contingent convertible triggering events for conversion from debt into equity are typically intended to avert a financial weakening of the entity. The automatic conversion from debt into equity helps increase capital when needed and lowers the debt-to-equity ratio. The automatic conversion of debt into equity may prove especially attractive to SIFIs who could otherwise be forced into restructuring. Because of the importance of the conversion feature of contingent convertible bonds for purposes of corporate governance improvements, the analysis in this Article will focus on

capital fails to convert).


102. Coffee, supra note 9, at 805; Duffie, supra note 96, at 4–5; Flannery, Stabilizing, supra note 89, at 3.

103. See Amar Bhidé, A Call for Judgment: Sensible Finance for a Dynamic Economy 291 (2010) (offering a proposal aimed at reducing risk-taking by amending current financial regulations to allow unregulated financial institutions to raise debt or equity, “but not on a short-term basis from the public or regulated fiduciaries”); Coffee, supra note 9, at 805 (averring that contingent capital can counter-leverage debt).

104. See Kaal & Henkel, supra note 14, at 234 (stating that “conversion of debt into equity could be an attractive alternative” for struggling financial institutions); Coffee, supra note 9, at 805 ("By definition, such a conversion [of contingent capital debt] averts ... bankruptcy .... ").
contingent convertible bonds with a conversion feature rather than a write-down.

Contingent convertible bonds have quasi-public good characteristics and are particularly suitable for purposes of corporate governance improvements in SIFIs. The conversion feature of these securities may have the ability to influence corporate governance in SIFIs. If sufficient volumes of contingent convertible bond issuances are combined with adequate design features, the conversion feature of contingent convertible bonds and the threat of dilution of equity positions in SIFIs could affect corporate governance in SIFIs. Measures to increase the effectiveness of the conversion feature of contingent capital, such as increased voting rights or sequential triggers, could further increase the impact these securities may have on corporate governance.

In part because of mounting pressure from politicians, policy makers, and legislators who demanded remedies for corporate governance shortcomings, European SIFIs have issued contingent convertible bonds. Contingent convertible bonds’ potential to address corporate governance shortcomings, albeit not fully utilized in the existing designs, may help with public relations because a contingent convertible bond issuance may signal to investors, politicians, and the general public that SIFI management is instituting safety-increasing measures that can help avoid future bailouts. Although the designs of recent contingent convertible bond issuances provide mostly for a write-down feature rather than a conversion to equity, the SIFIs who issued contingent convertible bonds seem to have recognized the market acceptance and investor demand for these hybrid securities. The market in

105. Kaal, supra note 14, at 140.
106. Id. at 140, 146.
107. Id. at 144.
108. Id. at 145.
109. See id. at 144–46 (outlining how contingent capital debt could be utilized to increase voting rights); Kaal & Henkel, supra note 14, at 255–56 (explaining how sequential triggers could “increase voting rights for holders of contingent capital”).
110. Kaal, supra note 14, at 125–26 (summarizing Swiss efforts to implement contingent capital rules and how these efforts precipitated Credit Suisse’s voluntary issuances of contingent convertible bonds); see also Goldman Sachs, supra note 12, at 18 (comparing the “few ‘loss absorbing’ securities . . . issued in the European market”).
contingent convertible bonds, however, is still in its infancy, and privately negotiated contingent convertible bond sales have so far not resulted in efficiently functioning contingent convertible bond designs that take corporate governance considerations into account. It is doubtful if market solutions and private ordering alone will produce contingent capital designs that help improve corporate governance in SIFIs.

IV. Incomplete Contract Theory

This Article expands the existing literature on the creditor-centered approach to executive compensation by recommending the inclusion of contingent convertible bonds in the calibration of executives’ compensation packages. Scholarly contributions in the context of the creditor-centered approach to executive compensation are often (implicitly) based on the classical contract model or the spot contract model. This can result in suboptimal and unrealistic outcomes. The classical contract model assumes a system of rules that deals with and legally guarantees all future eventualities. The parties to a contract negotiate and agree *ex ante* on all possible scenarios and eventually execute the contract as agreed. The contract in this model hopes to anticipate all eventualities and is not intended to be ambiguous. The model assumes that the comprehensive nature of the regulation leaves no discretion to the agents, making opportunistic behavior impossible. Transaction costs in the form of initiating, concluding

---

111. Many scholars who endorse the creditor-centered approach to executive compensation may implicitly use the classical- or spot-contract model of executive compensation. See, e.g., Bebchuk & Spamann, *supra* note 1, at 255–56 (examining banks’ incentive for risk-taking using only two periods: “the present, when managers make decisions, and the future, when gains or losses are realized and the manager gets paid” because using multiple periods for the analysis unnecessarily complicates the analysis without changes in the conclusions or other substantial benefits); Gordon, *supra* note 26, at 332–39 (explaining that shareholders now have more corporate governance power to push for “pay for performance” executive compensation); Hill & Painter, *supra* note 9, at 1175, 1186–95; Sepe, *supra* note 1, at 211–12 (summarizing the issues).

and enforcing the contract are not considered in the classical contract model.

Contrary to the assumptions of the classical contract model and the spot contract model, executive contracts have an informal relational dimension that goes beyond the written contract between the parties. The relationship between the executive as agent and the corporation and its shareholders as principal involves more than the contractual terms of the agreement. An executive compensation contract combines elements of a knowledge exchange and relational collaboration. Informal relational behavior between the principal and agent generates confidence in the relationship, creates institution-specific knowledge, and provides innovative capabilities. The informal relational elements of executive compensation contracts can often overshadow the legal terms of the agreement. Using a model for the analysis of executive contracts that is based on the single contract between the executive–agent and corporation–principal would ignore the informal relational element of this principal–agent relationship. Given the importance of the informal relational element of executive compensation contracts, the literature on the debt-centered approach to executive compensation should include informal relational-, behavioral-, and incomplete-contract theories.

These shortcomings in the analysis of executive compensation contracts under the classical contract model and spot contract model can be overcome with the relational or incomplete contract model in New Institutional Economics (NIE). NIE is a relatively young offspring of economic theory and shares core assumptions with the neoclassical model, such as


114. NIE’s assumptions are increasingly used in modern economic analysis of financial markets and financial rules. See, e.g., Hersh Shefrin, Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing 9–10 (2002) (pointing out that studying “behavioral phenomena” is integral to understanding inefficient markets); Andrei Shleifer, Inefficient Markets: An Introduction to Behavioral Finance 10 (2000) (stressing that investors are often irrational, that they “deviate from the standard decision making model in a number of fundamental areas,” and that they do not behave as would be expected under traditional theories that rely on efficient markets).
methodological individualism, scarcity of resources, and self-interested rational behavior. NIE, however, substitutes the assumption of full rationality with bounded rationality and opportunistic behavior, and underscores that information is systematically incomplete. NIE emphasizes the functioning, development, and improvement of institutions. Institutions are defined as general rules or sets of general rules, together with their enforcement mechanisms. NIE emphasizes the importance of informal institutions, such as social norms. Because corporate governance issues often involve formal and informal institutions, NIE is ideally suited to examine the efficiency of governance structures. Experimentation, observation, and rule revision in NIE's model are part of a


117. Id. at 88–89; FURUBOTN & RICHTER, supra note 115, at 5.


120. See FURUBOTN & RICHTER, supra note 115, at 7 (“[A]n institution will be defined as a set of formal and informal rules, including their enforcement arrangements.”); VOIGT, supra note 116.

121. See VOIGT, supra note 116 (recognizing that limiting the analysis to a subset of formal institutions would ignore important problems).
continuous process that avoids “optimal” or “stable” rules.122 Experimentation can result in a learning process that can improve the quality of governance structures. Because it is likely that underlying economic conditions will change, contingent convertible bond offerings to the public and to executives may benefit from a process of experimentation and learning.

In contrast to classical contract analysis, NIE’s incomplete contract model recognizes that contracts are inevitably incomplete and rely on control rights to minimize opportunistic behavior.123 This model takes opportunistic behavior and transaction costs into account.124 The degree of contractual incompleteness is influenced by the cost of contracting and contracting parties’ ability to anticipate opportunistic behavior.125 The theory of incomplete contracts under NIE can be considered part of the principal–agent approach because information before and after contracting is asymmetric and the agent has a certain amount of discretion making opportunistic behavior possible.126

At the core of the principal–agent relationship in executive compensation is the short-term interest of the manager/agent to generate a high income that conflicts with the long-term ownership interest of the shareholders/principals. The relational elements in executive compensation contracts may further increase the principal–agent problem. The combination of knowledge exchange and relational collaboration in executive compensation contracts makes opportunistic behavior of executives likely.127 Contracting parties are, however, limited in

122. See Kaal, supra note 14, at 137–39 (explaining the benefits of experimentation for the evolution of contingent capital rules).


124. See Williamson, supra note 119, at 30 (defining opportunism as “self-interest seeking with guile”); see also Furubotn & Richter, supra note 115.


126. See generally Furubotn & Richter, supra note 115.

127. Short-termism to maximize personal income through stock options, the Fuld Problem (see infra Part VI.B), and income inequality between senior
their ability to anticipate opportunistic behavior of agents. Therefore, control rights to limit opportunistic behavior become increasingly important. Because of incomplete information, information asymmetries in the principal–agent relationship, bounded rationality of contracting parties, the parties’ limited cognition and foresight, and transaction costs, control rights in executive compensation contracts cannot account sufficiently for ex post opportunism of agents. Adding contingent convertible bonds with an early trigger to executive compensation packages can create a corporate governance mechanism that helps address these shortcomings.128

Non-contractual behavior in the form of non-contractual norms, reciprocity, trust, friendship, reputation, altruism, interdependence, and moral obligations may be unenforceable through contractual agreements, but it can shape economic action.129 Non-contractual behaviors can harmonize conflicts and

executives and the rest of the workforce are only a few examples of opportunistic behavior of agents in the context of executive compensation.

128. See infra Part V.B.2 (discussing the design of early triggers in executive compensation).

help sustain relationships. Therefore, non-contractual behavior should be considered in the attempt to optimize manager incentives for corporate governance improvements.

V. Contingent Capital in Executive Compensation

The academic literature on executive compensation is largely silent on the use of contingent convertible bonds in executive compensation. Similarly, the literature on contingent capital has mostly ignored the possible application of contingent convertible bonds in executive compensation. The governance-improving features of contingent convertible bonds can be applied to executive compensation. Contingent convertible bonds in executive compensation can help address the core executive compensation issues that emerged after the global Financial Crisis. More specifically, contingent convertible bonds in executive compensation packages can improve suboptimal incentives for risk-taking by executives, the alignment of executives’ interests with those of different constituents, sustainability, and income inequality. This Article proposes the use of contingent convertible bonds with early triggers in executive compensation packages.

130. Powell, supra note 129, at 303.
132. See Kaal, supra note 14, at 295–96.
133. See Hill, supra note 22, at 24–31 (summarizing the core themes in recent proposals concerning executive compensation, including income inequality, incentive optimization, interest alignment, and sustainability).
A. Precedent Barclays

The proposal in this Article to use contingent capital bonds with early triggers in executive compensation is not a mere theoretical construct. European SIFIs have started to add contingent convertible bonds to executive compensation packages. The English bank Barclays has issued contingent capital securities to its executives as deferred incentive awards. The contingent convertible bonds issued under Barclays’s Contingent Capital Plan (CCP), however, would not be written down or converted into equity like other CCBs. Under Barclays’s CCP, its “synthetic CoCos” simply lapse when the capital ratio falls below 7%. More specifically, if Barclays’s Group Core Tier 1 capital ratio falls below the threshold, the executives will receive no coupon payment and the Contingent Capital Award (CCA) will

134. See Rob Cox, At Barclays, a Pay System That May Please, N.Y. TIMES, Dec. 6, 2010, at B2 (“CoCos would not merely constitute a compensation fig leaf. Throwing the securities into bankers’ stockings better aligns their interests with those of regulators hoping to avoid a repeat of the taxpayer bailouts of the last financial crisis.”); BARCLAYS ANNUAL REPORT, supra note 17, at 167 (“[D]eferred incentive awards for 2010 are made under the Share Value Plan (SVP) in the form of Barclays shares and under the Contingent Capital Plan (CCP) in the form of contingent capital awards.”).

135. E-mail from Mark Lane, Dir. of Corporate Comm’n, Barclays Capital, to Wulf Kaal, Assoc. Professor of Law, Univ. of St. Thomas Sch. of Law (Dec. 14, 2011, 3:52 PM CST) (on file with author and with the Washington and Lee Law Review); see also Jill Treanor, City Resists Barclays Chiefs’ ‘CoCo’ Bonuses: Bank Plans New Bonds to Boost Top Executives’ Pay, GUARDIAN Mar. 19, 2011, at 50 (“The cocos Barclays intends to use to pay its staff do not convert into equity, however, but merely fall away once the bank’s capital ratio falls below 7%.”); Tommy Wilkes & Sinead Cruise, Barclays Heads for Investor Clash over Pay, REUTERS Apr. 26, 2011, http://in.mobile.reuters.com/article/rbssFinancialServicesAndRealEstateNews/idINLDE73K0IP20110426?rpc=984 (last visited Nov. 14, 2012) (on file with the Washington and Lee Law Review); Jill Treanor, Barclays Faces Shareholder Anger over Bob Diamond’s Pay, GUARDIAN Apr. 20, 2011, at 28 (“Cocos are a new type of financial instrument that can convert into equity during times of severe stress and have been issued by a handful of banks to raise fresh capital from investors.”) “Barclays, though, intends to issue the cocos only to its staff. The Barclays cocos will not convert into equity but merely fall away once the bank’s capital ratio falls below 7%—which is why they are being called synthetic cocos.” Id.; see also Editorial, Bankers and Their Bonuses, N.Y. TIMES, Feb. 6, 2011, at WK.7 (“Cocos are long-term bonds that convert into equity if the bank hits a crisis. The idea is that paying bankers in bonds encourages them to keep the business solvent. This is even more so if a crisis triggers their conversion into shares that would become worthless in bankruptcy.”); BARCLAYS ANNUAL REPORT, supra note 17, at 182.
remain unvested. However, the unvested portion of the CCA can be released and paid out to the executives if after six months the Group Core Tier 1 capital ratio has recovered. In order to take into account the possible effect of any actions taken to address the shortfall in the capital ratio on shareholders, the release may be adjusted. No coupon will be awarded if there was a downward adjustment. Should the Group Core Tier 1 capital ratio not recover to above 7% five years after the suspension of the CCA, the CCA will lapse.

Barclays’s contingent capital award to executives without a triggering event into equity and a mere lapse is beneficial because it underscores the possible use of contingent convertible bonds in executive compensation. Given European proposals on the use of contingent convertible bonds to make SIFIs safer and avoid bailouts, Barclays’s contingent convertible bond issuance to executives may also signal its willingness to consider the

---

136. BARCLAYS ANNUAL REPORT, supra note 17, at 182.
138. Id.
139. Id.
possible uses of contingent convertible bonds to prepare the entity for future crises and systemic shocks. As the first mover in this context, it is also understandable that Barclays has decided to avoid the possible governance shakeup that could be associated with a contingent convertible bond issuance that provides a conversion feature.

Barclays’s issuance is commendable. Its CCA, as a deferred compensation award, is inside debt. Unlike inside debt in the form of traditional bonds, however, it does not create a fixed claim for managers with a stake in the firm’s liquidation value because it falls away when converted. Barclays’s CCA, therefore, also does not lower agency cost. Without a stake in the liquidation value of the SIFI, executives may also not limit their risk-taking. Barclays’s executives’ interests are aligned with their debt-holders’ because executives are now also debt-holders, albeit in a separate class with substantially higher coupon payments. However, because the CCA falls away upon conversion, the executives’ interests are still predominantly aligned with shareholders’ through equity-based compensation, not through contingent convertible bond holdings.

Barclays’s issuance of contingent convertible bonds without a conversion feature to its executives shows limited governance improvements. Without a conversion to equity, Barclays provides only limited incentives for its executives to lower risk-taking. In its current form, the CCA seems to be a mere compensation supplement for executives.

141. Contrast this outcome with the proposal in this Article to use contingent convertible bonds with early triggers in executive compensation. See infra Part V.B.2 (showing that contingent convertible bonds with early triggers can lower agency costs more than traditional inside debt, provide greater incentives to lower risk-taking, align executives’ interests with the interests of both debt-holders and shareholders, and lower income inequality).

142. See infra Part V.B.2 (explaining how contingent convertible bonds in executive compensation align the interests of executives with debt-holders’ interests before conversion while increasing incentives for sustainability and lowering risk-taking and income inequality). Conversely, upon conversion into equity, CCBs align executives’ interests with the interests of equity-holders when it is most needed and beneficial for the SIFI, that is, early before the entity becomes insolvent. Id.
B. Design of Contingent Convertible Bonds in Executive Compensation

Contingent convertible bonds issued to investors can have corporate governance implications as a result of the conversion feature and the threat of equity dilution following a conversion. Issuing contingent convertible bonds to executives as part of their compensation can have different governance implications. Perhaps most important for this Article is contingent capital's ability to generate powerful incentives for SIFI managers and thereby lower their risk-taking. This applies to contingent convertible bond issuances to investors but also, and more importantly, to the issuance of contingent capital to executives as part of their compensation.

Designing contingent convertible bonds in executive compensation to be similar to contingent convertible bonds issued to investors could result in suboptimal outcomes. Unlike contingent convertible bonds issued to investors, contingent convertible bonds may be issued to executives in volumes that may not suffice to dilute investors' equity holdings. The lower volume of contingent convertible bonds issued to executives may not provide a sufficiently strong equity infusion during a crisis. Without a design adjustment, contingent convertible bonds held by executives may not play a significant role in preparing SIFIs for future financial crises. Contingent convertible bonds in executive compensation should not be treated like other contingent capital securities. The design adjustment proposed in this Article helps to optimize the effectiveness of contingent convertible bonds in executive compensation and can improve corporate governance.

143. See Kaal, supra note 14.

144. Id.; Coffee, supra note 9, at 806; see also Dudley, supra note 90 (“If the bank encounters difficulties triggering conversion, shareholders would be automatically and immediately diluted. This would create strong incentives for bank managements to manage not only for good outcomes on the upside of the boom, but also against bad outcomes on the downside.”).

145. See Kaal, supra note 14.
1. Automatic Institution-Specific “Early” Trigger

The literature on contingent capital trigger designs focuses on the efficient calibration of triggering events. The efficient calibration of triggering events is central to the design of contingent capital because the trigger affects if and when the conversion takes place. The timing of conversion is crucial for possible corporate governance improvements.

The early trigger for contingent convertible bonds held by executives serves a different purpose than the trigger for contingent convertible bonds held by investors. The early trigger converts only the portion of executives' debt to equity, before investors' contingent convertible bonds are converted, when the entity is still sound on a micro-prudential basis. The purpose of an “early” or “strong” trigger design for CCBs held by executives is to establish an early warning system that is independent of

146. See, e.g., Coffee, supra note 9, at 806 (stating that contingent capital can help prevent the risk of the first step in a financial crisis); McDonald, supra note 95, at 2 (arguing for a conversion triggered by market prices); Pennacchi et al., supra note 13, at 7 (stating that CoCos can be effective if they are designed to convert prior to severe financial stress); Squam Lake Working Grp., supra note 91, at 4 (arguing for a long-term debt instrument that would convert to equity before a crisis); Suresh Sundaresan & Zhenyu Wang, On the Design of Contingent Capital with Market Trigger 4 (Fed. Reserve Bank of N.Y., Staff Rep. No. 448, Nov. 2011), http://www.newyorkfed.org/research/staff_reports/sr448.pdf (explaining that a market trigger for conversion can provide stability to banks and markets); see also Duffie, supra note 96, at 5 (arguing that the trigger needs to be set so as to eliminate debt claims before a liquidity crisis); Flannery, Stabilizing, supra note 89; Flannery, No Pain, supra note 89, at 1 (introducing “reverse convertible debentures” that would convert if the issuing firm’s capital level fell below a prespecified level); Paul Glasserman & Behzad Nouri, Contingent Capital with a Capital-Ratio Trigger 2–3 (Working Paper, 2010), available at http://ssrn.com/abstract=1669686 (examining conversion based on a capital-ratio trigger); Ceyla Pazarbasioglu et al., Contingent Capital: Economic Rationale and Design Features 18 (IMF Staff Discussion Note No. SDN/11/01, 2011), http://www.imf.org/external/pubs/ft/sdn/2011/sdn1101.pdf (arguing that contingent capital instruments should be considered part of a crisis prevention and management framework); Goldman Sachs, supra note 12, at 4 (stating that contingent capital is designed to operate before resolution mechanisms become involved); Swedish Ministry of Finance et al., Swedish Answers to the DG Internal Market and Services Working Document “Technical Details of A Possible EU Framework for Bank Recovery and Resolution” 43 (2011) [hereinafter Swedish Ministry of Finance], http://www.riksbank.se/Upload/Dokument_riksbank/Kat_publicerat/Remisser2011/Consultation_03031_1.pdf (arguing that bail-in tools may be used to minimize systemic risk).

147. The early warning nature of the early trigger design and the fact that
the capitalization needs of the entity and to provide corporate governance improvements. To ensure that the early warning objective of the early trigger is accomplished, it is important to distinguish between the two general categories of triggers discussed in the literature: (i) regulatory triggers and (ii) transactional and automatic triggers.

Regulatory triggers give regulators the authority to decide when to convert the contingent convertible bonds. The regulatory trigger may depend on a regulator’s determination that the respective bank is not viable without a public sector injection of capital or a write-off. It can also be based on the evaluation of a bank during a stress test conducted by regulators. Regulatory triggers may lead to market uncertainty and ad hoc decisions by regulators and result in adverse market responses. Because regulatory triggers generate the highest level of uncertainty, they may not be the best option for the design of contingent convertible bonds in executive compensation. Regulatory discretion in triggering the conversion could create unfavorable market movements against the entity. While a regulator may decide to trigger executives’ CCBs as an early warning sign in a pending crisis, the cost of supervision could be prohibitive and regulatory discretion could alienate managers. Similarly, even though a regulatory trigger could help avoid abuse by only the executives’ portion of CCBs gets converted to equity could make it easier for the SIFI to negotiate bridge loans because the entity is still sound on a micro-prudential basis. If investors’ CCBs get triggered, some lenders may be unwilling to provide bridge loans. In effect, the early trigger could help avoid lenders turning away from the company when a large portion of its debt is converted into equity.

148. The benefit of providing additional capital when needed derives predominantly from the conversion of contingent convertible bonds issued to investors, not the portion of contingent convertible bonds issued to executives.


150. See GOLDMAN SACHS, supra note 12, at 8 (describing the attributes of a conversion trigger based on regulatory discretion).

151. Id.

152. See Henkel & Kaal, supra note 149, at 255 (“[A regulatory systemic trigger] can be a trigger that converts CCS into equity upon, for instance, a regulator’s decision that additional capital is needed.”).
executives, regulatory triggers may insufficiently incentivize executives to lower risk because the executives would not have to self-monitor and adjust their risk-taking preferences to avoid the trigger.

Transactional triggers, which are also called institution-specific triggers, are privately negotiated terms for triggering events in bond contracts. They have the advantage of being flexible and tailored to the parties’ subjective needs. Automatic triggers are mostly privately negotiated terms in bond contracts that convert debt into equity when a certain capital ratio, stock price, CDS spread, index value, or other trigger is reached. Because institution-specific automatic triggers are flexible and independent from regulatory discretion, they constitute a good option for early trigger designs. However, market-based measures may be susceptible to market manipulation and banking runs. Accounting-based measures in institution-specific automatic triggers are arguably too infrequently updated to respond adequately in a financial crisis.

Early triggers for contingent convertible bonds in executive compensation could yield particular benefits if they convert to equity before the contingent convertible bonds that were issued to investors. Early triggers could be based on various combinations of design features including perhaps already-
existing designs of previously issued contingent convertible bonds to investors. An entity that has issued contingent convertible bonds to investors could simply lower the executives' trigger threshold in contrast with that of the CCBs that were issued to investors.\textsuperscript{159} In the case of automatic institution-specific triggers, lowering the trigger could be accomplished by adjusting the capital ratio, stock price, CDS spread, index value, or other triggering event for the executives’ CCBs to cause the conversion at an earlier point during the financial weakening of the respective entity.\textsuperscript{160}

In the case of contingent convertible bonds with a capital-ratio trigger,\textsuperscript{161} an “early” trigger for executives’ contingent convertible bonds could mean a 10% capital ratio whereas the “late” trigger for the contingent convertible bonds of investors could be around 8%.\textsuperscript{162} Regulatory capital requirements and industry standards for SIFIs or other banks that are categorized

---

\textsuperscript{159} Problems with trigger mechanisms have been discussed at length in the literature. See, e.g., Coffee, supra note 9, at 821, 827–29 (arguing that accounting-based measures may be too infrequently updated to respond effectively in a financial crisis, while regulatory triggers can lead to ad hoc decisions by regulators that result in market uncertainty and adverse market responses); McDonald, supra note 95, at 9–12, 22 (describing how market-based measures may be susceptible to market manipulation and banking runs).

\textsuperscript{160} Although it may be difficult to calibrate the trigger to provide for a specific time period before the conversion of investors’ CCBs, the early trigger should give sufficient warning of the weakening financial condition of the entity.

\textsuperscript{161} The capital ratio of an entity is the percentage of the entity’s capital to its risk-weighted assets. See Glasserman & Nouri, supra note 146, at 1–4 (discussing contingent capital with a capital-ratio trigger and partial and ongoing conversion).

\textsuperscript{162} This is just a numerical example that does not take other factors into account. The purpose of this Article is not to suggest specific design features for “early” triggers. Rather, the purpose here is to show that the nature of ownership can have an impact on trigger designs and their potential to improve corporate governance.
as “well-capitalized” may change over time. To be effective, early triggers should be well above the threshold for capital requirements under Basel III. But early triggers in the form of capital ratios should be independent of regulatory demands pertaining to capitalization levels.

2. The Benefits of “Early” Triggers

Automatic institution-specific early triggers for contingent convertible bonds issued to executives may increase the overall effectiveness of these bonds and improve executive compensation policies. The benefits of early triggers may partially depend on whether the entity issues contingent convertible debt only to executives or to executives and investors. The early trigger for the portion of contingent convertible bonds held by executives can be an early warning system and buffer for contingent convertible bond investors and shareholders. Early triggers in contingent convertible bonds issued to executives provide several advantages: they lower risk-taking by executives and increase shareholder monitoring. They also offer a better signal for default.

163. The recognition of contingent capital as Tier I Capital and the capital adequacy standards under Basel III may also influence the adequacy of an early trigger for contingent convertible bonds held by executives. See Kaal & Henkel, supra note 14, at 240 (examining the recognition of contingent capital as Tier I Capital).

164. Under Basel III, banks will have to hold better quality capital. Banks will have to hold minimum capital representing 8% of risk-weighted assets (RWA) as well as an additional capital buffer of 2.5% of RWA. 7% of RWA must be comprised of Tier 1 common equity. BASEL COMM. ON BANKING SUPERVISION, BANK FOR INT’L SETTLEMENTS, BASEL III: A GLOBAL REGULATORY FRAMEWORK FOR MORE RESILIENT BANKS AND BANKING SYSTEMS 1, 64 (2010), http://www.bis.org/publ/bcbs189.pdf.

165. See supra Part V.

166. Issuing contingent convertible bonds with the same design features to both investors and executives creates a risk of management opportunism if managers are heavily involved in the drafting process. Management opportunism can lead to socially suboptimal designs. See infra Part V.C (elaborating on design features to prevent abuse).

167. Investors would be able to anticipate a possible conversion of their CCBs if the early trigger resulted in a conversion of executives’ CCBs. The warning and signaling function of the early trigger would benefit CCBs investors and shareholders of the entity, which may be diluted if the investors’ CCBs are converted to equity.
risk and align executives’ interests with a more diverse group of constituents.168

Studies have shown that executives who hold more debt relative to their equity holdings take fewer risks when managing an entity.169 Increasing the debt portion of executive

168. These benefits could help make contingent convertible bonds generally more marketable. In the United States, where the tax treatment of these hybrid securities is still unclear, many investors still perceive contingent convertible bonds as a hybrid instrument with few attractive features despite a substantial coupon rate. See Satyajit Das, Investors Must Hope ‘Cocos’ Never Show Their Dark Side, FIN. TIMES, May 1, 2012, at 22 (“For investors, hybrids are a deeply subordinated investment with uncertain income and significant capital risk.”); cf. Liam Vaughan, Investors May Shun Banks’ Contingent Convertibles as Regulator Adds Limits, BLOOMBERG Dec. 13, 2011, http://www.bloomberg.com/news/2011-12-13/investors-may-shun-coco-bonds-approved-by-european-regulators.html (discussing negative investor reactions to contingent convertible bonds issued by European banks). The early trigger design of CCBs issued to executives could help signal the entity’s intent to address default risk and systemic risk, to lower risk incentives, to align executives’ interests with creditor interests, and to increase monitoring and other corporate governance improvements. This could generate increased investor interest and help create a market for contingent convertible bonds in the United States.

169. Edmans & Liu, supra note 82, at 77–78 (explaining how debt compensation aligns managers’ incentives with the interests of the firm’s creditors and emphasizing the importance of a balanced mix of equity and debt in compensation); Gerakos, supra note 8, at 11–12 (finding that pension benefits may reduce risk-taking); Tung & Wang, supra note 8, at 1 (“Theory predicts and empirical evidence supports the claim that as the proportion of CEO wealth held in the form of debt increases relative to CEO equity holdings, risk taking declines . . . .”); Wei & Yermack, supra note 8, at 3–5 (discussing the effect of debt compensation on volatility). Other studies reject the idea that corporate governance and executive compensation are correlated. See Core, Guay & Larcker, supra note 19, at 28 (“[A] number of scholars and practitioners either implicitly or explicitly take the view that contracting arrangements are largely inefficient and do not reduce agency costs.”); Mary Ellen Carter & Luann J. Lynch, An Examination of Executive Stock Option Repricing, 61 J. FIN. ECON. 207, 222 (2001) (finding no relationship between institutional ownership and the repricing decision); Michael B. Dorff, Does One Hand Wash the Other? Testing the Managerial Power and Optimal Contracting Theories of Executive Compensation, 30 J. CORP. L. 255, 268–69 (2004) (describing the Managerial Power Hypothesis and the correlation between executive compensation and performance); Marilyn F. Johnson et al., Stakeholder Pressure and the Structure of Executive Compensation 16–17, 38 (Working Paper. 1997), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=41780 (examining the effect of shareholder proposals on performance); see also Ronald C. Anderson & John M. Bizjak, An Empirical Examination of the Role of the CEO and the Compensation Committee in Structuring Executive Pay, 27 J. BANKING & FIN. 1323, 1324–26 (2003) (discussing the role of compensation committees in determining executive pay); Kam-Ming Wan, Independent Directors, Executive
compensation packages with contingent convertible bonds can help lower managers’ risk incentives.\textsuperscript{170} This effect could be even more pronounced depending on the proportion of CCBs included in the total compensation of executives. However, path dependencies in the executive compensation culture in the United States will likely make it difficult for executive compensation policies to move entirely to contingent convertible bonds or other debt instruments. It is more likely that compensation committees will continue to combine debt instruments with equity-based compensation.

The conversion of contingent convertible bonds from debt to equity, if triggered, would mean that executives hold a highly discounted equity interest in the entity.\textsuperscript{171} This may in turn depress the stock price of the respective entity.\textsuperscript{172} A negative effect on the stock price of the entity may lower the value of the


\textsuperscript{170}. Calcagno & Renneboog, \textit{supra} note 82, at 1796 ("[I]ncrease in the leverage of Anglo-American corporations has also stimulated the interest in the role of debt as a direct incentive device . . . including risky debt in the capital structure changes the ‘incentive to give incentives . . . .’"); Edmans & Liu, \textit{supra} note 82, at 77–78 (justifying the use of debt as efficient compensation and arguing that a debt bias can improve effort as well as deter risk-shifting); Krawcheck, \textit{supra} note 82, at 108–09 (suggesting that CEOs with large inside debt holdings will display lower levels of risk-seeking behavior); Listokin, \textit{supra} note 82 (proposing debt compensation as an incentive for managers during bankruptcy); Ortiz-Molina, \textit{supra} note 82, at 70 ("[T]he agency costs of debt can also be reduced by using convertible debt . . . ."); Edmans, \textit{supra} note 82, at R1 (suggesting debt compensation as a means to reduce risk); Bolton, Mehran & Shapiro, \textit{supra} note 19, at 2 ("[M]arket participants do indeed believe that linking executive compensation to default risk will reduce the riskiness of the firm."); Tung & Wang, \textit{supra} note 8, at 5 ("[O]ur empirical evidence provides a rationale for the use of inside debt compensation in structuring executive compensation in the banking context.").

\textsuperscript{171}. See \textit{infra} note 192 and accompanying text (discussing the potential for abuse).

\textsuperscript{172}. Depending on financial institutions’ implementation of contingent convertible bonds and the evolution of the market in CCBs, the potential effect of CCBs conversion on stock prices may be evaluated in future research. See Sundaresan & Wang, \textit{supra} note 146, at 25 (suggesting that under their design of contingent capital, where the state-contingent conversion ratio prevents value transfer, the prices would be kept “smooth’ at conversion”).
equity portion of executives' compensation packages. The conversion thus not only affects the debt portion of executives' compensation but also the equity portion after the contingent convertible bonds portion is converted and when equity is increasingly important to maintain the overall value of executive compensation. The combined effect could be a strong incentive for executives to take lower risks in order to avoid the triggering event. The implicit reduction in agency costs may be difficult to measure without actual implementation. The early conversion feature does not only create incentives for managers to lower their risk-taking; indeed, the implicit threat of financial loss for all contingent convertible bond investors, in combination with the threat of dilution to existing shareholders, could create overall increased pressure on managers to avoid the conversion of any portion of CCB, including the conversion of investors' CCBs.

The issuance of contingent convertible bonds with early triggers to executives may motivate shareholders, CCBs

173. This effect would be multiplied if the design of early CCBs triggers also prohibited the exercise of stock options in anticipation of the early trigger and after the triggering event. It could be technically difficult to delineate what “in anticipation of the early trigger” means. Defining a suitable time period for the prohibition before the early trigger could be equally difficult.

174. There is a risk that opportunism may lead managers to increase the risk profile of the entity after conversion to regain the lost equity value in their portfolio. However, such action is not likely to occur because an increase in the risk profile would affect the ability of the entity to obtain other forms of financing such as bridge loans. The increase in the risk profile after conversion is also improbable because the early conversion would likely increase monitoring by CCBs investors and shareholders. Boards may decide to let managers go upon the occurrence of an early trigger.

175. Managing to avoid the triggering event alone may be possible by ignoring the other interests of various constituents. However, this Article does not attempt to provide a holistic approach to corporate governance reform in SIFIs.

176. For purposes of this Article, “agency cost” is defined as the cost for the corporation as principal to supervise its executives as agents and protect investors and other constituents against agents’ opportunism. Executive compensation agreements can be seen as attempts to reduce agency costs. See Jensen & Murphy, supra note 7, at 138, 139–40 (explaining that executive compensation agreements and compensation awards are mostly attempts by the principal to minimize agency costs, i.e., minimize their agents’ opportunism and tendency to be risk averse, to invest in suboptimal or idiosyncratic projects, to shirk, etc.); cf. Bebchuk & Fried, supra note 65, at 1917–18 (discussing the manner in which some forms of compensation agreements increase agency costs by incentivizing risk-taking and explaining how this can be fixed).
investors, and other creditors to increase their monitoring. Because of an implicit expectation that the government will provide bailout funding due to the nature of the entity, ordinary SIFI creditors may have suboptimal incentives to monitor the performance of management. 177 These incentives may change if a SIFI issues contingent capital securities with a conversion feature for investors and with an early conversion feature for its executives. 178 Should the conversion of executives’ contingent convertible bonds occur, investors’ CCBs would likely be next in line for conversion. 179 In effect, the conversion of the contingent convertible bonds issued to executives provides early notice to


178. Once a SIFI has issued contingent convertible bonds with conversion features, creditors are likely to be aware of the triggering events in the executives’ portion of CCBs because the triggers become public information. If a SIFI issues CCBs, its intent is likely to avoid future bailouts. Creditors would be aware of the SIFI’s intent and the measures taken to avoid a bailout and may be less likely to rely on future bailouts.

179. The proximity of conversion of executives’ CCBs and investors’ CCBs will depend on the trigger design. If the entity uses an institution-specific early automatic trigger based on a capital ratio for executives’ CCBs and a similar trigger with a less aggressive triggering threshold for investors’ CCBs, the proximity of conversion would depend on the difference in capital ratio in the respective trigger designs of executives’ CCBs and investors’ CCBs.
investors that their CCBs may also convert into equity unless they address the underlying issues. Accordingly, managers would likely be subjected to increased scrutiny by investors who fear the conversion of their contingent convertible bonds into near-worthless equity. Shareholders who fear the loss of their entire investment should the entity go into bankruptcy may also put managers under increased pressure to lower their risk-taking. The impending threat of dilution for existing shareholders due to the possible conversion of the contingent convertible bonds held by investors may motivate existing shareholders to get actively involved in the governance of the entity. Shareholder voting on management proposals to address perceived concerns could increase. The implicit threat of financial loss for all contingent convertible bonds investors in combination with the threat of dilution to existing shareholders could create overall increased pressure on managers to avoid the conversion of any portion of the CCBs.

Early triggers for contingent convertible bonds in executive compensation packages may increase and optimize the signaling of default risk at a time when the risk of default is present but still somewhat remote. Various existing measures signal default risk. Existing signaling mechanisms for default risk, such as credit default swap pricing and Capital Asset management, do not seem to have provided sufficient protection for SIFIs during the past crisis. CAMEL ratings and credit default swap pricing did not suffice to signal default risk in the cases of Lehman Brothers, Bear Stearns, and Merrill Lynch. See generally Aline Darbellay & Frank Partnoy, Credit Rating Agencies and Regulatory Reform, in RESEARCH HANDBOOK ON THE ECONOMICS OF CORPORATE LAW (Claire A. Hill & Brett H. McDonnell eds., 2012); Tao Sun, Identifying Vulnerabilities in Systemically-

---

180. A possible downside of early conversion as a warning signal could be added pressure on the stock price, which could have negative effects in a pending crisis.


182. Existing signaling mechanisms for default risk, such as credit default swap pricing and Capital Asset management, Management, Earnings, and asset Liability management (CAMEL) ratings, do not seem to have provided sufficient protection for SIFIs during the past crisis. CAMEL ratings and credit default swap pricing did not suffice to signal default risk in the cases of Lehman Brothers, Bear Stearns, and Merrill Lynch. See generally Aline Darbellay & Frank Partnoy, Credit Rating Agencies and Regulatory Reform, in RESEARCH HANDBOOK ON THE ECONOMICS OF CORPORATE LAW (Claire A. Hill & Brett H. McDonnell eds., 2012); Tao Sun, Identifying Vulnerabilities in Systemically-
an entity’s default risk much sooner than previously proposed triggers for contingent convertible bonds issued to investors and other mechanisms. The benefits of this early default signal include more time for managers to (1) adjust to the current market conditions, (2) lower their risk-taking, and (3) deleverage in a comparatively liquid market environment. The improved signaling of default risk through the early conversion of executives’ CCBs may help address the systemic risks that SIFIs pose. Early triggers in CCBs, when used in executive compensation, can improve the core function of CCBs, that is, lowering systemic risk.

The hybrid nature of contingent convertible bonds and the early conversion from debt to equity aligns the interests of executives equally with creditors and shareholders. Before the unlikely but theoretically possible conversion into equity, contingent convertible bonds in executive compensation align the interests of executives with holders of traditional debt and debt in the form of contingent convertible bonds. Because executives would be holding securities with long-term maturities and coupon payments, executives would have incentives to manage the company with the interests of debt-holders in mind. How well these incentives work may depend on the proportion of debt in the compensation packages of executives. Managers’ level of risk-taking and their strategic management of the entity could become more focused on long-term and sustainable development as a result of the interest alignment between managers and debt-

183. Traditionally, because shareholders elected directors and senior executives played a major role in this process, executives’ interests were aligned with shareholder interests. See generally Bebchuk & Fried, supra note 1, at 655–56 (showing the enormous influence CEOs can have in the election of directors and other governance issues). The emphasis on equity-based instruments in executive compensation has made a large proportion of executives’ compensation dependent on stock performance, which aligns executives’ interests with the interests of shareholders. See Core, Guay & Larcker, supra note 19 (synthesizing the broad literature on equity compensation and executive incentives, and highlighting topics that seem especially appropriate for future research); Michael C. Jensen & Kevin J. Murphy, Performance Pay and Top-Management Incentives, 98 J. POL. ECON. 225, 226–27 (1990) (estimating the magnitude of the various mechanisms through which compensation policy can provide value-increasing incentives, including performance-based bonuses and salary revisions, stock options, and performance-based dismissal decisions); Meulbroek, supra note 19, at 5 (“Finance theory has long made the case for the use of equity-linked compensation plans as an effective means to align managers’ incentives with those of shareholders. In the last decade, finance practice, particularly in the United States, has embraced this prescription . . . .”); cf. Guido A. Ferrarini & Maria C. Ungureanu, Economics, Politics, and the International Principles for Sound Compensation Practices: An Analysis of Executive Pay at European Banks, 64 Vand. L. Rev. 431, 460 (2011) (explaining that the French banking model endured the crisis better than the traditionally liberal British model because of heavy regulation). Without a substantial portion of compensation in the form of debt instruments, managers may give into shareholder pressure to take higher risks for higher returns.

184. Managing the entity in the interest of debt-holders has implications for risk-taking, income inequality, and the sustainable development of the entity. See infra Part VI (discussing the impact of the proposed design on income inequality and sustainability).
holders and managers’ changed incentives. Because the early trigger for executives’ contingent convertible bonds also protects shareholders and their interest in the continuing existence of the entity, executives’ interests may be equally aligned with shareholders at a time when it is most needed.

If contingent convertible bonds replace a portion of equity-based compensation, contingent convertible bonds could lower income inequality and increase sustainability. If financial institutions begin using CCBs in executive compensation packages, depending on the calibration of the packages, the

---

185. Given that the average tenure of executives in the United States is less than seven years, interest alignment between managers and debt-holders is somewhat limited. However, debt in executive compensation could help shift executives’ management style so that it increasingly includes a long-term perspective. Managing for the long-term, in turn, could translate into longer tenures for executives. For a discussion of the increasing turnover rate of CEOs in the United States, see, for example, Steven N. Kaplan & Bernadette A. Minton, *How Has CEO Turnover Changed?*, 12 INT’L REV. FIN. 57, 58 (2012) (stating that a study of CEO tenure revealed that “[a]nnual turnover is 15.8% from 1992 to 2007, implying an average tenure as CEO of less than 7 years. . . . Since 2000, total CEO turnover increases to about 16.8%, implying an average CEO tenure of about 6 years.”); see also *The Doofus Factor: How Can You Tell a Good Board of Directors from a Bad One?* ECONOMIST, Sept. 17, 2011, at 69, available at http://www.economist.com/node/21529101 (“During the past decade the average tenure of chief executives has fallen to 6.6 years from 8.1 years, according to a recent study by Booz & Co, a consultancy.”).

186. See infra Part VI (elaborating on improved incentives for lowering risk, improved signaling of default risk, and improved monitoring).

187. Regardless of compensation policies, the strong relationship between managers and shareholders will likely endure. It is unlikely that firms will compensate managers entirely with debt. See infra Part VI.A (discussing contingent capital as inside debt).

188. The early trigger design could help align the interests of the two most powerful constituents in a financial institution at a time when it is most needed. Because executives become equity holders at a time when their actions should be most aligned with equity holders’ interests, that is, early before a possible insolvency of the entity, the early trigger design enables a shift in executive compensation and a corresponding interest alignment when it is most needed. Although the interests of managers are likely to be more aligned with those of shareholders, executives who hold hybrid securities may be incentivized to manage for the long-term and sustainable development of the entity and avoid volatility.

replacement of equity-based compensation with contingent convertible bonds could lower the overall compensation of executives. Although contingent convertible bonds would likely pay a substantial coupon rate,\textsuperscript{190} the return for executives would likely be incomparable with the return attainable with stock options and other equity-based compensation. If stock options are replaced with CCBs, not only would the total compensation for executives be lowered; short-termism and executives’ focus on quarterly stock price performance would also be disincentivized. Executives are more likely to consider the sustainable development of the entity when stock price appreciation no longer directly benefits them personally. However, it is important to note that path dependencies in the executive compensation culture\textsuperscript{191} in the United States could make the lowering of overall compensation for executives and the addition of new design elements in executive compensation difficult.

\textbf{C. Design Features to Avoid Abuse}

If SIFI executives are compensated with contingent convertible bond instruments, opportunism could lead them to manipulate the triggering event to obtain stock upon contingent convertible bonds’ conversion at a depressed price before or during a crisis.\textsuperscript{192} For instance, the conversion into equity at a

\textsuperscript{190} See Kaal, supra note 98 and accompanying text (discussing the coupon rate of between 7% and 9.2% paid by previously issued CCBs).


\textsuperscript{192} See Murphy, Walsh & Willison, supra note 140, at 7, 10 ("[T]he trigger metric could be undermined if it could be manipulated . . . ."). “With a trigger
value of $3 with a book value of $10 would create a substantial payoff. Opportunism may lead executives to disregard the impact of triggering the conversion on their reputation and career development prospects if the payoff is substantial.

Without rules and regulatory guidance regarding the design and issuance of contingent convertible bonds, SIFI executives may only be curtailed by their fiduciary duties. Existing fiduciary duties could prove insufficient to limit opportunism and abuse if the payoff for executives is substantial. Contingent convertible bondholders, regular bondholders, and shareholders would benefit from broader fiduciary duties.

metric based on a bank's equity price, there would be a risk that investors may short-sell a bank's equity to drive the equity price down in the absence of any change in the underlying value of a bank's assets and trigger a conversion event that results in a transfer of value from existing equity holders to precautionary contingent capital holders . . . .” Id. “[T]he risk of using market capitalisation to define the trigger event is that it could give investors an incentive to manipulate the equity price to trigger a conversion.” Id.


194. See Wall Street Fraud and Fiduciary Duties: Can Jail Time Serve as an
Adequate Deterrent for Willful Violations?: Hearing Before the S. Subcomm. on Crime and Drugs, Comm. on the Judiciary, 111th Cong. 835 (2010) (statement of John C. Coffee Jr., Adolf A. Berle Professor of Law, Columbia University Law School) (arguing that legislation is needed “to protect investors and to maintain market transparency and economic efficiency” and to return to the traditional norm that brokers should seek “to serve their clients (and not seek to profit from their losses)”; Danny Busch & Deborah A. DeMott, Liability of Asset Managers (Introduction), in LIABILITY OF ASSET MANAGERS (Danny Busch & Deborah A. DeMott eds., 2012); TAMAR FRANKEL, FIDUCIARY LAW (2011) (clarifying the theoretical underpinning for an expansive version of fiduciary duties and applying fiduciary theory to such contemporary problems as those in the securities industry and the professions, as well as to corporate issues such as executive compensation); Thomas Lee Hazen, Stock Broker Fiduciary Duties and the Impact of the Dodd–Frank Act, 15 N.C. BANKING INST. 47, 47 (2011) (“Although the existing framework for broker-dealer regulation is robust, it could be fine-tuned by possibly adding an express fiduciary duty requirement as well as more specific rule-based prohibitions.”); Douglas C. Michael, The Corporate Officer’s Independent Duty as a Tonic for the Anemic Law of Executive Compensation, 17 J. CORP. L. 785, 786, 824 (1992) (arguing that officers should have a “duty not to accept unreasonable compensation” and that courts should use officers’ fiduciary duties to engage in a sweeping review of the reasonableness of compensation); see also Lisa M. Fairfax, Spare the Rod, Spoil the Director? Revitalizing Directors’ Fiduciary Duty through Legal Liability, 42 HOUS. L. REV. 393, 394 (2005) (“[L]egal liability represents an essential mechanism for ensuring directors’ fidelity to their fiduciary duties and for questioning reform efforts that do not include such liability.”). For advocates of a fiduciary duty to bondholders, see, for example, LAWRENCE E. MITCHELL, PROGRESSIVE CORPORATE LAW (Lawrence E. Mitchell ed., 1995); Margaret M. Blair, Stakeholders as Shareholders, Ownership and Control: Rethinking Corporate Governance for the Twenty-First Century, 109 HARV. L. REV. 1150 (1996); William W. Bratton Jr., Public Values and Corporate Fiduciary Law, 44 RUTGERS L. REV. 675 (1992); Jensen & Meckling, supra note 30, at 305; David Millon, Communitarians, Contractarians, and the Crisis in Corporate Law, 50 WASH. & LEE L. REV. 1373 (1993); David Millon, Redefining Corporate Law, 24 IND. L. REV. 223 (1991); Lawrence E. Mitchell, A Critical Look at Corporate Governance, 45 VAND. L. REV. 1263 (1992); Wai Shun Wilson Leung, The Inadequacy of Shareholder Primacy: A Proposed Corporate Regime that Recognizes Non-Shareholder Interests, 30 COLUM. J.L. & SOC. PROBS. 587 (1997).

Contra Wall Street Fraud and Fiduciary Duties: Can Jail Time Serve as an Adequate Deterrent for Willful Violations?: Hearing Before the S. Subcomm. on Crime and Drugs, Comm. on the Judiciary, 111th Cong. 835 (2010) (statement of Larry E. Ribstein, Associate Dean for Research & Mildred Van Voorhis Jones Chair, University of Illinois College of Law); Larry E. Ribstein, Fencing Fiduciary Duties, 91 B.U. L. REV. 899 (2011) (arguing for a more precise definition and more limited application of fiduciary duties because their usefulness depends on differentiation from other duties that apply in other settings). For cases that discuss officers’ fiduciary duties, see generally Hill & McDonnell, supra note 193, at 133.
In practice, the proportion of contingent convertible bond instruments in the total mix of executive compensation packages will likely be limited, which may counteract possible abuse. A mandatory holding period for all equity securities held by executives in the entity they manage after the conversion of their contingent convertible bonds into equity takes place would help limit possible abuse. A broader approach to curtail abuse could entail the cancellation of any stock options or other equity-based pay arrangements in the compensation packages of executives upon the conversion of executives' CCBs. Ultimately, manager/agent opportunism that results from the design of CCBS triggers and managers' participation in the design may necessitate regulatory guidance in the form of general principles or best practice guidelines for the design of CCBs and their issuance to investors and executives.

Although executives are unlikely to be directly involved in designing their own contingent convertible bond awards, executives often have a strong involvement and influence in the executive compensation process. Assuming that executives act opportunistically within the bounds of their fiduciary duties, executives may not create socially optimal designs for contingent convertible bonds. Executives' opportunism and involvement in drafting the trigger could result in suboptimal early triggers in contingent convertible bonds issued to executives. If the trigger design allows executives to influence the early trigger, the suboptimal early trigger could increase the potential for abuse because executives may use the trigger to obtain cheap stock during a crisis. If executives are involved in drafting the features of contingent convertible bonds issued to investors, their opportunism could result in contingent convertible bond designs that do not result in a significant threat of equity dilution upon conversion and other governance-improving design features. Because regulators and others are unlikely to provide rules,

195. See Bhagat & Romano, supra note 1, at 2–3 (discussing equity-based compensation).
196. Bebchuk & Fried, supra note 1; Bebchuk & Spamann, supra note 2.
197. See supra Part IV (showing that the incomplete contract theory model acknowledges agent opportunism in the analysis of executive compensation policies).
guidelines, and best practice guidance for contingent convertible bond designs and issuances anytime soon, executives may continue to have a strong involvement in the design of contingent convertible bonds.\textsuperscript{199} Regulatory guidance on contingent capital designs and issuances may be needed to curtail the involvement of executives in the design of contingent convertible bonds and create socially optimal designs.\textsuperscript{200} Contingent convertible bonds with early triggers in executive compensation could help create socially optimal designs for contingent convertible bonds issued to investors.\textsuperscript{201}

\textbf{VI. Implications for Academic Debates}

Instituting contingent convertible bonds with early triggers in executive compensation packages adds a new perspective to the academic debate in the context of inside debt, the creditor-centered approach to executive compensation, and the design of triggering events.\textsuperscript{202} As a result of their early conversion into equity (or the threat thereof), contingent convertible bonds provide several advantages beyond the benefits of traditional inside debt instruments in executive compensation.\textsuperscript{203} Contingent convertible bonds also add benefits to the creditor-centered approach to executive compensation.\textsuperscript{204} In the context of the creditor-centered approach, adding contingent convertible bonds or replacing other debt instruments in executive compensation

\begin{footnotesize}
199. Institution- and industry-specific knowledge may make executives indispensable in drafting contingent convertible bonds. Regulatory guidelines and best practice guidance for contingent convertible bond designs could help ensure that executives are sufficiently involved but do not create socially suboptimal designs.

200. See Kaal, \textit{supra} note 14, at 139–41 (noting that private ordering and market mechanisms may not result in socially optimal designs for contingent convertible bonds).

201. See \textit{supra} Part V.B.2 (discussing the benefits of a design with early triggers).

202. See \textit{supra} Part VI.A–C (discussing contingent capital as inside debt, improving the creditor-centered approach, and the impact of ownership characteristics on trigger design).

203. See \textit{infra} Part VI.A (discussing contingent capital as inside debt).

204. See \textit{infra} Part VI.B (discussing ways to improve the creditor-centered approach).
\end{footnotesize}
packages with contingent convertible bonds can optimize managers’ risk incentives, especially in comparison with traditional debt instruments in executive compensation packages. In the context of the debate on trigger designs, contingent convertible bonds with early triggers show that the ownership characteristics of the contingent convertible bondholders can have an impact on the efficient design of contingent capital triggers.

A. Contingent Capital as Inside Debt

Empirical studies have demonstrated that risk-taking declines if executives hold more debt relative to their equity holdings. Inside debt may be defined as a corporation’s debt

205. See infra Part VI.B (discussing ways to improve the creditor-centered approach).
206. See infra Part VI.C (discussing the impact of ownership characteristics on trigger design).
207. See Calogno & Renneboog, supra note 82, at 1808–09 (maintaining that the increase in the leverage of Anglo-American corporations has stimulated interest in the role of debt as a direct incentive device for managers to generate stronger corporate performance and showing that including risky debt in the capital structure changes a principal’s “incentive to give incentives”); see also Edmans & Liu, supra note 82 (arguing that a debt bias can improve executives’ efforts as well as deter risk-shifting); Tao-Hsien Dolly King & Min-Ming Wen, Shareholder and Bondholder Governance, and Managerial Risk-Taking, 35 J. BANKING & FIN. 512, 530 (2011) (showing that strong bondholder governance incentivizes low-risk investments); Ortiz-Molina, supra note 82, at 78–83, 90–91 (examining how CEO compensation is related to firms’ capital structures and arguing that the hypothesis that debt reduces manager–shareholder conflicts can explain some—but not all—of the results); Sundaram & Yermack, supra note 8; Gerakos, supra note 8, at 23 (finding that pension benefits may reduce risk-taking); Tung & Wang, supra note 8 (showing that their empirical evidence provides a rationale for the use of inside debt compensation in structuring executive compensation in the banking context); Wei & Yermack, supra note 8. Other studies reject the idea that corporate governance and executive compensation are correlated. See Core et al., supra note 8, at 385–88; Dorf, supra note 169, at 5; Johnson et al., supra note 169, at 17, 38; Carter & Lynch, supra note 169, at 222 (finding no relationship between institutional ownership and the re-pricing decision); see also Anderson & Bizjak, supra note 169, at 1344; Wan, supra note 169, at 23 ("[T]here is no systematic evidence that board composition affects change in CEO compensation."). Some scholars argue that there is no role for inside debt in executive compensation because bonuses, salaries, and managerial reputation constitute adequate remedies to debt’s agency costs. See David Hirshleifer & Anjan V. Thakor, Managerial
hold by firm insiders.\textsuperscript{208} Inside debt, which includes pensions and deferred compensation,\textsuperscript{209} is already a substantial part of executive compensation in the United States.\textsuperscript{210} Unlike equity-based compensation, inside debt creates fixed claims for managers with a stake in the firm’s liquidation value, which reduces risk-taking incentives for executives and the associated agency costs.\textsuperscript{211} When managers have a stake in the liquidation value of the firm, they are more likely to increase their efforts in the vicinity of insolvency.\textsuperscript{212} Inside debt could provide adequate signaling of managers’ risk-taking.\textsuperscript{213}

Inside debt in the form of contingent convertible bonds has characteristics that could prove preferable to traditional inside debt.\textsuperscript{214} Managers’ inside debt stake in the firm’s liquidation

\begin{itemize}
\item \textsuperscript{208} Cassell et al., supra note 8, at 588 (“CEO inside debt holdings (pension benefits and deferred compensation) are generally unsecured and unfunded liabilities of the firm.”); Edmans & Liu, supra note 82, at 75 (defining inside debt as debt—or any security with payoffs very similar to debt—held by the manager, and contrasting it with outside debt, which is held by external investors); Jensen & Meckling, supra note 30, at 352 (defining inside debt as “debt held by the manager”).
\item \textsuperscript{209} Edmans & Liu, supra note 82, at 76 (“U.S. CEOs hold substantial defined benefit pensions. These are unsecured, unfunded obligations which, in nearly all cases, have equal priority with other creditors in bankruptcy and thus constitute inside debt.”); Sundaram & Yermack, supra note 8, at 1 (“The most common form of these intracompany IOUs are benefit pensions and deferred compensation.”); Tung & Wang, supra note 8, at 13 (defining bank CEOs’ inside debt as the present value of the CEO’s pension and deferred compensation balances).
\item \textsuperscript{210} See Sundaram & Yermack, supra note 8.
\item \textsuperscript{211} Edmans & Liu, supra note 82 (showing that the probability of default and the manager’s ability to affect liquidation values affect the appropriate amount of inside debt).
\item \textsuperscript{212} Tung & Wang, supra note 8, at 4.
\item \textsuperscript{213} Tung, supra note 9, at 35 (arguing that subordinated inside debt on the subsidiary level is preferable to debt of the holding company for signaling). Contra Bebchuk & Spamann, supra note 2, at 273–97 (proposing to pay executives through debt of the bank holding company).
\item \textsuperscript{214} See supra Part V.B.2 (elaborating on the benefits of an early trigger design).
\end{itemize}
value and the associated governance benefits\textsuperscript{215} incentive optimization\textsuperscript{216} and reduction of agency costs\textsuperscript{217} depend on the solvency of the respective entity. Inside debt without a conversion feature provides no mechanism to ensure the solvency of the entity. Contingent convertible bonds with a conversion feature offer the additional benefit of creating an early warning system and a buffer before insolvency that can help an entity avoid default.\textsuperscript{218} As a debt instrument with an early trigger before conversion, contingent convertible bonds in executive compensation have all the benefits of inside debt\textsuperscript{219} in addition to the benefits of an early trigger design.\textsuperscript{220}

Because the contingent convertible bonds in executive compensation packages would convert into equity early before insolvency, the value of the equity after conversion of the respective executives’ contingent convertible bonds before bankruptcy could still be higher than the liquidation value of

\begin{itemize}
\item \textsuperscript{215} Cassell et al., supra note 8, at 589 (“[O]ther studies find that inside debt holdings are associated with higher firm liquidation value . . . and lower credit default swap spreads . . . .”); Edmans & Liu, supra note 82, at 79; Sundaram & Yermack, supra note 8, at 1558; Tung, supra note 9, at 26.
\item \textsuperscript{216} Tung, supra note 9, at 3 (arguing that market pricing of inside debt is particularly sensitive to downside risk and that including inside debt in bankers' compensation packages could therefore give managers “direct personal incentives to avoid excessive risk”).
\item \textsuperscript{217} Id.; see also Jensen & Meckling, supra note 30; Sundaram & Yermack, supra note 8, at 1572 (“[D]ebt-based compensation reduces the agency costs of debt . . . we should observe a positive association between the CEO’s debt-to-equity ratio and the firm’s leverage.”); Edmans & Liu, supra note 82, at 79 (demonstrating that inside debt is a superior remedy to the agency costs of debt than the bonuses advocated by prior research).
\item \textsuperscript{218} See Kaal & Henkel, supra note 14 (discussing the use of contingent capital to create a buffer in the vicinity of bankruptcy).
\item \textsuperscript{219} See Edmans & Liu, supra note 82, at 92 (“Inside debt can be a more effective solution to creditor expropriation than salaries, bonuses, reputation and private benefits, owing to its sensitivity to liquidation value.”); Sundaram & Yermack, supra note 8, at 1558, 1583 (“Inside equity aligns managers with equity holders in good states, but inside debt aligns managers with debt-holders in bad states . . . . Debt-based compensation provides managers with interesting incentives to reduce the agency costs of debt.”); Tung & Wang, supra note 8, at 26 (“CEOs’ inside debt holdings preceding the Crisis are significantly positively associated with bank performance and significantly negatively associated with bank risk taking during the Crisis.”).
\item \textsuperscript{220} See supra Part V.B.2 (explaining the benefits of the early trigger design).
\end{itemize}
traditional inside debt. Unlike the liquidation value of traditional inside debt, the equity in executives’ portfolios after the conversion of the contingent convertible bonds can still be increased because the early trigger creates a substantial buffer before insolvency. Contingent convertible bonds as inside debt would not only have the benefits of inside debt before conversion, but would also provide significant incentives for management to maintain the value of equity after conversion.

Contingent convertible bonds in executive compensation could also help optimize inside debt instruments to incentivize lower risk-taking by managers. Similar to other inside debt instruments, the market price of contingent convertible bonds would likely be affected by managers’ risk-taking. Like subordinated inside debt on the subsidiary level, before conversion into equity, contingent convertible bonds can incentivize executives to lower their risk-taking because contingent convertible bond prices are sensitive to downside risks of SIFIs, including the risk of default. Contingent convertible

---

221. See Kaal & Henkel, supra note 14, at 233–34
A common denominator in the proposals on the use of contingent capital in the context of avoiding future crises could be the issuance of a . . . percentage of a financial institution’s long-term debt capital as convertible debt securities that convert into equity when triggered by financial weakening of the . . . institution. However, debt–equity conversion is not a new concept. The financial crisis has drawn increasing attention to this concept because the conversion of debt into equity could be an attractive alternative to forcing strained, but not insolvent, financial institutions into restructuring or liquidation.

222. See Kaal, supra note 14, at 293 (“Strained financial institutions may find the automatic conversion of debt into equity an attractive alternative to being forced into restructuring or liquidation.”).

223. See supra notes 211–12 and accompanying text (explaining how the conversion of contingent capital bonds creates incentives for managers).

224. See Tung & Wang, supra note 8, at 3 (discussing the pricing sensitivity of inside bank debt on the subsidiary level with regard to executive risk-taking). Contra Bebchuk & Spamann, supra note 1, at 269–74 (suggesting that managers’ risk-taking probably does not affect inside debt in the form of debt of the bank holding company to a great extent).

225. See Tung & Wang, supra note 8, at 35 (providing tables on subordinated inside debt at the subsidiary level).

226. See Murphy, Walsh & Willison, supra note 140, at 6 (arguing that the pricing of contingent convertible bonds could be a guide to the markets’ view on
bonds, however, provide an additional feature. Upon the conversion of the contingent convertible bonds, the downside pressure on equity places additional weight on the value of the equity portion of executives' pay packages. Upon conversion, the executives would not only forego the contingent convertible bonds in their compensation packages, but they would also hold converted equity and equity held prior to the conversion at a depressed value. The price-sensitivity-induced incentives for lowering risk before conversion and the new emphasis on the equity portion of the pay package after conversion create a combined effect on executives' incentives. This combined effect could create comparatively stronger incentives for executives to take fewer risks in order to avoid the triggering event. The emphasis in this design is on incentives that lower executives' risk-taking and help avoid the early trigger.

Critics argue that liquidity shocks and other exogenous factors could influence debt trading prices unrelated to managers' the riskiness of financial institutions).

227. The conversion of contingent convertible bonds signals to the market that the entity could be insolvent which might result in downside pressure on the stock price. See Sundaresan & Wang, supra note 146, at 6 ("[C]ontingent capital is essentially a junior debt that converts to equity shares when the stock price reaches a certain low threshold.").

228. See Murphy, Walsh & Willison, supra note 140, at 6 ("But it seems unlikely that precautionary contingent capital could be less costly than equity . . . . [I]f the private information is primarily about the downside risk faced by a bank, the values of equity and precautionary contingent capital could be similarly affected by this private information.").

229. See id. at 8 ("The presence of precautionary contingent capital could also risk creating systemic problems in other ways if bank equity holders or managers seek to avoid the trigger event . . . . [M]anagers could have an incentive to do this if they fear that conversion could lead to their replacement."). Managing to avoid the triggering event alone may be possible by ignoring the other interests of various constituents. See id. ("If the trigger metric depends on a bank’s ratio of capital-to-assets or risk-weighted assets, incumbent equity holders or managers could try to reduce assets to push the ratio up and away from the trigger value.").

230. Executives are unlikely to increase their risk-taking after conversion of their CCBs to salvage the equity value of their portfolio because of the public nature of the trigger, the entity's vicinity to bankruptcy, and managers' inability to obtain other forms of financing if they increase the risk profile of the entity. See supra Part V.C (suggesting a mandatory holding period upon the conversion of executives' contingent convertible bonds for all equity in the entity the executives manage).
risk-taking, making debt at the subsidiary level less useful for signaling. 231 Contingent convertible bonds should be less prone to liquidity shocks because of their inordinately high coupon rate of between 7% and 9.5%. 232 Although there is no assurance that oversubscribed issuances 233 guarantee future liquidity, the coupon rates of contingent convertible bonds could help hedge the liquidity risk. 234

Critics also allege that the inside debt theory of executive compensation ignores suboptimal short- and long-term incentives that may result from the inclusion of long-term debt in the

---

231. See Sepe, supra note 1, at 211 (arguing that the creditor-centered approach is flawed for three reasons: (1) banks’ debt obligations include, in large part, private debt; (2) bank-issued bonds trade at a markedly lower volume than bank equity; and (3) bank liabilities not included on balance sheets exist but their existence is difficult to demonstrate).

232. See Press Release, Credit Suisse, Credit Suisse Grp. Executes Agreement to Put in Place CHF 6 Billion of Tier 1 Buffer Capital Notes (Feb. 14, 2011), available at https://emagazine.credit-suisse.com/app/article/index.cfm?fuseaction=OpenArticle&aoid=300000&coid=293554&lang=EN; Press Release, Credit Suisse, Credit Suisse Group Places 7.875% Tier 2 Buffer Capital Notes (Feb. 17, 2011), available at https://publications.credit-suisse.com/app/article/index.cfm?fuseaction%20=OpenArticle&aoid=300500&coid=293551&lang=EN (“The Tier 2 BCNs were offered on a ‘Regulation S-only’ basis outside the US and other restricted jurisdictions in a minimum denomination of USD 100,000. The USD 2 billion Tier 2 BCNs will initially carry a coupon rate of 7.875% per annum.”). Regarding the Barclays issuance and coupon rate, see supra notes 134–35; see also Paul Clarke, BarCap’s 7% Coco Coupon Is Decidedly More Generous Than Most Deferred Bonuses, REUTERS, (Apr. 20, 2011), available at http://news.reuters.efinancialcareers.co.uk/News_ITEM/newsItemId-32101 (“The coupon rate of 7% is not just too generous for shareholders, it also outstrips the rate of interest being paid on most other deferred cash bonuses.”); Treanor, supra note 135, at 50 (“The Barclays cocos would pay a 7% coupon—or rate of interest—annually, not compounded.”).

233. See Katharina Bart, Credit Suisse Sells $2 Billion of Co-Cos to Public, WALL ST. J. (Feb. 17, 2011, 6:44 PM), http://online.wsj.com/article/SB10001424052748704546704576150861690164484.html (“A person familiar with the situation said the issue was oversubscribed.”) (on file with the Washington and Lee Law Review); Mary Watkins, Credit Suisse to Use ‘Cocos’ to Raise Sfr250M, FT.COM, (Mar. 7, 2012, 6:25 PM), http://www.ft.com/intl/cms/s/0/e77c968c-686e-11e1-b803-00144feabdc0.html#axzz1xqtwSc2 (“The bank last year raised $2bn using cocos in a heavily oversubscribed issue.”) (on file with the Washington and Lee Law Review); Murphy, Walsh & Willison, supra note 140, at 9 (stating that the CS issue was around eleven times oversubscribed); see also Kaal, supra note 14, at 312–15.

234. See Murphy, Walsh & Willison, supra note 140, at 15 (explaining the effects of varying coupon rates).
compensation packages of executives. They argue that long-term debt securities in the compensation packages of executives will not disincentivize high-risk, short-term transactions because managers can expect short-term gains that exceed the discounted value of their long-term debt securities. Adding contingent convertible bonds with early triggers as part of the inside debt portion of executives’ compensation packages could change managers’ incentives. Executives would no longer simply focus on the debt versus equity portion of their portfolio; they would also consider the effects of triggering events. In the day-to-day operation of the business, managers would need to consider the triggering event of a substantial portion of their outstanding debt instruments and the implications that a triggering event would have on the entity and their personal finances. Managing to avoid the early trigger would allow for enough risk-taking by managers to generate sufficient returns but, at the same time, would curtail risk-taking enough to avoid the negative effects of the triggering event.

235. See Sepe, supra note 1, at 223 ("Accordingly, the theory overlooks the perverse incentives long-term debt may produce both in the short- and long-run.").

236. See id. ("Tying managers’ financial rewards to debt securities with a long-term maturity will not induce managers to refrain from taking risky bets in the short run, because the expected short-term gains from these bets will tend to exceed the discounted value of managers’ debt holdings."). With equity-based compensation in managers’ compensation packages, managers may still be incentivized to drive up the stock price and exercise stock options at an opportune time. However, risk-taking generated by equity-based pay can be overcome if the calibration of debt- and equity-based compensation in the executive compensation package favors debt and requires a minimal holding period. See Bhagat & Romano, supra note 1, at 361–62 (discussing the equity-based part of the compensation package).

237. See supra Part V.B (discussing the effects of triggering events and management incentives). Triggering events in contingent capital securities can take various forms. The debate on what triggers should be used is ongoing. See Kaal, supra note 14, at 300 ("While institution-specific triggers would presumably grant most certainty to market participants, regulatory trigger designs could provide lower levels of certainty."); Kaal & Henkel, supra note 14, at 233–38 (examining the different trigger designs and their effects on risk-taking).

238. See supra note 169 and accompanying text (noting studies that explore how managers behave depending on the ratio of their debt to their equity).

239. These are ideal typical model assumptions. However, with the right trigger design, executives’ incentives and interests could be substantially

---

235. See Sepe, supra note 1, at 223 ("Accordingly, the theory overlooks the perverse incentives long-term debt may produce both in the short- and long-run.").

236. See id. ("Tying managers’ financial rewards to debt securities with a long-term maturity will not induce managers to refrain from taking risky bets in the short run, because the expected short-term gains from these bets will tend to exceed the discounted value of managers’ debt holdings."). With equity-based compensation in managers’ compensation packages, managers may still be incentivized to drive up the stock price and exercise stock options at an opportune time. However, risk-taking generated by equity-based pay can be overcome if the calibration of debt- and equity-based compensation in the executive compensation package favors debt and requires a minimal holding period. See Bhagat & Romano, supra note 1, at 361–62 (discussing the equity-based part of the compensation package).

237. See supra Part V.B (discussing the effects of triggering events and management incentives). Triggering events in contingent capital securities can take various forms. The debate on what triggers should be used is ongoing. See Kaal, supra note 14, at 300 ("While institution-specific triggers would presumably grant most certainty to market participants, regulatory trigger designs could provide lower levels of certainty."); Kaal & Henkel, supra note 14, at 233–38 (examining the different trigger designs and their effects on risk-taking).

238. See supra note 169 and accompanying text (noting studies that explore how managers behave depending on the ratio of their debt to their equity).

239. These are ideal typical model assumptions. However, with the right trigger design, executives’ incentives and interests could be substantially
B. Improving the Creditor-Centered Approach

Contingent convertible bonds with early triggers in executive compensation can help to improve the creditor-centered approach to executive compensation. Existing compensation practices have an effect on managers' risk-taking and risk preferences as well as the firm's long-term profitability and sustainability.240 Including improved. See, e.g., McDonald, supra note 95, at 3 (“The fact that the dual-trigger structure permits banks to sometimes fail addresses the concern that contingent capital would blunt the incentive effects of debt.”).

240. See Bebchuk & Fried, supra note 1, at 671 (“Unlike defined contribution plans, which force the employee to bear the risk of poor investment performance, defined benefit plans shift the risk of investment performance to the firm.”); Carl R. Chen, Thomas L. Steiner & Ann Marie Whyte, Does Stock Option-Based Executive Compensation Induce Risk-Taking? An Analysis of the Banking Industry, 30 J. BANKING & FIN. 915, 916 (2006) (“The compensation level and structure employed by each bank has implications for risk-taking and for the agency relation between managers and stockholders.”); Jeffrey L. Coles, Naveen D. Daniel & Lalitha Naveen, Managerial Incentives and Risk-Taking, 79 J. FIN. ECON. 431, 442–43 (2006) (providing empirical evidence of a strong causal relation between managerial compensation and investment policy, debt policy, and firm risk); Robert Haugen & Lemma Senbet, Resolving the Agency Problems of External Capital Through Options, 36 J. FIN. 629, 640 (1981) (“[T]here may remain an incentive for the manager to engage in either high or low risk investment programs. This is the well-known wealth transfer problem associated with the existence of risky debt in the capital structure.”); Jenson & Meckling, supra note 30, at 309–10 (“[E]xisting literature focuses almost exclusively on the normative aspects of the agency relationship; that is, how to structure the contractual relation (including compensation incentives) between the principal and agent to provide appropriate incentives for the agent to make choices which will maximize the principal's welfare . . . .”); Kose John, Anthony Saunders & Lemma W. Senbet, Perspectives on Bank Capital Regulation and Managerial Compensation, 19 J. BANKING & FIN. 735 (1995); Clifford W. Smith & Rene M. Stulz, The Determinants of Firms’ Hedging Policies, 20 J. FIN. & QUANTITATIVE ANALYSIS 391, 399 (1985) (suggesting that shareholders can affect management’s risk aversion through the design of compensation contracts); Bolton, Mehran, & Shapiro, supra note 19, at 1 (“[S]tructuring CEO incentives to maximize shareholder value in a levered firm tends to encourage excess risk taking.”); COUNCIL OF INSTITUTIONAL INVESTORS, TOP 10 RED FLAGS TO WATCH FOR WHEN CASTING AN ADVISORY VOTE ON EXECUTIVE PAY 1 (2010), available at http://www.cii.org/UserFiles/file/resource%20center/publications/March%202010%20-%20Say%20on%20Pay%20Checklist.pdf (“Poorly-designed incentives can promote excessive risk-taking and get-rich quick mentalities—key contributors to the financial crisis.”). Several empirical studies have explored the connection between managerial stock, option holdings, or both, and financial strategy or corporate focus (such as leverage, repurchase, or the extent of derivatives usage and hedging). See generally Anup Agrawal & Gershon Mandelker, Managerial Incentives and Corporate Investment and Financing Decisions, 42 J. FIN. 823
in executive compensation packages a basket of securities that represents a predefined percentage of the common shares could help address the shortcomings of existing executive compensation practices.241

Adding contingent convertible bonds to the basket of securities in executive compensation could further improve executive compensation practices. Like other debt securities in executive compensation, contingent convertible bonds may increase sustainability and can help overcome income inequality.242 If a SIFI issues contingent convertible bonds to

---

241. See Bebchuk & Spamann, supra note 1, at 253, 283–84 (discussing how executive compensation packages are one of several corporate governance measures that address excessive risk-taking and how regulating such incentives should be standard procedure in the future). Weighting debt securities, including contingent convertible bonds, heavily in the calibration of executive compensation packages may increase the positive effects of debt in executive compensation packages. See supra note 237 and accompanying text (expounding on the positive effects). At the same time, debt may not be the preferred form of compensation for executives. Calibrating executive compensation packages to account for desired incentives and governance improvements while giving sufficient incentives for executives to perform within expected parameters may require an institution-specific relational approach and learning from experience. See supra Part IV (discussing the benefits of the incomplete contract theory of NIE).

replace a substantial portion of its executives’ equity-based compensation, the coupon rate of contingent convertible bonds (between 7% and 9.5%).\textsuperscript{243} albeit higher than traditional debt instruments, could help address concerns over income inequality.\textsuperscript{244} For the remaining portion of executives’ equity-based compensation, a mandatory holding period upon the conversion of the contingent convertible bonds could provide incentives for sustainability.\textsuperscript{245} The coupon rate of contingent convertible bonds may be incomparable with the upside potential of equity-based compensation, especially for stock options. The coupon rate of contingent convertible bonds, however, is higher than the coupon rate of traditional debt instruments.\textsuperscript{246} This


Inequality today at the top end of the income distribution can be attributed to four different sectors of the economy—top executives of non-financial firms (Main Street); financial service sector employees from investment banks, hedge funds, private equity funds, and mutual funds (Wall Street); lawyers; and professional athletes and celebrities.

\textsuperscript{243} See \textit{supra} note 232 and accompanying text (describing Credit Suisse issuance statistics).

\textsuperscript{244} See \textit{supra} Part V (discussing how the governance-improving features of contingent capital bonds can be applied to executive compensation); \textit{supra} note 188 and accompanying text (explaining that contingent convertible bonds can lower income inequality and increase sustainability).

\textsuperscript{245} See Bhagat & Romano, \textit{supra} note 1, at 361 (suggesting a two- to four-year holding period for equity-based compensation).

\textsuperscript{246} See Markus Pelger, \textit{Contingent Convertible Bonds: Pricing, Dilution Costs and Effective Regulation} 7–11 (Coleman Fung Risk Mgmt. Research Ctr.,
higher coupon rate could make it a more attractive instrument in executive compensation. The attractive features of CCBs may help establish CCBs in the executive compensation culture in the United States and thereby improve corporate governance.

Because of the conversion feature, contingent convertible bonds may have additional benefits. Unlike traditional debt instruments in executive compensation, contingent convertible bonds provide incentives for increased monitoring by creditors and shareholders. They also align executives’ interests with the interests of shareholders and creditors. Most importantly, adding contingent convertible bonds or replacing other debt instruments in executive compensation packages with contingent convertible bonds with early triggers helps improve managers’ risk incentives. The conversion feature of contingent convertible bonds decreases executives’ risk-taking incentives because holding a substantial portion of debt rather than equity disincentivizes short-termism and executives’ focus on quarterly stock price performance. The early conversion feature also creates a risk that if executives do not lower their risk-taking and manage the entity well enough to avoid the trigger, they will receive equity at a point in time when the equity value will be substantially diminished.


247. See Murphy, Walsh & Willison, supra note 140, at 7 (“The opportunity to run before conversion could also reduce the incentives of precautionary contingent capital holders to monitor a bank’s risk-taking and impose market discipline.”).

248. See supra Part V (discussing how, because shareholders’ interests are aligned with the entity’s, executives’ interests are also aligned with the shareholders’ interests).

249. See supra note 230 and accompanying text (examining the effect of contingent convertible bonds on managers’ risk-taking).

250. The amount to be received upon conversion would depend on the trigger design. Most triggers currently discussed in the literature favor a conversion shortly before bankruptcy. If this is the case, the equity received by executives would be nearly worthless. If the SIFI chooses a contingent capital bond design with a write-down feature for its executive compensation packages, managers would perhaps have even less incentive to take risks because they would not get any financial benefit from the contingent capital bonds if they are triggered into a write-down. The beneficial effects could increase in proportion to the volume of contingent convertible bonds issued to executives.
Replacing a substantial portion of equity-based compensation with contingent convertible bonds can also help create incentives for managers to keep the business solvent. During the Financial Crisis, Lehman Brothers experienced a severe shortfall in solvency in part because its CEO Richard S. Fuld, Jr. had a substantial equity stake in the company and, in an effort to preserve his own interests, refused to pursue a dilutive capital infusion or sell the firm in order to avoid the firm's failure (the Fuld Problem). Motivated in part by a desire to address the Fuld Problem, Jeffrey Gordon suggests using convertible, equity-based pay to compensate executives. For example, executives in financial firms would receive a substantial part of their stock-related compensation in equity securities that convert to subordinated debt upon a triggering event. Gordon contrasts convertible equity-based pay with contingent convertible bonds and argues that contingent convertible bonds would not address the Fuld Problem.

If a financial institution issues contingent convertible bonds to the general public and its executives, and a large part of equity-based executive compensation is replaced with CCBs, SIFI managers may be more concerned with managing the entity to prevent the early triggering event than with averting an

251. See Executive Compensation, supra note 19, at 7 ("This might be called 'the Fuld Problem': a CEO who is reluctant to negotiate a large equity raise (or sell the firm) because the terms would massively dilute his personal equity stake and who instead may calculate that holding out [will be more beneficial to him or her].").

252. See id. at 11 ("[S]enior executives at financial firms should receive a significant portion of stock-related compensation in the form of equity that will convert into subordinated debt upon certain external triggering events . . . .'').

253. Id.

254. See id. at 11 n.18 ("Convertible equity-based pay bears a family resemblance to 'contingent convertible bonds' . . . . Among other features, 'co-cos' promote shareholder monitoring of managerial risk-taking by providing a credible threat of dilution in the event of financial distress . . . .').

255. See Kaal, supra note 14, at 321 (discussing how Credit Suisse and other SIFIs have already issued contingent convertible bonds to the public); see also supra note 134 and accompanying text (noting that Barclays has already issued contingent convertible bonds to its executives).

256. See supra Part V.C (explaining how SIFI executives might be led to manipulate triggering events if they are compensated with contingent convertible bonds).
equity infusion to avoid dilution of their own equity stake. If executives’ compensation packages do not include a large equity portion, managers have no incentive to avoid an equity infusion to preserve the value of their own equity. Even if a substantial portion of executives’ compensation remains equity-based, the early conversion of contingent convertible bonds affects not only the debt portion of executives’ compensation but also the value of the equity portion after the contingent convertible bonds portion is converted. The lower value of the equity portion after the conversion of the contingent convertible bonds, in combination with the negative impact of the CCBS conversion on the existing equity of the respective entity, makes it less likely that executives will fear the effect of raising additional equity, such as a negative impact on price and the dilutive effect.

Because contingent convertible bonds are addressed to the entire equity base and would, according to some proposals, become a (mandatory) feature of SIFIs’ balance sheets, the dilution and the effect on managers could be greater with convertible equity-based pay. However, the effect of the

257. See Bebchuk & Spamann, supra note 1, at 272 (“[T]he holder of [an out-of-the-money] option will be indifferent between a stable stock price and further losses of any magnitude. On the upside, only very large stock price gains will yield a positive payoff for the option holder.”).

258. See supra Part V.B (examining the consequences of an early conversion of contingent convertible bonds and the effects of such conversion on executives’ risk-taking). Opportunism may lead managers to increase the risk profile of the entity after conversion, enabling them to regain the lost equity value in their portfolio. On the other hand, an increase in the risk profile would affect the ability of the entity to obtain other forms of financing such as bridge loans. The increase in the risk profile after conversion is also unlikely because the early conversion would probably increase CCBs investors’ and shareholders’ monitoring of the entity. See Executive Compensation, supra note 19, at 11 n.18 (“Among other features, ‘co-cos’ promote shareholder monitoring of managerial risk-taking by providing a credible threat of dilution in the event of financial distress, because of the automatic conversion of a significant amount of debt into equity.”).

259. See id. at 11 (“The equity will convert into subordinated debt based on the value of the converted equity as of a period prior to the conversion moment . . . . This mechanism both imposes losses on senior management for deterioration in the firm’s financial condition while giving it a significant stake in avoiding further deterioration.”).

260. See id. at 11 n.18 (“Assuming that anti-dilution protection is scrubbed out of managerial compensation contracts, the dilution threat from co-co’s should also directly affect management behavior.”).
conversion from equity to debt will largely depend on the trigger design. There is currently no consensus on the ideal trigger design for convertible securities. Depending on the trigger design, a mandatory conversion of senior managers’ equity into subordinated debt on a valuation basis may not lead to a significant loss if the trigger from equity to subordinated debt comes after a likely significant loss in the share price of the entity. Although mandatory contingent convertible bond issuances with predefined triggering events could have better corporate governance outcomes, it is noteworthy that the contingent convertible bond market has evolved without mandatory issuances. Also, the Fuld Problem is only one of many corporate governance concerns pertaining to SIFIs. The approach in this Article is broader. The issuance of contingent convertible bonds to the public has great potential to improve several aspects of corporate governance in SIFIs.

261. See Kaal, supra note 14, at 300–01 (describing how different trigger designs pose different problems); see also Henkel & Kaal, supra note 149, at 251–57 (showing the uncertainty and risk involved in the different trigger designs).

262. See Executive Compensation, supra note 19, at 16 (“[A] mandatory conversion of senior managers’ equity into subordinated debt on valuation basis . . . [can] impose[] an immediate loss but . . . also [may] preserve[] incentives to prevent further deterioration of the firm.”).

263. See Kaal, supra note 14, at 317 (“The experimentation with different CCS rules and a mixture of market solutions, private ordering, and mandatory rules in different jurisdictions could help avoid a ‘stable rule’ and permit dynamic regulation.”).

264. See id. at 308–12 (discussing the evolving market in contingent capital).

265. See supra note 251 and accompanying text (defining “the Fuld Problem”).

266. Another similar problem is that “the incentives originating from corporate governance controls may not work in SIFIs.” Kaal & Henkel, supra note 14, at 242.

267. The common denominator between Gordon’s proposal and the proposal in this Article is the avoidance of SIFI resolution and distress in the financial sector. Compare Executive Compensation, supra note 19, at 2 (focusing instead on the ever-widening gap between executives’ interests and those of nonmanagerial shareholders), with supra notes 240–44 and accompanying text (examining contingent corporate bonds’ effect on risk-taking and income inequality).

268. See Kaal, supra note 14, at 321–22 (“While commercial motives are certain to have played a major role in the CCS issuance by Credit Suisse and
convertible bonds in executive compensation can solidify, and in
some cases, intensify the positive effects on corporate
governance.269

C. The Impact of Ownership Characteristics on Trigger Design

The literature on trigger designs of contingent capital focuses
on the efficient calibration of triggering events.270 The efficient
calibration of triggering events is central to the design of contin-
gent capital because the trigger affects if and when the
conversion of contingent convertible bonds takes place.271 The
efficient conversion of contingent convertible bonds into equity
has substantial implications for the effectiveness of contingent
convertible bonds and their ability to make financial institutions
safer.272

The literature on contingent capital trigger designs is largely
silent regarding the effect of trigger designs on corporate
governance.273 Contingent convertible bonds issued to
Barclays, the use of contingent capital to avoid public sector support and to
compensate executives could be a first step towards corporate governance
reform from within, with potentially larger implications.”).269

269. See supra Part V.B (explaining how contingent convertible bonds that
specifically include an early trigger design may optimize the effects on corporate
governance).

270. See Henkel & Kaal, supra note 149, at 251–52 (“The efficient
 calibration of triggering events is central to the design of contingent
capital. . . . Scholars discuss various trigger events that may be categorized as
follows: (1) transactional triggers, (2) automatic triggers, (3) statutory triggers,
and (4) regulatory triggers.”); Kaal, supra note 14, at 298–99 (“The optimal
design of CCS has been the subject of a long academic debate. Unresolved
questions include design features of CCS and the calibration of design features,
the mandatory or voluntary nature of contingent capital, the objectives of CCS,
market evolution, and the volume of CCS issuance, among others.”).

271. See Henkel & Kaal, supra note 149, at 251 (“The efficient calibration of
triggering events is central to the design of contingent capital. The optimal
design for a trigger event that converts debt into equity is unclear.”).

272. See Kaal & Henkel, supra note 14, at 224–25 (“Contingent capital is the
predefined conversion of financial institutions’ debt securities upon a triggering
event into equity securities. Pending contingent capital proposals are expected
to make financial institutions more resilient and avoid a future financial
crisis.”).

273. See Kaal, supra note 14, at 300–01 (discussing the potential benefits
and drawbacks of trigger designs, but not addressing trigger designs’ effect on
executives\textsuperscript{274} place a new emphasis on the impact of ownership characteristics on trigger designs. Trigger designs that may work well in financial institutions with the traditional mix of debt-holders and shareholders may be suboptimal if executives also hold debt instruments in the form of contingent convertible bonds.\textsuperscript{275} The nature of ownership of contingent convertible bonds may create different demands on the design features. Who owns the contingent convertible bonds can impact the efficiency, effectiveness, and corporate governance results of trigger designs.\textsuperscript{276} Recognizing that ownership characteristics can have an impact on the efficient design of contingent capital triggers can help adjust designs and avoid suboptimal outcomes.\textsuperscript{277}

\textbf{VII. Conclusion}

Contingent convertible bonds in executive compensation are not a mere theoretical concept. European SIFIs have started to add contingent convertible bonds to executive compensation packages.\textsuperscript{278} Path dependencies could make it difficult to adopt governance-improving elements in executive compensation policies in the United States.\textsuperscript{279} Contingent convertible bonds display many commercially attractive features that could help establish these hybrid securities in the compensation packages of executives in the United States. Like other debt securities in executive compensation, contingent convertible bonds can lower income inequality and incentivize the long-term and sustainable corporate governance).

\textsuperscript{274} See supra note 134 and Part V.A (reviewing Barclays’s issuance of contingent convertible bonds to executives).

\textsuperscript{275} Kaal & Henkel, supra note 14, at 301.

\textsuperscript{276} See Henkel & Kaal, supra note 149, at 252 (“Constituents favor trigger designs in accordance with their own utility preferences.”).

\textsuperscript{277} Adjusting the design of CCBs to take ownership characteristics into account could complicate this analysis, which already deals with a substantial number of parameters. However, the design of CCBs should not be compromised in order to avoid complexity.

\textsuperscript{278} See supra note 17 and accompanying text (discussing Barclays’s use of contingent convertible bonds in executive compensation).

\textsuperscript{279} See supra note 191 and accompanying text (identifying path dependence as an obstacle to the adoption of contingent convertible bonds in the United States).
Additionally, an early trigger design for contingent convertible bonds in executive compensation can help further improve governance shortcomings in SIFIs. Contingent convertible bonds with an early trigger design enable earlier signaling of default risk; they provide increased incentives for monitoring by creditors and shareholders as well as incentives for executives to lower their risk-taking.

280. See supra note 189 and accompanying text (explaining that if contingent convertible bonds replace equity-based executive compensation, contingent convertible bonds may lower income inequality).

281. See supra notes 221–23 and accompanying text (expanding on the benefits of an early trigger design).

282. See supra note 247 and accompanying text (discussing how early trigger designs can incentivize creditors and shareholders to monitor executives’ risk-taking).
Illuminating Innovation: From Patent Racing to Patent War

Lea Shaver*

Abstract

Patent law assumes that stronger protection promotes innovation, yet empirical evidence to test this “innovation hypothesis” is lacking. This Article argues that historical case studies hold unique promise to provide an empirical foundation for modern patent policy. Specifically, this Article uses the history of patents surrounding the light bulb to examine a recently articulated theory of “patent racing” as a justification for patent protection. Thomas Edison’s experience confirms that Mark Lemley’s racing model has substantial descriptive merit. Yet this case study also reveals the limits of the patent racing model. Looking past the initial finish line of patent filings to later litigation, the competition looks less like a race and more like a war.

* J.D., Yale Law School; M.A., University of Chicago; Associate Professor of Law, Indiana University Robert H. McKinney School of Law. I would like to thank Jack Balkin, Christopher Beauchamp, Joseph Blocher, Samuel Buell, Michael Carrier, Dennis Corgill, Eric Dannenmaier, Rochelle Dreyfuss, Laura DeNardis, Elizabeth Glazer, Laurence Helfer, Lauren Henry, Paul Israel, Eric E. Johnson, Mark Lemley, David S. Levine, Gerard Magliocca, Robert Mensel, Emily Morris, Adam Mossoff, Lisa Larrimore Ouellette, Jerome Reichman, Jake Sherkow, Rose Villazor, Carleton Waterhouse, and R. George Wright for their particularly helpful comments. Portions of this Article were presented at Duke Law School’s Legal Theory Workshop, Hofstra Law School’s Junior Faculty Forum, the Indiana University Maurer School of Law, the Information Society Project at Yale Law School, the Intellectual Property Scholars Conference at Stanford University, and on the Stanford University radio show Hear say Culture. This Article is made available to the public under the terms of a Creative Commons—Attribution license. A digital copy may be downloaded at no charge from the Social Sciences Research Network at http://ssrn.com/abstract=1658643. Copies and reprints should acknowledge first publication as: Lea Shaver, Illuminating Innovation: From Patent Racing to Patent War, 69 WASH. & LEE L. REV. (2012), available at http://law.wlu.edu/lawreview/ and http://ssrn.com/abstract=1658643.
The Article then proposes a new model of “patent warfare” resembling the board game Risk. In the game, competing parties assemble strategic assets, then turn to battle their rivals for world domination. Similarly, innovative technology companies assemble patent portfolios—initially for defensive purposes in the context of a dynamic and competitive field. As an industry matures, however, dominant players convert their shields into weapons to eliminate their competition. Just as nineteenth-century companies in the early electrical industry battled to control the light bulb, a new patent war is now emerging to control the smartphone. This anticompetitive endgame diminished next-generation innovation in electric light and now threatens the future of innovation in mobile computing.

A new appreciation of patent warfare should prompt increased skepticism of the “innovation assumption” at the heart of patent law. Patent protection clearly provides short-term benefits to innovation, but it also produces unanticipated long-term costs to competition and next-generation innovation. Further empirical research is needed to ascertain whether the benefits outweigh the costs or vice versa and what tweaks to the patent system might allow us to continue to capture the benefits while lowering the costs. In this continued effort, historical case studies will prove particularly helpful because they permit insight into the complex workings of patent law on an industry over a longer time horizon, revealing not only the short-term benefits but also the long-term costs.

Table of Contents

I. Introduction ................................................................... 1893

II. Testing the Innovation Hypothesis ............................... 1896
   A. Patents and Innovation ........................................... 1896
   B. Case Study as Methodology .................................... 1900
   C. Law and the Light Bulb .......................................... 1907

III. Dynamics of Inventive Races ........................................ 1911
   A. The Theory of Patent Racing ................................. 1913
   B. Support for Patent Racing Theory ......................... 1915
   C. Problems for Patent Racing Theory ....................... 1921
I. Introduction

More than a century after its introduction, the incandescent light bulb remains the defining icon of invention. Elegantly simple, this humble device was nonetheless a transformative technology. Just as spreadsheet software drove demand for personal computers and email clients drove demand for Internet service, the incandescent lamp was the “killer app” that drove demand for electricity. In the process, the light bulb paved the way for every other electronic appliance, including the radio, the television, and the personal computer. No less than the computer, the light bulb revolutionized our world.

This Article sheds light on the role that patent law played in the light bulb’s emergence. The story told here is not a conventional tale of great inventors, bright ideas, and the inevitable march of scientific progress. It is a story of corporate maneuvering and high-stakes litigation, as Thomas Edison and

---


2. See Arthur A. Bright, Jr., The Electric-Lamp Industry: Technological Change and Economic Development from 1800 to 1947 8–11 (1949) (explaining the importance of the incandescent lamp to the emergence of an electrical supply system, scientific and public interest in electricity, the radio, and economic growth through World War I generally). Edison himself played a great role in developing and building out the technology and infrastructure for electricity distribution. Of Edison’s patents related to electric light between 1881 and 1883, only 37% were related to the light bulb. The majority of his patents related to other elements of the electric power delivery system, including voltage regulation and engine governors (24%), transmission and distribution (14%), and dynamo design and operation (12%), among other aspects. Paul Israel, Claim the Earth: Protecting Edison’s Inventions at Home and Abroad, in Knowledge Management and Intellectual Property: Concepts, Actors and Practices from the Past to the Present (Graham Dutfield & Stathis Arapostathis eds., 2013) (forthcoming) [hereinafter Israel, Claim the Earth] (on file with author).
his competitors employed patents as weapons in their battle to dominate the electrical industry. Although famous for his workshop talents, Edison would be better understood as an early Bill Gates, his business success built on a keen appreciation of the power of aggressive intellectual property strategy and the frailty of antitrust law.\textsuperscript{3}

This story holds particularly important lessons for the modern information technology (IT) industry, where an epic patent battle is now underway.\textsuperscript{4} Although early incandescent

\begin{footnotesize}
\begin{itemize}
\item[3.] See generally William H. Page & John E. Lopatka, The Microsoft Case: Antitrust, High Technology, and Consumer Welfare (2007) (examining the antitrust case against Microsoft in the context of shifting debates about antitrust law); see also Israel, Claim the Earth, supra note 2 (explaining that Edison used patents as part of a larger strategy to establish industry dominance through his early work in the telegraphy industry, strategically identifying certain technological features of the electrical system as choke points for where to achieve “the commercial monopoly of the general [electricity] distribution business”).
\end{itemize}
\end{footnotesize}
light bulbs and modern smartphones are very different technologies, they have much in common from the perspective of patent strategy. Understanding the patent war over the light bulb can help us understand what is at stake in the current war over the smartphone.5

Edison’s lamp can also shed light on one of the most challenging questions asked in patent scholarship today. Is patent protection on the whole helpful or harmful to innovation? And how might the law be tweaked to achieve an even better balance between protection and competition? The light bulb can illuminate these questions by testing existing theory and helping to improve on it. For more than two centuries, U.S. patent law has assumed that patents promote innovation.6 Yet for all the profound impact that patent law has on industry and society, we have little empirical evidence to test this “innovation assumption.” In fact, there are good reasons to believe that patents may also impede innovation by creating barriers to competition.7 The assumption should more prudently be treated as a hypothesis in need of empirical testing.

Part II of this Article argues that careful case studies, particularly those drawn from history, may best advance our understanding of patent law’s impact on innovation. The remainder of the Article develops such a case study, focusing on Thomas Edison’s efforts to market an incandescent light bulb. Part III takes up a recently articulated theory of patent racing, using the case study to test and refine the racing model. Part IV explores the lessons of this case study for the innovation hypothesis. Here, a new model of “patent warfare” is advanced, which not only explains the experience of the early electrical

5. See infra Part II.A (discussing Judge Posner’s decision to dismiss the recent suit between Apple and Motorola for lack of an adequate, acceptable remedy).


7. See infra Part II.A (discussing the economic and social costs of a patent system).
industry, but also reveals the high stakes of the new patent war now breaking out in the smartphone industry.

II. Testing the Innovation Hypothesis

A. Patents and Innovation

For more than two centuries, the notion that patent protection encourages technological innovation has remained the conventional wisdom. In 1830, Anglo-American legal philosopher Jeremy Bentham captured its essence: “In new inventions, protection against imitators is not less necessary than in established manufactures protection against thieves. He who has no hope that he shall reap will not take the trouble to sow.” A half-century later, the same notion is reflected as the popular wisdom of industrial Americans in Mark Twain’s *A Connecticut Yankee in King Arthur’s Court*. In this tale, a New England
mechanic from the 1880s travels back in time to Camelot, where he becomes a valued advisor to the English king. On the subject of technological innovation, Twain’s hero relates:

[T]he very first official thing I did, in my administration—and it was on the first day of it, too—was to start a patent office; for I knew that a country without a patent office and good patent laws was just a crab, and couldn’t travel any way but sideways or backwards.\textsuperscript{11}

As an empirical matter, however, the benefits of patent protection for innovation are far less certain. Scholarly skepticism about the anticipated benefits of patent protection also has deep roots.\textsuperscript{12} In the 1950s, the noted economist Franz Machlup advised Congress:

If we did not have a patent system, it would be irresponsible, on the basis of our present knowledge of its economic consequences, to recommend instituting one. But since we have had a patent system for a long time, it would be irresponsible, on the basis of our present knowledge, to recommend abolishing it.\textsuperscript{13}

Fifty years later, noted scholar Richard Posner lamented that economists have still not determined whether the benefits of intellectual property protection outweigh the costs.\textsuperscript{14}

---

\textsuperscript{11} Mark Twain, \textit{A Connecticut Yankee in King Arthur’s Court} 107 (1st ed. 1889). Credit is due to Mark Adelman for leading me to Twain’s work. For a more serious interview with Mark Twain suggesting that the Yankee hero’s statement reflects Twain’s own beliefs rather than simply words thrust upon a satirical strawman, see Mark Twain, \textit{Mark Twain: The Complete Interviews} 295 (Gary Scharnhorst ed., The University of Alabama Press 2006) (using historical evidence himself to suggest that patent protection encourages not so much invention itself as the diligent perfection of invention into a commercially practical technology). Twain was also a patent holder. \textit{See Improvement In Scrap-Books, U.S. Patent No. 140,245 (filed May 7, 1873) (issued June 24, 1873) (disclosing a self-pasting scrap-book).}


\textsuperscript{14} See, e.g., Richard A. Posner, \textit{Intellectual Property: The Law and Economics Approach}, 19 J. Econ. Persp. 57, 59 (2005) (“Unfortunately, economists do not know whether the existing system of intellectual property rights is, or . . . whether any other system of intellectual property rights would be, a source of net social utility, given the costs of the system and the existence
and legal scholars have recently devoted great theoretical attention to clarifying the ways in which patent protection may inhibit innovation, as well as incentivize it.\textsuperscript{15}

\begin{itemize}
Patent law and policy, however, continue to operate on the assumption that patents not only incentivize greater innovation, but that the patent incentive is substantial such that innovation would be greatly diminished if the incentive were removed or even weakened. By and large, the field of patent scholarship also treats this assumption as presumptively correct, while acknowledging that future research might one day call it into question. This willingness to assume—at least for the time being—that patents work as intended has the same practical effect as an evidentiary presumption: it shifts the burden of proof onto the challengers of the conventional wisdom. There are virtues to adopting this assumption. One such virtue is enabling patent scholars to focus productively on calibrating less-entrenched aspects of patent doctrine and on suggesting reforms that may be implemented in the near future.

Embracing this assumption too fully, however, carries an unacceptable risk. Patent law is a tremendously influential aspect of modern economic regulation. What if its central assumption is wrong? So far we have justified the substantial social costs of patent protection—litigation costs, anticompetitive effects, higher prices for consumers, barriers to adoption of new technologies—on the basis that these sacrifices are worthwhile in order to achieve faster innovation. If patent protection offers much fewer benefits for innovation than widely supposed, the law is imposing these substantial costs needlessly. And if patent protection actually has a net negative impact on innovation, the real costs to economic growth may be appalling in scale. Because the risks of being


16. Indeed, evidence suggests that patent law’s reliance on the innovation assumption has, if anything, increased in recent years. See Gaia Bernstein, In the Shadow of Innovation, 31 CARDOZO L. REV. 2257, 2259–60 (2010) (documenting an increasing use of the phrase “innovation” in intellectual property opinions beginning in the 1980s).

17. See, e.g., Lisa Larrimore Ouellette, Do Patents Disclose Useful Information?, 25 HARV. J.L. & TECH. 531, 532–35 (2012) (surveying the mixed empirical support for the innovation hypothesis without taking sides and proceeding to focus on a more politically open question: What emphasis should patent doctrine place on disclosure as a central policy goal?).

18. See, e.g., BESSEN & MEURER, supra note 15, at 6–7 (comparing the economic impact of tangible property law to economic incentives potentially created by patent laws).
wrong are so great, the innovation assumption should more prudently be treated as the innovation hypothesis, in need of further empirical testing.

B. Case Study as Methodology

A major challenge for researchers seeking to test the innovation hypothesis is the complexity of the phenomenon. Changes in patent policy take time to play out in the marketplace and affect different industries in different ways. At the macroeconomic level, it can be very difficult to separate out impacts due to changes in patent policy from those due to other causes. Cross-national comparisons are further complicated by the trend toward harmonization of patent law, which limits opportunities for empirical evaluation of differing regimes. Even defining “innovation” as a quantifiable outcome has proven elusive. These challenges make it difficult to reliably measure patent law’s impact through large-scale quantitative analysis.

19. See, e.g., Burk & Lemley, supra note 15, at 1576–78 (explaining how innovation and therefore the impacts of patent policy differ based on the industry studied).

20. See, e.g., Bessen & Meurer, supra note 15, at 11 (“Many interrelated factors can influence R&D spending, innovation, and the resulting social welfare, so it is difficult to disentangle these to determine the independent influence of patents.”).

21. Economist Petra Moser has creatively approached this problem by analyzing evidence of innovation predating the modern trend toward international harmonization of patent law. See Petra Moser, How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World’s Fairs, 95 AM. ECON. REV. 1214, 1231–33 (2005) (concluding that, because patent protection appears to enhance innovation in some industries but impede it in others, variance in patent regimes across countries may result in optimal global results by encouraging the development of comparative advantages).

In my first article, I explored some of the problems inherent in macroeconomic measures of innovation, particularly the problematic reliance on rates of patent filing as the measure of innovation.23 Rather than relying on numerical data, the article suggested that careful case studies are likely to prove more enlightening at the present stage of research.24 An empirical case study steps back from the overwhelming web of data to pick up the thread of a single “case,” which might be the life of a particular patent, technology, company, or industry.25 The

---


24. See id. (suggesting that researchers focus on “high-value innovation” or narrow the focus of their study to one field rather than attempt to measure innovation across an entire society).

25. The empirical “case study” has no necessary relationship with the “case method” widely used in legal research and education. Literature on case studies as an empirical research methodology has been careful to distinguish its subject from the quite different use of cases as a teaching tool in law and business. See, e.g., Robert K. Yin, Case Study Research: Design and Methods 4–5 (4th ed. 2009) (describing a “teaching case” as one designed to establish a framework for discussion and debate among students and noting that the focus on his work is on the use of case studies as a research method). In contrast to the case method, the case study need not take a particular legal dispute or judicial opinion as its focus of inquiry. Of course, the “case” that is the object of empirical study might be a legal case, such as the Supreme Court’s hearing of the Incandescent Lamp Case, 159 U.S. 465 (1895). But the “case” might also be a patent (such as Edison’s ‘898 Patent), a technology (the light bulb), a patentee (Thomas Edison), a firm (General Electric), or a sector (the early electrical industry). Any specific, complex, functioning thing may be a “case” for study. See Robert E. Stake, The Art of Case Study Research 2 (1995) [hereinafter Stake, CASE STUDY RESEARCH]. My interest in understanding how patent law influences technological innovation led me to define my case around one technology—the early light bulb. Many different legal disputes, patents, patentees, and firms influence innovation in any particular field of technology. Therefore, focusing on any one particular opinion, patent, inventor, or company would likely narrow the inquiry too much and miss essential aspects of the larger innovation dynamics. A concrete technology (such as the light bulb or the smartphone) appears to be the simplest unit of investigation that allows us to seek answers...
researcher follows the path of that one subject to see where it leads and attempts to interpret its lessons. 26 A leading articulation of the case study methodology in the social sciences proposes that “case studies are the preferred method when (a) ‘how’ or ‘why’ questions are being posed, (b) the investigator has little control over events, and (c) the focus is on a contemporary phenomenon within a real-life context.” 27 These three conditions accurately capture the challenges associated with asking how patent law impacts technological innovation.

To be sure, a case study of any one technology has only limited utility for answering our larger questions about innovation. The dynamics of the particular technology selected may not be representative of all sectors or even of the general trend. Multiple data points are needed. It should also be acknowledged that case studies are fundamentally acts of interpretation. 28 To guard against investigator bias, these interpretations should be subject to peer review, public debate, to questions about technological innovation and diffusion, without hiding the complex operations of patent law in the real world.

26. Within the substantial social scientific literature employing and commenting on case study methodology, some researchers prefer case studies because of normative commitments that achieving objective, accurate explanation of complex social phenomena is unattainable or undesirable. From these researchers’ point of view, there is no “truth” in social science, only “interpretation.” See Robert E. Stake, Case Studies, in THE SAGE HANDBOOK OF QUALITATIVE RESEARCH 435, 439 (Norman K. Denzin & Donna S. Lincoln eds., 2d ed. 2000) [hereinafter Stake, QUALITATIVE RESEARCH] (describing some of the criticisms of case study research regarding their value to generalizability). This is not my point of view, nor is it my motivation for adopting this methodology. Quite to the contrary, the present study assumes that there is a correct answer to the question, “Does patent protection promote technological innovation?” and that social science can help us find that correct answer. Of course, the correct answer is unlikely to be a simple “yes” or “no.” Rather, the accurate answer to this question is likely to be complex, to be filled with caveats, and to inevitably have some margin of uncertainty. The correct answer is likely to be: “Under certain circumstances, yes, patent protection does appear to promote innovation; and under other circumstances, no, patent protection appears to be counterproductive.” Careful case studies can help illuminate these complex dynamics and help to form provisional answers, which may in turn be further tested and verified through quantitative analysis.


28. See, e.g., Stake, CASE STUDY RESEARCH, supra note 25, at 8 (“According to one highly respected writer on qualitative studies, Fred Erickson, the most distinctive characteristic of qualitative inquiry is its emphasis on interpretation.”).
and independent studies of other cases, a phenomenon referred to in the case study literature as “triangulation.”29 Over time, however, the collective work of many scholars may build a robust body of case studies, which could provide a strong empirical basis to support or call into question key predictions and assumptions about the impact of patents on innovation. Even before this critical mass is reached, individual case studies can offer a helpful reality check on our theoretical models of how patent incentives work. Do theoretical predictions that seem to make sense in the abstract actually fit with what can be observed in the historical record? Or is there a significant gap between theory and fact, at least in the particular cases available to date?

The present work builds upon a body of empirical case studies that have already been developed by patent scholars looking at a number of modern technologies. For example, Lisa Ouellette has examined the disclosure function of patents in spreading scientific knowledge through the nanotechnology sector.30 Graham, Merges, Samuelson, and Sichelman have drawn findings about the role of patents in assisting startups, drawing on surveys of more than one thousand high-tech entrepreneurs in sectors ranging from biotechnology to software.31 Samuelson and Scotchmer have examined the economic impact of reverse engineering in the semi-conductor and software industries.32 Health economist Heidi Williams has used interviews and statistical analysis to document a negative impact

29. See id. at 111–15 (describing the triangulation protocols to “increase credence in the interpretation” and “to demonstrate commonality of an assertion”).

30. See, e.g., Ouellette, supra note 17, at 552–71 (drawing on a survey of more than 200 nanotechnology researchers and case studies of selected nanotechnology patents to evaluate the impact of patents on scientific disclosure).


on downstream innovation for gene sequences covered by intellectual property.\textsuperscript{33} Fiona Murray and colleagues documented a similar negative impact in a comparison of laboratory mouse varieties subject to differing intellectual property restrictions.\textsuperscript{34} The above case studies each have a quantitative component, focusing on a particular set of data to answer particular questions about their cases. These studies illustrate the point that the empirical case study need not be purely qualitative.

A more recent trend in patent scholarship is to explore historical, as well as current, case studies.\textsuperscript{35} Thus, Gerard Magliocca has harvested lessons from the behavior of patent


\textsuperscript{34} See Fiona Murray et al., \textit{Of Mice and Academics: Examining the Effect of Openness on Innovation} 4–5 (Nat'l Bureau of Econ. Research, Working Paper No. 14,819, 2009), \textit{available at} http://www.nber.org/papers/w14819 (determining that research using mice species subject to greater intellectual property restrictions resulted in less “downstream” innovation than research using less-restricted species).

trolls with regard to nineteenth century agricultural machinery.\textsuperscript{36} Adam Mossoff has delved into patent thickets surrounding the early sewing machine industry.\textsuperscript{37} Kara Swanson uncovered the hidden dimensions of “the corset case” to offer a gendered perspective on patent law.\textsuperscript{38} And Christopher Beauchamp picks up the telephone as an object of cultural and legal disputes over inventorship.\textsuperscript{39}

So far, historical scholars have not typically focused their attention on the innovation hypothesis, and innovation scholars have not generally favored historical case studies.\textsuperscript{40} Yet historical case studies may be critical to testing patent law’s innovation hypothesis. The unique advantage of a historical case study is that the passing of time allows researchers to observe both short-term and long-term results. This is crucial because the full impact of patent protection plays out over several decades. In the near term, incentives are perceived, a technology is developed, patents are issued, and products come to market. In the longer term, the new industry matures, litigation ensues, the first-generation patents expire, and second-generation innovation takes place. When studying cutting-edge technological sectors, our analysis is necessarily limited to the near term; the longer-term effects remain in the future, out of sight. This shortened horizon may introduce a systematic blind spot in our


\textsuperscript{40} Indeed, historical inquiry comes with challenges for the researcher interested in assessing the impact of patent law on innovation. Quantitative data may not have been recorded and might be impossible to gather now. Participant observation and interviewing are similarly unavailable as methodologies.
understanding of patent law’s true impact. The benefits of patent protection are visible immediately in the form of enhanced research and development spending, cutting-edge technologies, and new products on market. The long-term costs to competition, access, and next-generation innovation may take place later, out of sight.41

Taking a historic case study such as the light bulb allows us to examine the impact of patent law in the fullness of time—not just at the initial stage of research and development, but throughout the term of protection and past the expiration of first-generation patents. This greater perspective can allow researchers to observe long-term dynamics that would have remained hidden in the future when studying a more current technology. In the case of the light bulb, these long-term dynamics included massive litigation,42 the consolidation of a previously competitive industry,43 and a half-century lull between when incandescent-light technology was first patented and when it finally became widely accessible to the American public.44 Perhaps it should come as no surprise, then, that patent scholars centrally concerned with the innovation problem have heavily relied on the very limited historical research already at hand.45

41. See Bernstein, supra note 16, at 2311 (distinguishing the first step of innovation from the second step of diffusion and concluding that intellectual property law has been insufficiently concerned with the latter).
43. See id. at 155–62 (describing the consolidation of the incandescent lamp industry following the upholding of Édison’s light bulb patent).
44. See generally DAVID E. NYE, ELECTRIFYING AMERICA: SOCIAL MEANINGS OF A NEW TECHNOLOGY, 1880–1940 (1992) (tracing the history of electricity’s spread across America from its initial development in the early 1880s to its broad spread to ordinary Americans only in the 1930s and 1940s).
45. See, e.g., BESSEN & MEURER, supra note 15 (discussing patents as a form of property right and concluding that patents fail as property in rewarding innovators and encouraging investment in innovation); BOLDRIN & LEVINE, supra note 15 (analyzing the impact of intellectual monopolies created through patents and copyrights on the economic motivations for innovation); Mark A. Lemley, The Myth of the Sole Inventor, 110 Mich. L. Rev. 709 (2012) [hereinafter Lemley, Sole Inventor] (examining the problem of simultaneous or near-simultaneous invention as it relates to patent law theory).
History remains the best evidence available to us, in a world where controlled experiments are impossible. This promising body of evidence deserves more careful and extensive investigation.

C. Law and the Light Bulb

Many different technologies might have been the focus of an attempt to historically examine the innovation hypothesis. Why choose the light bulb? Robert E. Stake has distinguished three approaches to case study research: “intrinsic, instrumental, and collective.” First, a researcher may choose to undertake a case study for intrinsic interest: not to understand a general phenomenon or build theory, but because the case itself is of inherent interest. Alternatively, a case may be explored with the primary goal of gaining an understanding of larger questions removed from the particular instance—the instrumental case study. My aim in undertaking this case study is decidedly instrumental. My research is driven by the desire to understand how patent law impacts technological innovation and diffusion. Many different technologies could have served that instrumental purpose; the light bulb offered simply one convenient starting point. In the future, several thoughtfully selected inventions might be explored through a collective case study.

For many Americans, however, the case of Thomas Edison and the light bulb holds substantial intrinsic interest. People are interested in knowing more about the invention of the light bulb, for its own sake. Choosing this particular case thus offered me a vehicle to explore the innovation hypothesis, with the hope of finding a wider audience. A second advantage of the enduring intrinsic interest in the light bulb and Thomas Edison is that a

46. Stake, Qualitative Research, supra note 26, at 438.
47. See id. at 437 (noting that the case is undertaken not “because the case represents other cases or because it illustrates a particular trait or problem, but because, in all its particularly and ordinariness, [the] case itself is of interest”).
48. See id. (“The case is of secondary interest, it plays a supportive role, and it facilitates our understanding of something else.”).
49. See id. (defining the collective case study as an instrumental study extended to several cases, which are selected as a sample to be representative of a broader class of phenomena the researcher wants to understand).
substantial body of scholarship already exists on this case. This Article draws significantly on that existing scholarship to look for answers to a previously neglected research question. The availability of a wealth of primary and secondary materials on this case also means that researchers who share my interest in the innovation hypothesis have access to the same resources. Others may interpret these materials independently, critique my conclusions, and offer alternative explanations: such “triangulation” is an important check on case study work that seeks to answer an empirical question.

Discussion of Edison’s work seems an almost inescapable element of any treatise on innovation. Yet the Great Inventor has so far attracted only a little attention from legal scholars. The Supreme Court’s opinion in the *Incandescent Lamp Patent* case is familiar to many patent attorneys, students, and scholars. It is excerpted in several patent casebooks and makes regular appearances in law review footnotes. Yet the case’s

---


51. For the exception that proves the rule, see Barak Y. Orbach, *Prizefighting and the Birth of Movie Censorship*, 21 Yale J.L. & Human. 251, 251 (2009) (examining extensively Edison’s role in adopting self-regulation as a mode of content censorship in the early motion picture industry). Orbach’s interest was censorship and freedom of expression, yet his article’s insight into Edison’s business approach to motion picture technology is interesting from an intellectual property perspective as well. For instance, Orbach traces Edison’s tendency to publicly exaggerate the progress of an invention and his inability to accurately predict the profitability of a technology at the time of its invention. See id. at 275–76 (recalling that Edison often used the press to promote uncompleted inventions without foresight into the financial benefits of such publicity). Edison also receives several paragraphs of attention in recent works by Mark Lemley and Brian Love; these works are discussed in detail in Part III of this Article.


53. The case quickly became known by the subject matter of the patent, rather than the parties’ names. See, e.g., Walter Malins Rose, *Notes on the United States Reports: A Brief Chronological Digest of All Points*
historical context, the litigation’s implications for the electrification of America, and the case’s potential larger lessons for innovation policy remain only dimly unappreciated. Indeed, the case name itself is a misnomer. There is no single patent corresponding to the incandescent lamp. Thomas Edison himself took out more than 100 patents on technical components of the incandescent lamp; his assistants and competitors also held many light bulb patents. Indeed, the patent at the center of the Supreme Court case was not Edison’s but that of a rival. More often, we invoke Thomas Edison and his light bulb in passing—treating Edison as an icon, rather than as a subject of inquiry. 
This Article seeks to go further, examining the invention, patenting, marketing, and litigation of the incandescent light bulb through the close lens of patent racing theory.

Before beginning this deeper inquiry, however, a brief introduction is in order. Many of us were taught as children that Thomas Alva Edison invented the light bulb. The full story is more complicated. A primitive light bulb had first been demonstrated more than seventy years before Edison’s work began. In the intervening decades, crucial improvements in vacuum technology and electricity generation were developed that made electric light a newly promising technology. Beginning in 1878, Edison led a team of engineers who worked together to produce one of the first commercially practical systems of interior lighting. This team, however, had serious competition in the laboratory and the marketplace. The competition was so intense, in fact, that for many years after greater depth, calling into question the iconic image. See id. at 722–23 (“It seems clear, however, that Edison did not ‘invent’ the light bulb in any meaningful sense.”). Lemley also uses the Incandescent Lamp Patent case as an illustration of the doctrine of enablement in the context of follow-on innovation. See Lemley, Economics of Improvement, supra note 53, at 1002–03 (discussing the court’s decision to invalidate Sawyer and Mann’s patent infringement claim against Edison on the grounds that Sawyer and Mann’s patent for the use of any vegetable fiber as a filament did not preclude Edison from making further investigation using bamboo).

56. A number of works aimed at children present Thomas Edison as an archetype of the heroic inventor See, e.g., Laurie Carlson, Thomas Edison For Kids: His Life and Ideas: 21 Activities (2006); Inez Nellie McFee, The Story of Thomas A. Edison (1922); William H. Meadowcroft with Autobiographical Note by Mr. Edison, The Boy’s Life of Edison (1911); Francis Rolt-Wheeler, Thomas Alva Edison (1915).

57. See Bright, supra note 2, at 36 (“One of the first demonstrations of electrically induced incandescence was made in 1802 . . . .”).

58. See id. at 38–42 (discussing the contribution energy advancements and vacuum developments had on the ultimate success of incandescent lighting).


60. At the exhibitions of the International Electrical Congress in Paris in 1881—held after Edison had filed several patents in the field but before he had his first paying customers—inandescent lamps were displayed by Swan, Lane-Fox, Maxim, and Edison. Thomas P. Hughes, Networks of Power: Electrification in Western Society, 1880–1930, at 50 (1983) [hereinafter Hughes, Networks of Power].
light bulbs began to be widely manufactured and sold, it remained unclear who had won. Even today, British schoolchildren are taught that Joseph Swan, not Thomas Edison, won the race to invent the light bulb. Once electric light was ready for commercialization, many companies competed with Edison’s in the marketplace. Compared to his rivals, however, Edison was particularly successful in litigating his portfolio of patents related to electric light technology. Edison leveraged his patent portfolio and courtroom successes to obtain mergers or market sharing agreements with most of his competition. A fifteen-year litigation campaign advancing claims based on a number of patents produced a stunning consolidation of a previously competitive market for light bulb manufacture. By 1910, General Electric—the successor to Edison’s electric light companies—would control 97% of all light bulbs sold in the U.S. market. The remainder of this Article examines both the early patent race and the later patent war in greater detail.

III. Dynamics of Inventive Races

One contribution historical case studies can make is to test existing theoretical models of patents’ roles in incentivizing innovation. Part III of this Article uses the case study of the light
bulb to test, critique, and refine one theory of “patent racing,” which has been offered as one account of how patent protection may incentivize faster innovation.\footnote{For a fuller picture of variations on the theory of patent racing and related models offered to explain how patents incentivize faster innovation, see John F. Duffy, \textit{Rethinking the Prospect Theory of Patents}, 71 U. Chi. L. Rev. 439, 440–47 (2004) (“T]he race to claim patent rights becomes a race to diminish the patentee’s rents by dedicating the invention to the public sooner.”); Mark F. Grady & Jay I. Alexander, \textit{Patent Law & Rent Dissipation}, 78 Va. L. Rev. 305, 305–10 (1992) (identifying that the case law suggests that “patent protection applies only to inventions that signal large possibilities for improvement relative to the rent that a patent gives to the inventor”) (emphasis omitted); Edmund W. Kitch, \textit{The Nature and Function of the Patent System}, 20 J.L. & Econ. 265, 265–67 (1977) (arguing that the patent system increases the output from resources used for technological innovation).} As articulated by Mark Lemley, patent racing theory centers on an underappreciated fact that proves problematic for other theories of patent-motivated innovation: most inventions are not breakthroughs at all, but ordinary engineering work that achieved its goal just a short time before the rest of the competition.\footnote{Lemley, \textit{Sole Inventor}, supra note 45, at 710 (discrediting the “canonical story” of the individual inventor and suggesting that new technologies are oftentimes invented simultaneously by independent teams).}

In using the light bulb case study to test this theory, this Article draws upon a vast secondary literature describing the development of incandescent light. These historical accounts, however, predate patent racing theory, tend to discuss patents and patent law only tangentially, and are generally authored by writers more interested in Thomas Edison as a personality than in theoretical questions of innovation policy. The process of sifting through this literature is thus one of searching for needles in a haystack—finding those snippets of relevant information here and there in the corpus, historical asides that now enjoy a second life as data points completely unintended by their original documenters. In all cases, I have made a point to critically question whether the original author’s account of the historical “facts” should be treated as reliable or not. As much as possible, this account relies only on data that can be confirmed through reference to more than one source.

Held up to the empirical light of Edison’s bulb, Lemley’s patent racing model proves to have substantial merit as a descriptive account. The racing metaphor captures a significant,
and often overlooked, reality about the dynamics of innovation in the shadow of patent law and is helpful to understanding the development of the light bulb. In its present form, however, patent racing theory also has weaknesses. A closer look at the light bulb suggests that the better metaphor for understanding technological innovation may be neither a sprint nor a marathon but the board game Risk. Patent law does not provide clearly defined finish lines and criteria for judging winners, but rather a set of strategic resources for waging longer-term corporate battles. Ultimately, the experience of the light bulb leaves me less optimistic than Lemley that patent racing will fare any better as a normative justification for patent protection. Rather, a fuller understanding of the inventive race to control the light bulb should leave us even more skeptical of the innovation hypothesis.

A. The Theory of Patent Racing

In The Myth of the Sole Inventor, Mark Lemley explores the underappreciated dynamic of “simultaneous invention” and the problems it poses for the conventional economic justifications of patent protection. In case after case, Lemley’s article illustrates that multiple inventors working on the same technological problem have arrived at the same solution at nearly the same time. The archetypal case here is the telephone: Alexander Graham Bell and Elisha Gray reportedly filed their competing

---

68. I am indebted to Robin Feldman at the UC Hastings College of the Law for the invitation to think of patent rights not as property claims in the traditional sense but as chips to be presented at the bargaining table when shares of ownership in a technology are eventually being resolved. I first encountered this presentation of patent law as an audience member when she gave a plenary presentation at the 2009 Intellectual Property Scholars Conference, and it greatly shaped my thinking about this project and the present Article. That presentation has now evolved into an excellent book, see ROBIN FELDMAN, RETHINKING PATENT LAW (2012). My theory of patent warfare builds on Professor Feldman’s view of patent strategy, although the metaphor I have chosen—war, rather than bargaining—reflects my view of the dynamic as less collaborative and more destructive than her own.

69. See Lemley, Sole Inventor, supra note 45, at 710–11 (arguing for a change in patent theory to accommodate the reality of simultaneous invention).

70. See id. at 735 (concluding, after a historical analysis, that it is “rare” for one inventor to develop a “wholly new product that no one else achieved at roughly the same time”).
Simultaneous invention may also take a less obvious form, wherein a single product builds on incremental inventive contributions by so many different parties that it truly has no identifiable inventor.\textsuperscript{72}

The overwhelming presence of simultaneous invention, Lemley argues, calls into question the dominant theoretical justifications for patent protection.\textsuperscript{73} If invention is the product of ordinary engineering work rather than extraordinary labors of genius, why do we award such an enormous prize to the party that achieved it first, if only by a few days?\textsuperscript{74} Lemley then proposes an alternative theory that might better justify patent protection, which begins from the reality of simultaneous invention, rather than from the myth of the sole inventor.\textsuperscript{75} This “patent racing” theory proposes that patent incentives accelerate innovation by increasing the stakes of the race to invent.\textsuperscript{76} All parties in the competition know that the first-place winner will take away a great prize, and the runner-up may be shut out completely. The enhanced competition these heightened stakes engender, Lemley suggests, may motivate greater investment of resources and thus accelerate the overall pace of innovation.\textsuperscript{77}

\textit{The Myth of the Sole Inventor} actually includes the light bulb among its historical examples of simultaneous invention, along the incremental model.\textsuperscript{78} Lemley correctly characterizes Edison’s

\begin{flushleft}
\textsuperscript{71} See id. at 710, 720–22, 751.
\textsuperscript{72} See id. at 714–15 (claiming that the incremental process of invention often leads to recognition of the “first person to make one significant step in the chain” and ignorance of “the developments that precede and follow it”).
\textsuperscript{73} See id. at 737 (“The overwhelming presence of both independent invention and incremental contribution calls the basic incentive story into serious question.”).
\textsuperscript{74} See id. at 752 (“If two or more putative investors invest that money in research and development in an effort to beat each other to the market, all but the first to invent will have wasted that money . . . .”).
\textsuperscript{75} See id. at 750 (considering a patent system that recognizes the importance of cumulative innovation).
\textsuperscript{76} See id. at 753–54 (describing the benefits of a patent race, including the quicker progress of inventions, the possibility of a wide variety of approaches to solving a problem, and the improved quality of inventions that a deadline might encourage).
\textsuperscript{77} See id. at 753 (“[A] patent race should . . . cause inventions to be made sooner than they otherwise would be . . . .”).
\textsuperscript{78} Id. at 722–23.
\end{flushleft}
inventive contribution as “an incremental one: one in a long chain of improvements.” Indeed, the contributions of Edison’s research team took the form of a chain of improvements as well, with multiple patents filed on various aspects of the light bulb and its related technology as they were developed, ultimately building a robust patent portfolio surrounding electric light technology. A closer look at the development of the incandescent light bulb reveals strong support for some aspects of patent racing theory, and also suggests ways in which the framework requires complication and revision.

B. Support for Patent Racing Theory

First, the support: Thomas Edison was clearly in an inventive race. He had multiple contemporaneous rivals pursuing the exact same inventive goal: a commercially viable light bulb. According to Conot, at least thirty-one U.S. Patents in the area of incandescent light had already been granted and not yet expired. Edison thus began work on the incandescent light bulb in 1878, at a time when the field was already populated by competing inventors and prior patents. This competition continued after Edison entered the field, and multiple parties filed important patents in the United States and other jurisdictions. Edison ultimately began large-scale electric light

79. Id. at 723.
80. See BRIGHT, supra note 2, at 44 (listing six other inventors working towards the development of incandescent electric lighting).
81. See CHARLES BAZERMAN, THE LANGUAGES OF EDISON’S LIGHT 247 (1999) (“[I]n the 15 years before Edison was granted his first patent for incandescent light, 31 patents had already been granted in that area . . . .” (citing ROBERT E. CONOT, A STREAK OF LUCK 214 (1979))).
82. See FRIEDEL & ISRAEL, ELECTRIC LIGHT, supra note 59, at 7–8 (“As early as 1841, Frederick De Moleyns, an Englishman, received a British patent for an incandescent lamp using both carbon and platinum. In 1845 an American, J.W. Starr, not only patented two forms of incandescent lamp . . . but also traveled around England giving exhibitions and promoting his inventions.”); see also William Sawyer, Elec. Lamp, U.S. Patent No. 235,459 (filed Sept. 20, 1880) (issued Dec. 14, 1880) (claiming a particular mechanism for fitting an electric lamp to a bracket).
83. See BAZERMAN, supra note 81, at 247 (“[I]n the 15 years before Edison was granted his first patent for incandescent light, 31 patents had already been granted in that area . . . .” (citing ROBERT E. CONOT, A STREAK OF LUCK 214.
service in New York City in September 1882. By that point, several other companies were also mass-producing electric light bulbs—and wielding their own patent portfolios.

Offering further support for the validity of Lemley’s metaphor, these parties also perceived themselves to be in a race. They were acutely self-conscious about their competition, sought out every available bit of information about the other teams’ progress, and were generally quite concerned about who seemed to be ahead at any given moment. So were their

---

84. Edison’s Electric Light, N.Y. TIMES, Sept. 5, 1882, at 8.

85. See, e.g., Passer, The Electrical Manufacturers, supra note 42, at 191 (“Under [Hiram] Maxim, the [United States Electric Lighting Company] began production of incandescent-lighting plants for isolated use in 1880.”). In August 1881, various light bulb manufacturers, including the Swan, Maxim, and Lane-Fox teams, in addition to Edison’s, had demonstrated their wares at the Exposition Internationale de l’Électricité (Paris Exhibition). See Letter from Otto A. Moses, Edison Emp. (Aug. 26, 1881), in 6 The Papers of Thomas Alva Edison: Electrifying New York and Abroad 144–45 (Paul Israel et al. eds., 2007) [hereinafter PAPERS OF EDISON VOL. 6] (assuring the receiver of the letter that the Edison exhibit outshone all the rest). By 1882, several systems competed with Edison’s. See Letter from Charles Mott to Arnold White (July 5, 1882), in PAPERS OF EDISON VOL. 6, supra, at 584–85 (referring to competing providers of incandescent light installations, including Siemens, Swan, “and others”); see also Letter from Frank J. Sprague, a member of Edison’s research team, to Edward Johnson, another Edison employee (Sept. 21, 1882), in PAPERS OF EDISON VOL. 6, supra, at 655–57 (“Crompton informed me privately that the [Swan] Company were turning out 15,000 lamps a week, and I am inclined to believe him.”); Bazerma, supra note 81, at 247–48 (discussing the patents wielded by other manufacturers).

86. This was true from the first months of Edison’s work on incandescent light. See, e.g., Letter from E.W. Griffiths, counsel retained by Edison, to Thomas A. Edison (Oct. 30, 1878), in 4 The Papers of Thomas A. Edison: The Wizard of Menlo Park, 1878, at 657–58 (Paul Israel et al. eds., 1998) [hereinafter PAPERS OF EDISON VOL. 4] (“Mr. Lowrey directs me to send you a few hurried notes of a conversation had with Mr. Sawyer this morning.” The enclosure then details particular technological accomplishments in which Sawyer claims to have anticipated Edison’s team.) The very next day, Edison directed his attorney to conduct a thorough patent search. Letter from T.A. Edison to Lemuel Serrel (Oct. 31, 1878), in PAPERS OF EDISON VOL. 4, supra, at 658 (“Please make out a list of all US patents from the earliest ages to the present time on Electric light and send them for me [sic] Want them right off.”). Edison’s concern to keep apprised of his competition continued as the endeavor moved from the research stage to commercialization. A report from an Edison employee gives a sense of the competition research being conducted:

The Swan Company is pushing forward, and in this connection let me say that Edison must not abate the work on the lamp, Swan is
investors.\textsuperscript{87} The progress made by the team of William Edward Sawyer and Albon Man posed particular problems for Edison.\textsuperscript{88} Between 1880 and 1883, the competitors were embroiled in a hard-fought priority battle at the U.S. patent office over which team had been first to achieve a key insight.\textsuperscript{89} The dispute went against Edison, and the issuance of the Sawyer and Man patent in 1885—along with Westinghouse’s entry in the industry—set the stage for a decade of extensive and expensive litigation.\textsuperscript{90}

Given the existence of other competitors, one very plausible interpretation of this history is that Edison’s participation in this race merely accelerated the arrival of a commercially practical light bulb. Gas lighting was an extremely lucrative industry,\textsuperscript{91} and arc electric lighting was already finding commercial making continual advance, and has recently commenced making lamps by a new process, which are the finest lamps I ever saw. They are not yet public, but Crompton, who is now one of the Swan Directors showed me one. . . . This new lamp is claimed to have 100 ohm resistance hot, and it can easily be increased, and to require but .6 of an empere, and 36 vol-tamperes, for 20 candle power. I think perhaps this is a little high, but is a wonderfully fine lamp, and I think can be made more cheaply than Edison’s, and has fully as much life. He is making other lamps of 40 candle and has recently made lamps of 300 ordinary candle power.

Letter from Frank J. Sprague, in \textit{PAPERS OF EDISON VOL. 6}, supra note 85, at 657. See also Swan, supra note 61, at 19 (describing Swan’s response to the growing publicity afforded to the competing systems of electric lighting).

\textsuperscript{87} See, e.g., FRIEDEL & ISRAEL, supra note 59, at 26–29 (describing growing pressure from Edison’s investors at the earliest stages of his research to demonstrate some advantage over existing teams of researchers).

\textsuperscript{88} See generally BAZERMAN, supra note 81, at 247–49 (describing Edison’s practice of resolving conflicting patent claims over the incandescent light by combining forces with the businesses holding competing patents, but noting that this was not successful universally, and the Sawyer Man competition was particularly vexing).

\textsuperscript{89} See id. at 248–57 (detailing the patent interference proceedings and later litigation between the Edison and Sawyer-Man teams).

\textsuperscript{90} See id. at 248 (discussing the beginning of multiple infringement suits involving Edison and his competitors); \textit{infra} Part IV.A (same); see also Israel, \textit{Claim the Earth}, supra note 2, at 20 (“Edison and his company used their strong patent position to enforce licensing agreements with customers of both central and isolated stations and to attack competitors. By 1885, they were involved in nearly one hundred patent office proceedings and lawsuits involving electric lighting.”).

\textsuperscript{91} BRIGHT, supra note 2, at 20.
application. The immense market potential of electric light in the household context was widely realized. For electric engineers of the day, the challenge of “subdividing the electric light”—adapting it for domestic use—was an obvious goal, attracting significant attention and effort even before Edison entered the field. From the perspective of market incentives and of technological groundwork already laid, a commercially practical light bulb may have been a near inevitability.

Edison’s participation, however, probably greatly increased the fervor of this race—at least between his 1878 entry and 1881, when Edison moved from Menlo Park to New York to oversee commercial introduction. Prior to Edison’s entry, limited resources were being invested in pursuit of the incandescent light bulb. The leading American competition was the Sawyer–Man team. William Sawyer was a journalist turned inventor funded primarily by New York lawyer Albon Man, who organized the Electro-Dynamic Light Company for this purpose. The company had allocated $4,000 to Sawyer’s research between March and June 1878. Edison’s talent for raising venture capital would produce a much bigger war chest: nearly $130,000 in research funds were directed to his electric lighting work between 1878 and 1881. Edison’s entry into the field not only attracted significant new investments to his own team, but also motivated increased investments to his competitors. In September 1878, the precise month that Edison threw his hat into the ring, the Sawyer–Man company set up a second workshop.

Indeed, Edison’s entry may have been the defining event that truly kicked off the race for the light bulb as a serious one. When Edison filed his first light bulb patent in September 1878, he gave

---

92. See, e.g., id. at 32 (“Arc lighting was making commercial as well as technical progress in this country [the United States] by the time incandescent lighting appeared on the scene in 1880.”).

93. HUGHES, NETWORKS OF POWER, supra note 60, at 31–32.


95. Id.

96. Id.

97. Id. Adjusted for GDP deflation, Edison’s $130,000 in research expenses from 1878 to 1881 translates to more than $2.5 million in current dollars. Id.

98. Id.
an interview to the New York Sun in which he claimed definitively—if not accurately—to have solved the problem of incandescent light.99 A brief excerpt serves to illustrate the effect:

Mr. Edison says that he has discovered how to make electricity a cheap and profitable substitute for illuminating gas. Many scientific men have worked assiduously in that direction, but with little success. . . . It has been reserved for Mr. Edison to solve the difficult problem desired. This, he says, he has done within a few days. . . . “I have obtained it through an entirely different process from which scientific men have ever sought to secure it. They have all been working in the same groove, and when it is known how I accomplished my object, everybody will wonder why they have never thought of it, it is so simple . . . .”100

A few days later, Edison began financial negotiations for investments.101 A letter privately assured one potential investor that “all serious difficulties have been overcome and there is every reason to believe that [Edison] has discovered the means of giving an electrical light suitable for every use at a vastly reduced cost as compared with the cost of gas.”102

Edison thus claimed to not only be ahead of his competition in the inventive race for incandescent light, but to have already won. Stock prices in gas companies tumbled, indicating that at least some people took these claims seriously.103 More sophisticated investors, however, knew better than to take Edison’s initial claims at face value.104 They suspected that the solution was not yet in hand and even doubted whether Edison

---

100. Id. at 83.
101. BAZERMAN, supra note 81, at 10; FRIEDEL & ISRAEL, supra note 59, at 14.
103. See FRIEDEL & ISRAEL, supra note 59, at 23 (“[I]n early October the news from America caused a panic in gas shares in London, where Edison’s reputation for wizardry was unequaled.”).
104. See id. at 14 (“[T]he story was read by some of the Wall Street moneymen who had already learned to be wary of Edison’s technical genius.”).
would ultimately be the first to achieve it. Indeed, the technique Edison claimed in U.S. Patent No. 214,636, “Improvement in Electric Lights,” was ultimately unsuccessful. The line of research reflected in this first patent filing was discarded entirely and the design never made it into an actual product. Edison would not be able to publicly demonstrate a successful bulb—utilizing a different underlying technology—until more than a year later, at the end of December 1879.

Although Edison’s claims were premature, they could not be ignored. Lemley suggests that patents may increase the incentives in a race as both a carrot and a stick. From the carrot perspective, runners race because they believe that they will finish first and obtain a patent reward. From the stick perspective, runners are motivated to go faster because they fear that their competitors will win, obtaining patent rights that shut them out of a lucrative market. Looking at the race for the light bulb, the stick dynamic seems to be strongly in play. Edison’s 1878 claims represented a threat that he would solve the problem of the light bulb in the near future. That threat seemed uniquely

105. Potential investors in Europe asked to see a working model of the invention, which Edison promised to send through his agent George Gourad. See Letter from Thomas A. Edison to Theodore Puskas (Oct. 5, 1878), in THERESA M. COLLINS & LISA GITELMAN, THOMAS EDISON AND MODERN AMERICA: A BRIEF HISTORY WITH DOCUMENTS, at 85–86 (2002) (hereinafter A BRIEF HISTORY). Two days later, however, a telegram from Gouraud himself to Edison vainly pleaded for a demonstration or some helpful detail of the breakthrough that could be publicly shared to bolster confidence before panicked British investors moved to support a rival system. See Telegram from George Gouraud to Thomas A. Edison (Oct. 7, 1878), in A BRIEF HISTORY, supra note 105, at 86. Edison cabled back advising Gouraud to “let them go ahead,” as he could not yet provide the requested proof. Id.; see also FRIEDEL & ISRAEL, supra note 59, at 14 (describing the hesitancy of investors in financing Edison’s new invention).

106. BAZERMAN, supra note 81, at 186–89. Bazerman deftly provides a clear explanation of the technique claimed by Edison in this first patent, which even a non-engineer can understand, as well as reproducing the patent specification and claims, as issued. Id.

107. See e.g., id, at 183–85; FRIEDEL & ISRAEL, ELECTRIC LIGHT, supra note 59, at 111–14; ISRAEL, A LIFE OF INVENTION, supra note 94, at 187–88.

108. See Lemley, Sole Inventor, supra note 45, at 755 (describing the different motivations for inventors to move quickly in developing their invention).

109. See id. (describing the role a patent race can play as a “stick” by motivating inventors to move quickly to avoid “the risk of losing a race and being excluded from competition in that market”).
credible coming from Edison because of his public reputation as an inventor and because of his obviously passionate commitment to this particular technological challenge.\footnote{See Friedel & Israel, \textit{supra} note 59, at 29 (suggesting that initially, Edison’s chief asset was self-confidence, largely shared by the world at large).}

Edison’s entry into the race on these terms helped to create an air of inevitability surrounding incandescent lighting. The perfection of the technology suddenly seemed close at hand. Investors scrambled to pick sides and stake claims in what now appeared poised to be an important new industry.\footnote{See id. at 14 (examining the various investors who decided to finance Edison’s operation, including individuals associated with the telegraph and gas companies).} Gas lighting companies stood to come out as losers no matter which team won the race. This made it all the more important for them to throw some of their great resources into backing a winning team in order to hedge their existing investments. The race was definitively on and gathering momentum. The key points of Lemley’s theory—the prevalence of simultaneous invention, the role of patent protection in raising the stakes, and the power of competition to motivate increased investments in research and development—all find support in the race to perfect the light bulb.

\textit{C. Problems for Patent Racing Theory}

Despite all this support in the historical record, Lemley’s theory of patent racing has been vehemently disputed by business professor John Howells and his co-author, Ron D. Katznelson.\footnote{See John Howells & Ron D. Katznelson, A Critique of Mark Lemley’s “A Myth of the Sole Inventor” 4–5 (Sept. 15, 2011) (unpublished manuscript), \textit{available at} http://bitly.com/Lemley-Critique (on file with author).} These critics specifically take issue with Lemley’s use of the light bulb as an example of incremental invention.\footnote{See id. (disputing Lemley’s claim that Edison was not a sole inventor by highlighting the distinct strengths of Edison’s invention in contrast with those of his competition).} They argue that Edison’s contribution in fact “unlocked the field,” that it cannot be placed on par with contributions by any other inventor, and...
that Edison was deserving of a “pioneer” patent. My own examination of the light bulb case puts me on the side of Lemley rather than his critics. The central mistake made by Howells and Katznelson is to treat judicial opinions and patent claims as reliable evidence of who really did what. These legal documents emphatically assert that Edison’s inventive contributions were without peer. Such assertions, however, should be treated with great skepticism. The historical claims about Edison’s contributions, which Howells and Katznelson present as fact, are better understood as a deeply mythologized “narrative of invention” that Edison himself forged in preparation for litigation. The model of incremental innovation more accurately fits the true history of the light bulb’s development.

Contrary to Howells and Katznelson’s position, my concern with Lemley’s patent racing theory is that the model may not go far enough to reflect the reality of incremental innovation. Lemley describes simultaneous invention as existing in two models. The first model is when two inventors are working on the same technology, and the contest is to see who will produce it first. Lemley offers the telephone as the archetype of a patent race: two inventors almost literally racing to the patent office door. But the second model is probably more common. This is when the process of invention is characterized by roughly

114. Id. at 6. The concept of a “pioneer patent” was popular in American patent doctrine at the late nineteenth and early twentieth centuries. Proponents argued that when an inventor had made a “pioneering” contribution to a new field of technology, he should be entitled to broad claims covering not only the specific embodiment they had produced but also later generations of the technology that others developed. For a critique of this doctrine—and argument as to its enduring relevance—see Brian J. Love, Interring the Pioneer Invention Doctrine, 90 N. CAROLINA L. REV. 379, 389 (2012).

115. See Howells & Katznelson, supra note 112, at 4–7 (describing the reasons that Edison’s patent for the incandescent light was superior to his competitors).


117. See Lemley, Sole Inventor, supra note 45, at 731 (“[I]nventions . . . are in fact generally the products either of simultaneous independent invention or of incremental development from multiple sources, or both.”).

118. Id. at 712–15.

119. Id. at 720–22.
contemporaneous contributions by so many different parties that it becomes impossible to declare any one true “father” of the technology. The race for the light bulb, at least, best fits this pattern of incremental innovation. Even in the case of the telephone, the narrative of the race to the patent office may be more a product of litigation stance than of the underlying reality of technological development.

If the pattern of incremental innovation is indeed the prevailing one, as I suspect, we need a reconceptualization of racing theory that deemphasizes the patent office door as the finish line. Doing so, however, destabilizes the metaphor. Central to the concept of a race is the existence of a clearly defined finish line, at which point the winners and losers are declared based upon a simple and objective measure of performance. Indeed, the very purpose of the race as an athletic convention is to enable this objectivity by clearly defining the terms of the contest in advance. A political campaign also meets these two criteria of a traditional race: a clearly defined endpoint (election day) and an objective methodology for determining the winner (tabulating votes). Inventive races, however, may not conform to this model. The race for the light bulb, at least, lacked these two key characteristics.

120. Id. at 715–16.
121. For a comprehensive account of innovations in electrical lighting between 1880 and 1890, well situated in their economic and political context, see generally HUGHES, NETWORKS OF POWER, supra note 60.
122. See Beauchamp, Who Invented the Telephone, supra note 39, at 877 (highlighting the difficulty of “the question of where, on the continuum of conceptualization, experimentation, ‘reduction to practice,’ and commercialization, the invention of the telephone should be rightly marked”).
123. The inventive race for the light bulb has one other similarity with political races, however: the dynamics of self-fulfilling prophecies. In a political campaign, contenders seek to create the perception that they are in the lead, in order to win over those potential supporters who simply want to throw in their lot with the likely winner. Edison thus cultivated the media and exaggerated his own progress to discourage investors from backing his rivals. See FRIEDEL & ISRAEL, supra note 59, at 13. Similar to a political campaign, impressions about which inventive team is ahead can translate into additional resources helpful for getting ahead, making such prophecies self-fulfilling. Mercifully, even the longest political campaign finds its end within eighteen months, and a victor is clearly anointed. Not so with an inventive race, in which it may take more than a decade for courts to declare the winner.
Indeed, this particular race was so closely and fiercely contested that for many years—even years after light bulbs were widely being manufactured in the United States—there was no clear winner. Throughout the 1880s, multiple companies, each holding patent portfolios from different inventive teams, claimed strenuously that their champions had prevailed. 

Ultimately, these conflicting claims had to be settled by the law. In doing so, the patent office and the courts had to apply legal criteria that are infinitely more complex and subjective than those commonly used to judge athletic and political races. Courts also had to make this determination, not immediately after the relevant finish line, but many years later, upon an imperfect evidentiary record developed by deeply interested parties. In one crucial trial, for instance, Edison’s attorneys persuaded the judge that William E. Sawyer had in fact never made the lamp claimed in the disputed patent. Because ten years had passed between the claimed experiments and the Incandescent Lamp Patent trial, Mr. Sawyer was no longer alive to testify personally as to the events.

Beyond mere problems of evidence, Robin Feldman argues persuasively that subjectivity and indeterminacy are inherent features of patent interpretation. In Feldman’s words, “[p]atents cannot possibly delineate the boundary of an inventor’s rights because those rights will be established in relation to products that have yet to be created at the time of the patent grant.” Patent lawyers face the daunting task of choosing words to capture the essence of a technology that by definition

124. See, e.g., Bazerman, supra note 81, at 247–48 (describing the uncertain and overlapping claims of ownership over the light bulb technology and the increasingly competitive production of light bulbs based on these various patents).

125. See id. at 248–50 (describing the interference battle between Edison and the Sawyer-Man team at the patent office in 1881–1883 and later litigation); Passer, The Electrical Manufacturers, supra note 42, at 151–59 (describing the beginning of patent litigation over the incandescent lamp in 1885 and its continuation for many years); Consol. Elec. Light Co. v. McKeesport Light Co., 159 U.S. 465 (1895) (invalidating a Sawyer-Man patent under which any light bulb with a filament of carbonized vegetable matter would have been deemed infringing).


127. See generally Feldman, Rethinking Patent Law, supra note 68.

128. Id. at 3.
has never before been described, in order to compare it in the future to other technologies that do not yet exist. 129 “Taken together, this set of circumstances ensures that a patent could never grant a definitive and clearly bounded set of rights.” 130 Feldman concludes that a patent must be understood not as a clear marking of the “metes and bounds” of one party’s intellectual property, but rather as an invitation to bargain with others over the scope and value of the patent rights. 131 In the end, “the extent of the rights cannot truly be known until the day that the patent expires.” 132 Keeping Edison’s experience in mind, we might also characterize this inherent uncertainty of patent claims as not just “an invitation to bargain,” but also as an invitation to litigate.

Because of problems of evidence, indeterminate claim language, and shifting standards of legal interpretation, the nature of the competition in a patent race may be much less objective and meritocratic than the metaphor initially suggests. Are inventive races meaningfully similar to the 100-meter dash or a presidential primary? Or are they more like an Olympic gymnastics competition, in which the American and Soviet judges may arrive at wildly different conclusions? 133 Whether we view the patent race as a fair and objective contest or not has great implications for the degree of confidence we place in the patent system to pick winners and losers, and therefore, to structure the proper incentives for technological innovation. Lemley’s patent racing model lessens the power of the myth that inventions appear out of thin air, and would not have existed but for the contributions of the credited inventor, a misperception that too easily favors very strong patent rights. Without great caution, however, the patent racing model might replace an old myth with a new one: that of the patent race as a fair and clean contest. Patent law on the ground is not so neat and tidy.

129. Id.
130. Id.
131. Id.
132. Id. at 5.
133. After all, American courts picked Thomas Edison, and British courts picked Joseph Swan.
The patent racing model also carries a second risk: it tells only part of the story and leaves out what might be the most important part, from the perspective of innovation. Patent racing theory best describes the early stages of development of a new technology. In the latter stages of commercialization, improvement, and diffusion of that technology, the nature of the competition shifts very distinctively.\textsuperscript{134} This latter dynamic, which I term “patent warfare,” can also be understood as a contest or game. But its nature is quite different from the patent race described by Lemley, and its implications for innovation are not nearly so positive. A look at what happens next in the story of the light bulb—after the technology was perfected enough for Edison and others to begin to commercialize it—will make this point crystal clear.

IV. Learning from the Light Bulb

A. From Patent Racing to Patent War

In Part III, the early contest to develop and bring to market an incandescent light bulb was used to test the theory of patent racing. The history largely supports the theory of patent racing, albeit with skepticism as to any implication of a clear finish line. After the light bulb came to market, however, the dynamics of the competition among inventive teams shifted dramatically. Here, the more accurate metaphor is neither a sprint nor a marathon, but the strategic board game \textit{Risk}.\textsuperscript{135} In this game, the board resembles a map of the world, with varying territories. In the early stages of the game, players amass tokens representing armies and position them strategically on the board to fortify their positions. Eventually, the game shifts into attack mode, as

\textsuperscript{134} I am indebted to Gaia Bernstein for highlighting an important distinction between innovation and diffusion of new technologies. \textit{See}, \textit{e.g.}, Bernstein, \textit{supra} note 16, at 2290–91 (distinguishing the innovation and diffusion stages of a technology’s introduction and arguing that intellectual property law neglects the latter).

players vie to conquer each other’s territories. Ultimately, one party pushes out all the competitors and achieves the goal of the game: world domination.

Similarly, industrial contenders may spend years amassing a patent portfolio and income streams that will eventually fund their litigation war chest. At a certain point, when one party thinks it is the right amount ahead, it begins to close in on its opponents. Both in *Risk* and in patent warfare, the strategic campaigner should begin by confronting its weakest competitors because those victories will be easiest. The victor may also amass strength by absorbing the assets—patent portfolio and market share—of the vanquished foe. Finally, the two strongest competitors confront each other, as the Edison and Westinghouse companies faced off in the litigation that reached the Supreme Court in the *Incandescent Lamp Patent* case.

To paraphrase the classical military theorist Clausewitz: Patent litigation is the continuation of business strategy by other means. Patent racing is helpful for understanding how patents impact innovation leading up to the patent filing. But the long-term impact of patent law on industrial competition requires a very different metaphor. Patents acquired in the racing stage become pawns in a battle to achieve market dominance and secure competitive position. Contrary to the metaphor of the race, it is not who is first to the patent office that wins, but who is last left standing at the end of litigation.

In the early days of the light bulb’s commercial use, patents proliferated but did not pose a barrier to competition. Economic historian Harold C. Passer points out that Edison had a significant share of the market early on, but not anything like exclusive control. In Passer’s words: “In the beginning, very little attention was paid to patents, and manufacturers could make the entire incandescent system. The dynamo could not be


137. *See Passer, The Electrical Manufacturers,* *supra* note 42, at 206 (“Of the 250,000 [incandescent lights] in use in 1885, over 200,000 were Edison lights on Edison equipment.”).
patented, and lamp patents were not respected.”\textsuperscript{138} By 1890, however, Edison’s market share had fallen to less than half.\textsuperscript{139} Patent warfare was crucial to reversing that trend and securing Edison’s control of the light-bulb market. In the first year of litigation, Edison’s company was involved in nearly one hundred patent disputes, including administrative proceedings.\textsuperscript{140} Between 1885 and 1895, more than twenty major opinions were issued in U.S. cases disputing Edison’s claims to control of the incandescent light bulb.\textsuperscript{141} By 1901, Edison’s side had spent about $2,000,000 on more than two-hundred lawsuits related to its lighting patents.\textsuperscript{142} The Sawyer–Man infringement suit against Edison alone generated more than 5,000 pages of documents, not including the patent application and interference stages.\textsuperscript{143} Nor was the litigation confined to the American shores. Foreign patents and litigation, especially in Great Britain and in

\textsuperscript{138} Id.
\textsuperscript{139} Id.
\textsuperscript{140} Israel, Claim the Earth, supra note 2, at 20.
\textsuperscript{142} \textsc{Bright}, supra note 2, at 136–37.
\textsuperscript{143} \textsc{Bazerman}, supra note 81, at 239–40 (“The conflict . . . moved to the courts, where it produced eight volumes of documents totaling more than 5000 pages.”).
Germany, formed an important part of Edison's transnational litigation strategy.

As companies in the field of incandescent light technology negotiated cross-licensing deals and takeovers, patents on many aspects of the electrical lighting system played important roles. But in U.S. litigation one of Edison's many patents on the light bulb would ultimately play the most important role. Issued on January 27, 1880, U.S. Patent No. 223,898 ('898 Patent) recognized Thomas Alva Edison’s “improvement in electric lamps,” including a claim on “[a]n electric lamp for giving light by incandescence, consisting of a filament of carbon of high resistance, made as described, and secured to metallic wires, as set forth.” The '898 Patent was just one among dozens of patents that Edison sought on the light bulb technology, and it was not initially clear that it would prove so central. Competitors in the industry strongly doubted the validity of its claims. That changed crucially when a Pennsylvania court, in the course of invalidating a rival patent held by Sawyer and Man, and in dicta, gave Edison's '898 Patent a new interpretation so broad as to render virtually every incandescent light bulb infringing. Subsequent litigation produced a stunning consolidation of the previously very competitive American electric industry.

---

144. See BRIGHT, supra note 2, at 86. See generally Israel, Claim the Earth, supra note 2.

145. Thomas Alva Edison, Elec. Lamp, U.S. Patent No. 223,898 (filed Nov. 11, 1879) (issued Jan. 27, 1880). The patent contains three additional independent claims, which are omitted here for the sake of brevity. Id.

146. See BAZERMAN, supra note 81, at 86–89 (explaining that, despite the large number of Edison patents, only the '898 Patent was deemed "crucial").


148. See Consol. Elec. Light Co. v. McKeepart Light Co., 40 F. 21 (C.C.W.D. Pa. 1889), aff’d, 159 U.S. 465 (1895) (invalidating the Sawyer–Man Patent); PASSER, THE ELECTRICAL MANUFACTURERS, supra note 42, at 155 (“The meaning of the decision seemed to be that the Edison patent could be the basis for an absolute monopoly of incandescent-lamp manufacture in the United States.”); see also BAZERMAN, supra note 81, at 250–57 (offering an insightful account of the McKeepart trial).

149. See BRIGHT, supra note 2, at 84–93; see also BAZERMAN, supra note 81, at 239–40 (explaining that, after the litigation with its industry competition, the Edison companies and their successor, General Electric, emerged as “the primary owner of light technology as the industry expanded”).
Court’s 1895 decision, all U.S. producers either purchased licenses from General Electric or merged with it.150

Passer’s economic study documents the transformation of the light bulb market before and after the patent litigation. In the early years of the electrical industry, he writes:

Patents were not significant in either the product or the manufacturing process. In these conditions—a standardized product sold mainly to business firms, relatively free entry, many buyers, and more than fifteen sellers—the market closely approached the economist’s concept of pure competition. The competition was almost entirely in prices, and these were driven down to cost by the additional supply from new firms. But these freely competitive conditions did not prevail for long. The year 1888 marked the end of pure competition in arc carbons, and 1896 [the year after The Incandescent Lamp Patent was decided] saw the termination of pure competition in incandescent lamps.151

By 1897, General Electric was the undisputed leader, and Westinghouse was its only serious rival.152 The two companies agreed to a truce in the form of patent cross-licensing and product pricing agreements.153 Smaller competitors were invited to join

150. See BRIGHT, supra note 2, at 136–37 (explaining that the combination of “patent victory, consolidations, and patent and marketing agreements” made Edison the indisputable leader of the domestic incandescent-lamp industry).
151. PASSER, THE ELECTRICAL MANUFACTURERS, supra note 42, at 350. Passer details forces contributing to the demise of competition including both non-patent and patent factors. Of the latter, he writes:

Patents also influenced competitive conditions. They were a prime cause of the numerous consolidations and mergers which finally resulted in only two full-line producers. Furthermore, patents permitted competition to take place on a system basis instead of with reference to single items of equipment. Manufacturers could refuse to allow the use of particular patented apparatus except as a part of a complete lighting or power system.

Id. at 352.
152. See BRIGHT, supra note 2, at 144 (“In 1887 the General Electric Company was unquestionably the leader of the American electric-lamp industry. It had recently entered into a general cross-licensing arrangement with the Westinghouse Electric & Manufacturing Company, its largest competitor . . . .”).
153. See id. at 12 (“A general patent-licensing arrangement between General Electric and Westinghouse in 1896 and subsequent specific licenses established a pattern in the lamp industry which was substantially maintained until 1945.”); see also W. BERNARD CARLSON, INNOVATION AS A SOCIAL PROCESS: ELIHU THOMSON AND THE RISE OF GENERAL ELECTRIC 271–301 (1991).
the Incandescent Lamp Manufacturers Association, organized by General Electric. These members agreed to divide up the markets and avoid price competition. The result was that G.E. sold half of the bulbs in the U.S. market, and the smaller companies divided the other half. These smaller companies, dependent on G.E. patent licenses for their existence, put no downward pressure on prices and were discouraged from contributing their full potential to next-generation innovation. In Bright’s words: “They gave only the appearance of competition.” Through incorporation, patent litigation, licensing deals, and eventually purchases of stock, by 1910, G.E. controlled 97% of the market. Only then could Edison finally rest, his empire secure.

History remembers Edison as a workshop inventor, a tinkerer—an image established, in no small part, by Edison’s own efforts. A more accurate historical portrait begins with recognizing Edison as an entrepreneur and empire builder.

154. See Bright, supra note 2, at 144 (explaining that General Electric organized the Incandescent Lamp Manufacturers Association, obtaining the cooperation of a large proportion of the industry, including six “formerly bitter rivals,” who joined to avoid intense price competition).
155. Id. (“The members of the association agreed among themselves to fix lamp prices, both wholesale and retail, to divide business and customers, and to set terms of sale.”).
156. Id. at 457. (“The role of the licensees was largely passive . . . add[ing] little to the vitality of the lamp industry; and they aided General Electric in retaining its hold over the industry. They did not provide any real downward pressure on prices. They gave only the appearance of competition.”).
157. Id.
158. Id. at 147–48. An antitrust investigation subsequently found misconduct, but the remedies did not require the company to seriously change its business model. General Electric would continue to dominate the light bulb market for decades longer. See id. at 253–56.
159. In an unguarded moment, Edison's voice was captured on his own phonograph in an exchange suggesting Edison saw himself as a “hustler.” Stross tells the story best:

One occasion when Edison cast off the expectations of others in his middle age was when he met Henry Stanley, of “Dr. Livingston, I presume” fame, and Stanley's wife, who had come to visit him at his laboratory in West Orange, N.J. Edison provided a demonstration of the phonograph, which Stanley had never heard before. Stanley asked, in a low voice and slow cadence, “Mr. Edison, if it were possible for you to hear the voice of any man whose name is known in the history of the world, whose voice would you prefer to hear?” “Napoleon’s,” replied Edison without hesitation.
Among his contemporaries working on the light bulb, Edison had a unique focus on commercialization of the emerging technology, rather than merely securing patents. Edison was no mere inventor, but a different breed altogether: the inventor-entrepreneur. From the beginning, Edison intended not merely to license his light-bulb patents, but to commercialize the technology himself; indeed, much of his energies would be consumed in supervising the rollout of his products in the marketplace. In this way, he had much more in common with Steve Jobs or Bill Gates than with the average patent seeker.

As an entrepreneur, Edison understood that he would have to confront his competition in the marketplace sooner or later, no matter how advantageous his patent portfolio. He likely recognized the business advantages that come with being the first

“No, no,” Stanley said piously, “I should like to hear the voice of our Savior.”

“Well,” explained Edison, “You know, I like a hustler.”

Randall Stross, Edison the Inventor: Edison the Showman, N.Y. TIMES, Mar. 11, 2007, at B1. Strictly speaking, Edison’s use of the word “hustler” describes Napoleon, not himself. But Edison’s selection of Napoleon as the historical voice he would most like to hear strongly suggests that the inventor considered the empire-maker a personal role model. In describing Napoleon as a hustler, I believe Edison was also describing his vision of himself. It was a self-image successfully conveyed to others. Edison collaborator Francis Jehl would later celebrate Edison as the Napoleon of the team’s efforts.

160. See Passer, supra note 42, at 83 (“From the economist’s viewpoint, the most significant aspect of Edison’s activities in electric lighting was his concern at every step with economic factors.”); see also Hughes, Networks of Power, supra note 60, at 29 (“From the start, he clearly realized . . . that his system would have to be economically competitive, and thus he conceived of the problem to be solved by invention as inseparably technical and economic.”).

161. See Hughes, Networks of Power, supra note 60, at 18–22 (describing Edison’s work as an inventor-entrepreneur). Harold C. Passer had earlier written of the engineer-entrepreneur, of which Edison was a leading example. See Passer, supra note 42, at 356–60. “Edison was an engineer-entrepreneur on a full-time basis when he turned to the electric light. He made the move consciously and chose electric lighting as the best of a number of alternatives.” Id. at 357.

162. See Hughes, Networks of Power, supra note 60, at 39–42 (explaining Edison’s strategic design and location of the Pearl Street Station of the Edison Electric Illuminating Company in the financial district of New York City).

163. See Passer, supra note 42, at 83 (“Nearly all of [Edison’s] inventive activities were directed toward making his system as efficient as possible. If his system were to be successful . . . it would have to survive the market test that every product in a free economic system has to meet.”).
to bring a product to market. Moreover, Edison was not just an experienced inventor and entrepreneur, he was also an experienced patent litigator, having participated in patent disputes related to many of his earlier inventions. Because of this past experience, Edison likely appreciated how difficult a task the courts would ultimately face in sorting out the competing patent claims surrounding incandescent light and realized that the party with greater market share and greater income would hold an important strategic advantage in the very expensive process of litigation.

Perhaps Edison ultimately won the patent war over electric light because, uniquely among the parties, he understood the true nature of the contest. Throughout the inventive race to the marketplace, each patent filing was merely a mile marker in an ultra-marathon of uncertain length. No one knew at the time of filing which patent claims would turn out to be key, once the market and the litigation began to take shape. 164 This is not to say that patents were unimportant. They were stockpiled for the future, when the competition ultimately shifted from an inventive race to patent war. This battlefield was the true finish line, and the competition here took place as much on the dimensions of litigation and business strategy as on the technical merits. In retrospect, the initial patent race was not an athletic competition with a clearly defined finish line, allowing judges to objectively declare the winner. It was an arms race.165

B. The Current Smartphone Patent War

This Article’s model of patent litigation, as resembling a game of Risk, was inspired by historical evidence. But the

164. See Bright, supra note 2, at 86 (detailing the situation before the patent war began, stating that by 1887 the “validity of the patents granted in 1880 had not yet been tested”). See generally Feldman, Rethinking Patent Law, supra note 68, at 40–74 (suggesting that this tremendous uncertainty as to future meaning and value at the time a patent is filed and granted is inherent to the nature of patent law).

metaphor of patent warfare has also recently been used to describe emerging litigation in the IT industry. According to competition and antitrust law scholar Michael Carrier, “The smartphone industry today is characterized by a thicket of patents and wars based on those patents. Every day brings a new lawsuit or development between Apple, HTC, Microsoft, Motorola Mobility (MMI), Nokia, and Samsung.”

Carrier is neither the first nor the last to refer to the developing pattern of smartphone litigation as patent “warfare.” Journalists appear to have taken the cue for this military metaphor from the litigants themselves. As the metaphor suggests, the goal of this patent litigation is to crush the opponent and force surrender.

Ironically, this new patent war is breaking out in the very sector of the economy where patents may have played the smallest role in incentivizing innovation. When the computer industry first emerged in the 1960s, the United States Patent and Trademark Office (USPTO) resisted software patent applications as a matter of policy, and computer companies themselves

---

166. Carrier, supra note 4, at 1.
167. See, e.g., McMillan, supra note 4 (“But in practice, patents are weapons. Technology companies load up on patents like Cold War nations stockpiling nuclear bombs, hoarding them for use when an important market is at stake.”); Oberlander, Stabe & Bernard, supra note 4 (“Patent wars are raging in the smartphone industry. What began as Apple v. Google Android conflict has turned into a vast legal quagmire involving everyone from Amazon to ZTE.”); Vascellaro & Ramstad, supra note 4 (“Chief executives . . . will meet in San Francisco Monday in a court-directed session aimed at settling their smartphone patent war. But a deal seems unlikely, people familiar with the matter and others tracking the battle say.”).
169. See Pamela Samuelson, Benson Revisited: The Case Against Patent Protection for Algorithms and Other Computer Program-Related Inventions, 39 EMORY L.J. 1025, 1038–39 (1990) (“It was the [Patent Office] Commission’s judgment that patent protection for computer program innovations was not desirable. The Commission’s 1966 report . . . recommended legislation that expressly excluded computer programs from patent protection.”).
supported this approach. Even in the 1990s, some leading software firms opposed the granting of patents in their field. Software engineers also overwhelmingly expressed the view that patents were unnecessary and potentially harmful in their field. Nevertheless, the patentability of software has now been firmly established in the United States, and the number of software patent applications filed and granted has dramatically increased. Europe has been even slower than the U.S. to grant patent protection to software but is moving in the same direction. It is unclear whether the modern availability of

170. See id. at 1028 n.3 (describing industry input to the presidential commission in 1966); see also id. at 1143 (providing an overview of the early legal changes regarding the patentability of software).
174. The European Patent Convention states that methods of “doing business[] and programs for computers” are not eligible for patent protection. European Patent Office [EPO], European Patent Convention art. 52(3), Oct. 5, 1973, 13 I.L.M. 268. That prohibition has been significantly modified, however, by more recent legal sources. The European Patent Office routinely grants software patents. See EPO, Guidelines for Examination in the European Patent Office, Part C, Chapter IV, §§ 2.1, 2.2 (June 2005) (indicating that the prohibition on the patentability of items under Art. 52(2) should be narrowly construed); see also Computer Program Product/IBM, T 1173/97-3.5.1 (EPO Bd. of App. July 1, 1998) (noting that “a computer program product is
Even under the newer pro-patenting rules, the prevailing dynamic of patenting in the software industry was “defensive”—with companies seeking patent portfolios as leverage to protect themselves against suits by others.176 Strong patterns of cross-licensing and a “gentleman’s agreement” not to sue kept the major firms from litigating their patents.177 In this context, it seemed that the major corporations had relegated patent law to an incidental role in the industry’s business models. Mobile telephony was built upon collaborative industry standards relying on voluntary patent pooling. The Internet was built upon nonproprietary protocols. These were understood as deliberate choices made by industry consensus to limit the power of patent law and thereby avoid its negative consequences.178

Industry insiders and patent scholars alike assumed that the peace was secured by the logic of mutually assured destruction. Alex Blumberg and Laura Sydell summed up the conventional

not . . . excluded under all circumstances” under Article 52(2) and (3)). This leaves it up to each nation’s courts to judge the validity of such patents in subsequent litigation. For a current overview of U.K. and German case law on the validity of software patents, see Susan J. Marsnick & Robert E. Thomas, Drawing a Line in the Patent Subject-Matter Sands: Does Europe Provide a Solution to the Software and Business Method Patent Problem?, 34 B.C. INT’L & COMP. L. REV. 227 (2011).

175. See generally Bessen, Software Patents, supra note 171 (drawing on empirical research to conclude that the spread of software patents has not produced greater innovation, only greater litigation). It should be noted, however, that software has long enjoyed copyright protection. The important difference is that copyright protection does not prevent competitors from reverse-engineering the same function, so long as they write original code.

176. See, e.g., Chien, supra note 165, at 303–10 (discussing the tendency of companies in the current software industry to pursue defensive patenting strategies); see also Baio, supra note 4 (recounting his participation in a software patenting effort he believed at the time was defensive, only to later find his firm at the center of offensive use of such patents).


178. See id. (discussing the patent-licensing arrangements in the technology industry that sought to avoid patent law’s negative impact).
wisdom of industry insiders and followers when they reported in 2011:

All the big tech companies have started amassing troves of software patents—not to build anything, but to defend themselves. If a company’s patent horde is big enough, it can essentially say to the world “If you try to sue me with your patents, I’ll sue you with mine.” It’s mutually assured destruction. But instead of arsenals of nuclear weapons, it’s arsenals of patents.\(^{179}\)

By 2012, however, it became clear that the truce had broken down. What is less clear as this Article goes to press is what the consequences of this patent war will turn out to be. Will it indeed be mutual destruction? Or will one party emerge victorious, and with what damage to each side? And what will be the collateral damage to the public interest?

The smartphone patent war might be viewed as simply one more episode in the inevitable dynamics of economic uncertainty in a competitive marketplace. We might choose to remain confident in the legal system to sort things out properly, and in the markets to rebalance themselves. The lesson of the light-bulb war, however, cautions against such confidence. If history repeats itself in the smartphone patent war, we could once again see the consolidation of an important high-technology industry, with the attendant decline of competition.\(^{180}\) If one company establishes dominance in the smartphone sector, it will not only control the device itself but also have a market-dominant position from which to influence the markets for related services: mobile telephony, software, social networking, online advertising, mobile banking, and online retail. Smartphone dominance could be used to extract rents from and consolidate those related industries, subject only to the limits imposed by effective enforcement of antitrust law.\(^{181}\)


\(^{180}\) See supra Part IV.A (noting that the electric industry, following the patent warfare involving the light bulb, consolidated into General Electric, thus effectively preventing competition).

\(^{181}\) See Carrier, supra note 4 (discussing the role that antitrust law has begun to assume in limiting market dominance in the high-technology industry,
To be sure, the fact that patent warfare led to effective monopoly in the early electrical industry is no guarantee that the same result will occur in the smartphone sector. The light bulb and the smartphone are different technologies, and these differences may produce different outcomes in the two industries. It is also possible that reforms in patent and antitrust law since the turn of the century have lowered the risks for this kind of patent-precipitated consolidation of a market. Nevertheless, the very fact that a similar pattern of patent warfare is exhibited in two very different industries, at two very different points in time, suggests significant commonalities. Rather than adopt a wait-and-see approach when so much may be on the line, what can be done proactively to reduce the risk of consolidation and loss of competition in the smartphone industry?

At least one federal judge has already flagged the risk that patent litigation poses to industry competition and consumer welfare in the smartphone sector. In June 2012, Judge Richard A. Posner issued a remarkable ruling in a particularly high-profile smartphone patent suit between Apple and Motorola (which was acquired by Google in August 2011). Judge Posner indicated his intention to dismiss the suit with prejudice without granting a jury trial, on the grounds that, even if the disputed patents were valid and infringed, there was insufficient evidence to support damages, and injunctive relief would be counter to the public interest. This holding may be interpreted as clever procedural handling by Judge Posner to squash a case in which one or both parties were attempting to misuse patent law to damage its competitor company out of proportion to any real harm done. On the heels of the decision, Judge Posner—acting in his capacity as a law professor—also initiated a media effort to

---

182. See Posner, Too Many Patents, supra note 4 (noting the potential risks of certain patenting tactics).

183. Judge Posner is one of a number of Article III judges specially designated as volunteers to hear patent disputes at the trial court level, having received special training in the handling of these complex cases.

highlight the potential of patent law to damage competition and innovation, calling for dramatic patent reform.\textsuperscript{185} The suit before Judge Posner, however, was just one among many. In August 2012, Judge Lucy Koh presided over a highly publicized trial brought by Apple in the Northern District of California.\textsuperscript{186} The lawsuit targeted Samsung, the leading manufacturer of smartphones running Android software, which were increasingly overtaking the Apple iPhone in popularity.\textsuperscript{187} After just three days of deliberation, the jury found Samsung to have infringed Apple’s utility patents, design patents, and trade dress.\textsuperscript{188} The jury awarded over $1

\textsuperscript{185} See, e.g., Posner, \textit{Too Many Patents}, supra note 4 (discussing the negative consequences of patent law).


\textsuperscript{187} Apple had an early advantage in the smartphone market because of lead-time and network effects. Apple encouraged independent software developers to produce “apps” that could run on Apple’s iOS system. The availability of these apps became a major feature driving consumers to purchase the iPhone. None of the competing smartphone manufacturers commanded sufficient market penetration to attract similar app development, until Google introduced the Android operating system. Google widely licensed the Android software to a number of smartphone manufacturers, including Samsung. The emergence of the Android platform enabled other manufacturers to effectively compete with the iPhone. By mid-2011, Android phones were outselling iPhones more than two-to-one. See IDC, \textit{Android and iOS Surge to New Smartphone OS Record in Second Quarter} (Aug. 8, 2012), http://www.idc.com/getdoc.jsp?containerId=prUS23638712 (last visted Oct. 21, 2012) (reporting Q2 2011 shipments of 50.8 million Android phones and 20.4 million iOS phones) (on file with the Washington and Lee Law Review). By the middle of 2012, however, the gap had widened; four Android phones were being sold for each iPhone. See id. (reporting Q2 2012 shipments of 104.8 million Android phones and 26.0 million iOS phones.) At the time of the trial, Samsung was the leading producer of Android phones. See id. (“Android’s success in the market can be traced directly to Samsung, which accounted for 44.0% of all Android phones shipped [worldwide] in [the second quarter of 2012] and totaled more than the next seven Android vendors’ volumes combined.”).

billion in damages.\textsuperscript{189} Having tested the appeal of their claims before one jury, Apple has probably gained a stronger hand to negotiate concessions from other smartphone manufacturers running the Android system. This case was unique, however, in that the accused products not only incorporated software features described by Apple’s utility patents but were also very similar in external appearance. This latter fact led to infringement claims based on Apple’s design patents and on a theory of trade dress dilution.\textsuperscript{190} The jury was not asked to distinguish the damages it attributed to each theory of infringement.\textsuperscript{191} A different result might thus be reached in another case in which the products were not similar in external appearance.

The Federal Circuit has also shown an inclination to lower the stakes of the smartphone patent war by making it increasingly difficult to block a competitor’s phone on the basis of a minor software patent. For many years, courts routinely awarded not only monetary damages but also an injunction against future sales of the infringing product, whenever the patent holder prevailed on the merits.\textsuperscript{192} This created a problem of “patent holdup,” whereby patents on minor features could be used to block the sale of an entire product.\textsuperscript{193} The Supreme

\begin{footnotesize}
\footnotesize
\begin{enumerate}
\item[189.] See Amended Jury Verdict Form, supra note 188, at 15 (stating the damages awarded to Apple); Levine, \textit{Apple Trial} (awarding a jury verdict of $1.05 billion dollars, which could be tripled because the jury found the company acted willfully).
\item[190.] The San Jose jury concluded that the iPhone trade dress (registered and unregistered) was famous and was willfully diluted by Samsung’s products. See Amended Jury Verdict Form, supra note 188, at 10–12, 14.
\item[191.] See Amended Jury Verdict Form (asking jurors to answer “yes” or “no” for whether each accused product violated each patent and the trade dress, and also “[w]hat is the total dollar amount that Apple is entitled to receive from Samsung on the claims on which you have ruled in favor of Apple?”).
\item[192.] See, e.g., MercExchange, L.L.C. v. eBay Inc, 401 F.3d 1323, 1339 (Fed. Cir. 2005) (granting an injunction and restating the “general rule that courts will issue permanent injunctions against patent infringement absent exceptional circumstances”), rev’d, 547 U.S. 388 (2006).
\end{enumerate}
\end{footnotesize}
Court’s 2006 decision in *Ebay v. MercExchange*^{194} questioned the routine grant of injunctive relief in patent cases and called for a more demanding inquiry into the equities of each case.^{195} In the wake of this case, lower courts have been more inclined to deny injunctive relief, even where a patent holder seems likely to prevail on the merits.

In the Apple–Samsung case described above, Judge Koh refused to enjoin the sale of four Samsung products in December 2011, despite holding that Apple was likely to prevail in proving patent infringement.^{196} Koh reasoned that Apple had failed to prove a sufficient “nexus” between the patent Samsung infringed and the alleged harm of lost iPhone sales.^{197} On appeal, the

big product from the market because of a small patent harms consumers, and blocking a large number of lawful components and features from the market along with the infringing one distorts competition.

In another common situation, a patent holder sues a large number of companies, threatening to hobble an entire industry. An injunction would exclude many participants from the marketplace and dramatically reduce competition.

In both these scenarios, the harm to consumers and competition from an exclusion order is greater than the contribution made by the individual infringing component.

---


195. See *id.* at 393–94 (holding that injunctive relief was inappropriate based on a case-specific analysis of four equitable factors); see also *Winter v. Natural Res. Def. Council, Inc.*, 555 U.S. 7, 20 (2008) (applying the *MercExchange* standard to deny a preliminary injunction in a case not related to patent law).


197. See *id.* at *20–21, *39–40. Judge Koh explained her decision with reference to the “leveraging” reasoning offered in *Ebay*:

Even though Apple has shown Samsung’s products likely infringe Apple’s valid ‘381 patent, and that Apple and Samsung compete in the same market for new smartphone customers and for tablet computer consumers, Apple has offered no evidence that Samsung’s infringement of the ’381 patent is likely to cause irreparable harm. . . . Indeed, Apple has failed to establish a relationship between any alleged loss of market share, customers, or goodwill, and the infringement of the ’381 patent.
Federal Circuit endorsed Koh’s articulation and application of a nexus requirement. Following this guidance, Judge Koh again conducted a nexus analysis in June 2012, in the second case brought by Apple against Samsung in her court. This time, Koh found the nexus requirement to be satisfied and granted an injunction barring sales of Samsung’s Galaxy Nexus smartphone. Judge Koh held that Apple had met its burden of proof with respect to one particularly important patent, U.S. Patent No. 6,381,942.

Justice Kennedy, in his concurring opinion in eBay, instructed courts to be cognizant of the nature of the patent being enforced and the economic function of the patent holder when applying the equitable factors:

> When the patented invention is but a small component of the product the companies seek to produce and the threat of an injunction is employed simply for undue leverage in negotiations, legal damages may well be sufficient to compensate for the infringement and an injunction may not serve the public interest.

The Court finds this argument persuasive. While Apple undoubtedly uses the patent and produces goods in the same market, Apple has neither alleged, nor established, that the ’381 patent is either necessary to, or a core functionality of, the products that it seeks to enjoin. Nor has Apple shown that consumers’ purchasing decisions are based on the existence of a snap back feature protected by the ’381 patent.

Id. at *39–40 (citations omitted).

198. See Apple, Inc. v. Samsung Elecs. Co., 678 F.3d 1314, 1324 (Fed. Cir. 2012) We hold that the district court was correct to require a showing of some causal nexus between Samsung’s infringement and the alleged harm to Apple as part of the showing of irreparable harm. To show irreparable harm, it is necessary to show that the infringement caused harm in the first place. Sales lost to an infringing product cannot irreparably harm a patentee if consumers buy that product for reasons other than the patented feature. If the patented feature does not drive the demand for the product, sales would be lost even if the offending feature were absent from the accused product. Thus, a likelihood of irreparable harm cannot be shown if sales would be lost regardless of the infringing conduct.


200. See id. The “Nexus” phone was thus saved by the “nexus” requirement. This may not be a complete linguistic coincidence. Although other courts had previously articulated this concept, Judge Koh appears to be the first to use the term “nexus” to describe it.
Patent No. 8,086,604, which was related to the iPhone’s “Siri” personal assistant feature.  

In October 2012, however, the Federal Circuit ruled on appeal that the Galaxy Nexus injunction was improperly granted. The Federal Circuit insisted that the nexus requirement be applied even more stringently. According to the circuit-level opinion, Apple had failed to demonstrate that the infringement drove consumer demand for the Galaxy Nexus because Samsung’s feature incorporating the ‘604 patent was not among the top five reasons that consumers choose that phone. Articulated in this way, the Federal Circuit’s “causal nexus” standard seems extremely difficult to satisfy, at least in the context of smartphones. This ruling may signal the end of the injunction as a weapon in the smartphone patent war.  

Even as the Federal Circuit has made it more difficult to block a competitor’s phone through the courts, however, litigants are looking to other venues. Colleen Chien and Mark Lemley point out that patent holders are increasingly circumventing the new higher standards for injunctive relief by taking their cases to the International Trade Commission (ITC) instead. The ITC is a U.S. administrative agency that is empowered to issue “exclusion orders” blocking products it deems to be infringing from entering the U.S., achieving the same result as an injunction. According to Chien and Lemley, the ITC is increasingly the battleground of choice in the smartphone patent war and continues to be liberal in granting exclusion orders.  

Given the myriad of patents relevant to smartphone technology, an aggressive and persistent litigator can launch

---

201. See id. at *55–56.  
203. See id. at *10–11 (“Apple’s own survey evidence shows that unified search is not one of the top five reasons consumers select Android smartphones. In this light, the causal link between the alleged infringement and consumer demand for the Galaxy Nexus is too tenuous to support a finding of irreparable harm.”).  
204. See Chien & Lemley, supra note 193, at 102–03. (“[T]he ITC is busier with patent cases than it ever has been before.”).  
205. See id. at n.15 (citing individual cases involving Apple, Samsung, Sony, LG, HTC, Motorola, Nokia, and RIM). “In 2011, every major smartphone maker was embroiled in an ITC dispute.” Id. at 104.
many volleys and advance through attrition. Edison’s own litigation campaign met with modest success for many years, until he achieved a surprising victory in a German court, which then formed the basis for the next round of decisive litigation in the U.S. Similarly, the smartphone patent war is now being waged on many fronts, with suits active in England, Germany, South Korea, Japan, and elsewhere, and is far from over.206

C. Implications for Innovation

Supplementing the theory of patent racing with a theory of patent warfare casts a more skeptical light on the innovation hypothesis. The initial fervor to claim patent rights in the early electrical industry stimulated greater investment in first-generation research.207 This is consistent with the predictions of Lemley’s patent racing theory.208 Over the long term, however, patent protection also led to a dramatic reduction of competition in the electrical industry.209 As the industry matured, the more well-funded companies used patent warfare to target and eliminate their competition.210 Regardless of who emerges as the winner in Risk, the endgame is always the same: one party controls the entire board. In the real-life version, there may also be a détente where two parties agree to tolerate each other in separate spheres—as General Electric and Westinghouse divided the U.S. and British markets. Either way, the very competition that stimulated faster innovation during the racing stage is ultimately brought to an end.

206. See Carrier, supra note 4, at 3 (noting the importance of the German courts with respect to patent law).

207. See supra Part III.B (supporting the proposition that the initial fervor to claim patent rights stimulated substantial investment in first-generation research).

208. See supra Part III.A–B (noting Lemley’s predictions and discussing how the historical record of the beginnings of light bulb inventing is consistent with Lemley’s predictions).

209. See supra Part IV.A (noting the negative effects of patent law on industry research).

210. See supra Part IV.A (discussing the tactics of well-funded companies to use patent litigation to eliminate competition).
Patent warfare may harm innovation in at least three ways. First, litigation is a resource-intensive distraction of capital and human resources. The Edison interests, for example, spent $2 million on litigation at a time when the company's net worth was only $12 million. These financial resources, as well as Edison's own talents, might otherwise have been invested in improving the product and delivering it more efficiently to consumers. Second, a litigious environment can discourage the risk-taking of investing in next-generation innovation because would-be innovators fear they will be legally prevented from producing the newer technology. Third, as the pressure to stay one step ahead of the competition is eased, the competitive motivation to keep improving the technology is diminished. Bright notes this dynamic in the early electrical industry: as the light bulb litigation heated up, technological innovation dimmed. In the aftermath of patent warfare, General Electric was able to prevent any other light bulb manufacturer from exerting downward pressure on prices; the only real competitive pressure came from further advances in the substitute technology of gas lighting. This is not to say that light bulb innovation ground to a halt. General Electric still had profit-based incentives to cut its own production costs by introducing more efficient technological improvements. But the company no longer risked

---

211. See Bright, supra note 2, at 86 (estimating that "between 1885 and 1901 the Edison company and its successors spent about $2,000,000 on well over two hundred infringement suits under its lamp and lighting patents"). Defendants of the lawsuits probably had to spend nearly the same amount.

212. See id. ("The final step in the transition of the Edison companies from domination by Thomas A. Edison to domination by financiers took place in 1889 [merging] the remaining separate Edison development and manufacturing companies into the Edison General Electric Light Company. . . . The new company had a capitalization of $12,000,000 . . . .").

213. See id. at 138 (indicating that "[t]he lengthy and expensive patent struggle in the lamp industry from 1885 to 1894 was a serious damper on progress in lamp design, although process improvement continued").

214. See id. at 457 (discussing the realities surround General Electric's dominance, including the volume of the market controlled by General Electric, the lack of developmental work available to competitors, and the generally passive role licensees).

215. See id. at 455 (noting that General Electric "possessed immense ability to achieve technological advances . . . [and] . . . had strong incentives to make improvements in lamp design and production methods").
being overtaken by a competitor if it lagged behind in the innovative endeavor. As competition in the electric industry slowed and prices for equipment and service remained high, millions of Americans literally waited in the dark. More than a half-century would pass after Thomas Edison filed his first patents before electric light would find its way into the average American home.

Importantly, the potential long-term damage to next-generation innovation is not limited to the twenty years of the modern patent term. By the time first-generation patents become obsolete, newer patents already exist to take their place. When Edison’s patents expired, General Electric was able to use newer ones to extend its dominance. The result in the case of the light bulb was that General Electric would control the electric lighting industry for many decades longer. Even as Arthur A. Bright published his thesis on the electric-lamp industry in 1949, General Electric’s monopolization of the electric lighting industry remained a problem. Once patent warfare produces the consolidation of an industry, reintroducing competition is not a simple matter.

This anticompetitive endgame is my central concern with the use of patent racing theory as a new justification for patent protection. The racing metaphor paints a picture of objectivity and fair competition. In the case of the light bulb, however, the race was merely a prelude to the war. Patents almost certainly add fervor to inventive races at a certain stage of research and development. But at a later stage, patents also become weapons that competitors use to threaten and dominate their competition, long before the public policy goal of wide diffusion of the new technology has been fully attained. This may also have the perverse effect of ultimately limiting next-generation innovation. In the end, enthusiasm for the competitive advantages of races may be an argument against patent protection, rather than an argument in its favor.

216. See id. at 458 (noting General Electric’s ability to exert control over the patent market).

217. See id. at 458–60 (noting General Electric’s ability to exert control until World War II “with the aid of a strong patent position, commercial and technical aggressiveness, strong control over lamp parts and machinery, extensive advertising, and good public relations”).
The story of the light bulb thus turns out to be a cautionary tale against reliance on the innovation assumption. The downsides of patent protection are probably systematically underappreciated. Partly this has to do with our methodology of study. Empirical study of recent fields of technology is best at highlighting the initial, short-term impacts of patent protection. These include assistance in recruiting capital to fund research and product development, and in outsourcing production chains—both of which are positives for innovation. Only historical study can reveal the longer-term impacts that patent litigation produces on an industry. These may be systematically more negative as patents become weapons in attempts to monopolize a field. Further research is needed to determine whether patent wars are typical or exceptional features of technological development in the shadow of patent law, identify what factors motivate different outcomes in patent wars, and clarify the impact of industrial consolidation on next-generation innovation as well as diffusion.

V. Conclusion

Although patent law is founded on the assumption that patent protection encourages innovation, there is little empirical support available to test this “innovation hypothesis.” This Article has argued that historical case studies can offer a stronger empirical basis for patent scholarship by illuminating the true impact of patent law on innovation throughout the life of a patent term. The present case study demonstrates this potential, both by suggesting refinements to one theory of patent racing, and by proposing a new theory of patent warfare resembling the board game Risk. Patent warfare transformed the electrical industry at the close of the nineteenth century and has now erupted in the smartphone sector. The case study of the light bulb shows that patent law can have both positive and negative impacts on technological innovation, and we ignore the complexity of these dynamics at our own peril.

218. See supra Part II.B (noting case studies of patenting that support the proposition that patent law harms industry and innovation).
The Law School Critique in Historical Perspective

A. Benjamin Spencer*

Abstract

Contemporary critiques of legal education abound. This arises from what can be described as a perfect storm: the confluence of softness in the legal employment market, the skyrocketing costs of law school, and the unwillingness of clients and law firms to continue subsidizing the further training of lawyers who failed to learn how to practice in law school. As legal jobs become increasingly scarce and salaries stagnate, the value proposition of law school is rightly being questioned from all directions. Although numerous valid criticisms have been put forth, some seem to be untethered from a full appreciation for how the current model of legal education developed. Indeed, a historical perspective on legal education is sorely missing from this debate, as many of the criticisms merely echo charges that have been lodged against legal education for well over a century, but do not draw lessons from how those former critiques ultimately failed to deliver fundamental change. This Article reviews the historical development of legal education in America, including the critiques and reforms made along the way, to gain insight that will inform our own efforts to make law schools better at preparing lawyers for practice.

* Associate Dean for Research, Frances Lewis Law Center Director & Professor of Law, Washington and Lee University School of Law; Chair, Virginia State Bar Section on the Education of Lawyers. Copyright © 2012 A. Benjamin Spencer. This Article was written during my time as a Visiting Professor at the University of Virginia School of Law, which I would like to thank for its generous grant assistance that enabled this research and for the wonderful research assistance provided by its reference librarians. I would also like to thank those who were able to give helpful comments on the piece, including participants of a workshop on this piece at the University of Virginia School of Law and the Cardozo School of Law.
Table of Contents

I. A Perfect Storm ............................................................. 1951

II. From Blackstone to Langdell ........................................ 1960
   A. Private Reading and Office Apprenticeships ........ 1961
   B. Professorships in Law ............................................. 1964
   C. The Litchfield Law School ....................................... 1966
   D. University Law Schools ........................................... 1968
   E. Langdell and the Modern American Law School .... 1973
       1. The Case Method ............................................... 1973
       2. The Cloistered Law Professor ........................... 1975
       3. The Socratic Method .......................................... 1977
       4. Other Reforms ................................................... 1978
       5. Spreading the Langdellian Model ..................... 1979

III. Over a Century of Critique and Reform ........................ 1982
   A. Early ABA Reports .................................................. 1982
   B. The Redlich and Reed Reports on Legal Education .... 1988
   C. Twentieth Century Critiques and Reforms ............ 2001
   D. Contemporary Critiques .......................................... 2009

IV. The Current State of Legal Education .......................... 2015
   A. The Curriculum ....................................................... 2020
   B. The “Signature Pedagogy” of Law School ............... 2026
   C. Assessment in Law School ...................................... 2039
   D. Law School Faculty .................................................. 2047

V. The Next Century in Legal Education ............................ 2054

VI. Conclusion ...................................................................... 2061
I. A Perfect Storm

“The training received by American lawyers to-day has thus become a curious complex. It cannot be understood unless its main features, and the causes that have produced them, are presented in their historical sequence.”

—Alfred Zantzinger Reed, 1921

Legal education is under attack. The value of a law degree is being questioned given the deterioration of the traditional legal job market and the substantial and growing size of the student

1. ALFRED Z. REED, TRAINING FOR THE PUBLIC PROFESSION OF THE LAW: HISTORICAL DEVELOPMENT AND PRINCIPAL CONTEMPORARY PROBLEMS OF LEGAL EDUCATION IN THE UNITED STATES, WITH SOME ACCOUNT OF CONDITIONS IN ENGLAND AND CANADA, Bulletin No. 15, at 4 (1921) [hereinafter REED REPORT].


I regret law school, but more specifically I regret not fully considering the debt I would be facing after law school and weighing that against the prospect of committing my 20s to paying off that debt. Finding out I had chosen the wrong profession for me was bad enough, but being trapped in that career for many years just to break even was worse.

loan debt of recent graduates. Further, law schools are being charged with failing to prepare their graduates adequately for practice. Thus, we have what appears to be a perfect storm in

status was known, only 68.4% had jobs requiring bar passage, with only 64% overall being full-time legal positions. See Press Release, Nat’l Ass’n for Law Placement (NALP), Class of 2010 Graduates Saddled with Falling Average Starting Salaries as Private Practice Jobs Erode 3 (July 7, 2011), available at http://www.nalp.org/uploads/PressReleases/Classof2010StartingSalaryFindings PressRelease.pdf (providing a summary of the employment status of Class 2010 law students).

5. See William D. Henderson & Rachel M. Zahorsky, The Law School Bubble, 98 A.B.A. J. 30, 30 (2012) (“In 2010, 85 percent of law graduates from ABA-accredited schools boasted an average debt load of $98,500 . . . .”). Student loan repayment burdens are worse for students graduating from law schools placed at the lower tiers of the U.S. News rankings, as they tend to have median annual income well below graduates of higher ranked schools. See NALP Found. for L. Career Res. & Educ. & Am. B. Found., After the JD: First Results of a National Study of Legal Careers 42 (2004) [hereinafter NALP, After the JD] (providing an in-depth study of the legal profession, including demographics, work satisfaction, compensation, professional mobility, and legal education).

6. See, e.g., William R. Rakes, Conclaves on Legal Education: Catalyst for Improvement of the Profession, 72 Notre Dame L. Rev. 1119, 1119 (1997) (“The practicing bar is demanding that law schools provide more training to prepare graduates to hit the ground running when they enter practice.”); John Caher, N.Y. State Bar Asks ABA to Support ‘Practice Ready’ Law School Education, LAW.COM (Aug. 5, 2011), http://www.law.com/jsp/article.jsp?id=12025095 95910&slreturn=1 (last visited Aug. 25, 2012) (“Legal education should have more of an emphasis on making sure graduates are ready to practice law . . . . It is something that has been de-emphasized, and it shows. Our research and our own experience show that graduates are less prepared to practice law.”) (on file with the Washington and Lee Law Review); David Segal, What They Don’t Teach Law Students: Lawyering, N.Y. TIMES, Nov. 19, 2011, https://www.nytimes.com/2011/11/20/business/after-law-school-associates-learn-to-be-lawyers.html?_r=1&pagewanted=all (last visited Sept. 11, 2012) (“What they did not get, for all that time and money, was much practical training. Law schools have long emphasized the theoretical over the useful . . . .”) (on file with the Washington and Lee Law Review). Throughout this Article, I will use the phrase “prepared for practice” or “practice-readiness” to refer to students who have sufficient mastery of the specialized knowledge, skills, and values of the legal profession to serve as competent attorneys and counselors without supervision. Thus, practice-readiness is not simply about having the technical skills required for practice but also about having a solid understanding of legal doctrine, acute analytical abilities, and an understanding of the ethical framework within which lawyers must operate.

7. David Thomson uses this phrase to describe the current situation in legal education in which changes in the legal market, dissatisfaction with law schools, and a new generation of students with more technological savvy are combining to make this a critical moment for change. See David I. C. Thomson,
legal education: Law school graduates are under-employed, over-indebted, and under-prepared for practice.

As the economic downturn has forced legal service providers to deliver services with leaner staffs or outsourced resources, technology has developed to the point that fewer attorneys are needed to complete many legal tasks than was the case in the past. For example, artificial intelligence has advanced such that computerized document review has become faster, cheaper, and of a higher quality than traditional human review. This means that many lost legal jobs, related to discovery work or other legal tasks, may not be coming back. Additionally, law firms and

---


9. See Karen Sloan, Elite Firms Seem to Have Lost Their Appetites, Nat’l Law J. (Feb. 27, 2012), http://www.law.com/jsp/nlj/PubArticleNLJ.jsp?id=1202543428334&sreturn=1 (last visited Aug. 25, 2012) (“Some of that more routine legal work that used to be handled by a lot of associates is now being done by contract attorneys or outside providers.”) (on file with the Washington and Lee Law Review); THOMSON, supra note 7, at 16 (“Many corporations send routine legal work to large shops of relatively low-paid attorneys in India. Law firms are doing it too—in a recent study, 80 percent of the largest firms admitted to having outsourced projects.”).


11. See, e.g., id. (“Research has shown that, under the best circumstances, manual review will identify about 70 percent of the responsive documents in a large data collection. Some technology-assisted approaches have been shown to perform at least as well as, if not better, at far less cost.”).


I argue that the market is increasingly unlikely to tolerate expensive lawyers for tasks (guiding, advising, drafting, researching, problem-solving, and more) that can equally or better be discharged by less expert people, supported by sophisticated systems and processes. It follows, I say, that the jobs of many traditional lawyers will be substantially eroded and often eliminated.
corporate law departments have recognized that there are some legal tasks that do not necessarily require the same level of training or expertise to complete. Thus, some firms have established differentiated career tracks that create nonpartner-track staff attorneys who receive lower compensation for completing more routine and less demanding legal tasks. Corporate law departments have kept more work in-house with their employees or used low-paid contract attorneys to do routine legal work. Two commentators cited statistics predicting that the paucity of legal jobs will be stark going forward, as the American Bar Association (ABA)-approved law schools will be producing roughly 45,000 graduates annually to fill only 25,000 lawyer positions.

13. See Rama Lakshmi, U.S. Legal Work Booms in India, WASH. POST, May 11, 2008, at A20 (noting that in 2008, the legal outsourcing industry had grown “about 60 percent annually” for the past three years because corporate firms outsource work spurred by the passage of e-discovery laws).


15. One corporate general counsel reports, “[i]t’s simply a situation where, for the most part, law firms have priced themselves out of a whole bunch of work I used to have them do . . . . If work is going to repeat at all, I’ll bring the expertise in-house. My in-house teams have simply gotten much bigger, and my outside counsel use has gone down . . . .” Jay Pinkert, Straight Talk from General Counsels on How to Win Their Business, SHATTERBOX (Feb. 16, 2012), http://www.shatterbox.biz/2012/02/straight-talk-from-general-counsels-on-how-to-win-their-business/ (last visited Aug. 25, 2012) (quoting Eric Whitaker, General Counsel of Tesla Motors) (on file with the Washington and Lee Law Review).

16. See Ashby Jones & Joseph Palazzolo, What’s A First-Year Lawyer Worth?, WALL ST. J., Oct. 17, 2011, at B1 (“Of course, the more menial tasks still need to get done. But many corporate legal departments are either farming them to their own employees or giving them to so-called contract attorneys, lower-cost outside lawyers who work independently from the large law-firm ecosystem.”).

This bursting of the legal job market bubble has laid bare two major deficiencies in legal education. First, its costs have risen to levels that could only be sustained as long as cheap student loans and high-paying legal jobs were available, conditions that are now eroding. Second, the Great Recession has led clients and thus law firms to have less capacity to subsidize the on-the-job training of law graduates that they had been expected to provide, revealing deficiencies in the ability of


18. See Nancy M. Jackson, Student Loans Becoming Scarce, FOX BUS. (Sept. 20, 2011), http://www.foxbusiness.com/personal-finance/2011/09/20/student-loans-becoming-scarce/print (last visited Sept. 10, 2012) (“[T]he money to boost the Pell Grant program was raised by making cuts to the federal student loan program, eliminating the interest subsidy on loans to graduate students after July 1, 2012, and eliminating the Education Department’s Direct Loan Program repayment incentives.”) (on file with the Washington and Lee Law Review).

19. See Larry E. Ribstein, Practicing Theory: Legal Education for the Twenty-First Century, 97 IOWA L. REV. 1649, 1659 (“Law partners who have to spend more time tending to their books of business have less time for building the firm’s value through activities like training younger lawyers. Also, increased competition reduces firms’ freedom to bill training time to clients.”); see also Jones & Palazzolo, supra note 16 (citing a survey reporting that more than 20% of responding in-house legal departments refuse to pay for the work of first- or second-year attorneys); id. (quoting the general counsel of a company as saying “[T]he training someone on putting together an asset-purchase agreement shouldn’t be done on our nickel”); THOMSON, supra note 7, at 18 (“[F]irms are less and less interested in taking on this role . . . . [W]ith annual salaries for new attorneys ballooning in some cities to $150,000 and more, firms tend understandably to be less patient with new associates who need significant training before they are truly useful.”).

20. See Carrie Hempel & Carroll Seron, An Innovative Approach to Legal Education and the Founding of the University of California, Irvine School of Law, in THE PARADOX OF PROFESSIONALISM: LAWYERS AND THE POSSIBILITY OF JUSTICE 169 (Scott L. Cummings, ed., 2011), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1851702& (“For the better part of the twentieth century, there has been an informal division of labor between law schools, which are in the business of credentialing knowledge, and first employers responsible for passing on the skills of day-to-day practice.”).
law schools to adequately prepare a sufficient number of their students to handle legal matters for clients.21

But this latter critique is not a new development; over the past 130 years we have heard from many sources22 that law schools are not truly fulfilling their obligation to prepare students for legal practice.23 This would strike many outside the profession as odd since—as Judge Richard Posner once suggested and as most would assume—the “basic focus” of law schools should be “the training of practicing lawyers.”24 This is far from a truism, in

21. See, e.g., N.Y. ST. B. ASS’N, REPORT OF THE TASK FORCE ON THE FUTURE OF THE LEGAL PROFESSION 38 (2011) [hereinafter REPORT OF THE TASK FORCE ON THE FUTURE OF THE LEGAL PROFESSION] (“Too many law students and recent graduates are not as well prepared for the profession as they might be.”).


23. Long ago, Albert Harno, in his 1953 study on legal education in America, noted the criticisms that were leveled at law schools, which, he wrote, “all can be grouped under one heading, that the schools do not adequately prepare students for the tasks they will have to perform in the practice.” Albert J. Harno, Legal Education in the United States 137 (1953); see also MacCrater Report, supra note 22, at 5 (“[S]urveys understandably indicate that practicing lawyers believe that their law school training left them deficient in skills that they were forced to acquire after graduation.”); Cramton Report, supra note 22, at 8 (“Chief Justice Burger and others have spoken, in recent years, of a serious problem of ‘incompetency’ among those lawyers trying cases before the federal courts and among the trial bar generally.”).

24. Richard A. Posner, The Present Situation in Legal Scholarship, 90 Yale L.J. 1113, 1129 (1981); see also Erwin Chemerinsky, The Ideal Law School for the 21st Century, 1 U.C. Irvine L. Rev. 1, 20 (2011) (“[L]aw schools exist preeminently for training students to be lawyers . . . .”). Harno defined the mission of legal education as the production of “well-trained and capable lawyers—lawyers who are skilled in legal procedures, who are versatile in the tasks of the law, who have an understanding and a vision of the purposes and mission of the law, and who are guided by a sense of moral responsibility.” Harno, supra note 23, at 164.
light of the perspective expressed by many that law schools should be primarily scholarly institutions in which the law can be studied and understood as an academic and intellectual pursuit rather than professional schools that provide vocational training.\textsuperscript{25} To be sure, there is ground to allow both perspectives;\textsuperscript{26} a law school must train lawyers but also can be “a

\textsuperscript{25} The late Charles E. Clark, Second Circuit Judge, former Yale Law School Dean, and father of the Federal Rules of Civil Procedure, was a harsh critic of efforts to involve law schools in vocational training in practical legal skills:

\begin{quote}
I regard the repetitive attempts to coerce law schools into offering so-called practical training as at best curiously naive, and in general at odds with sound concepts of legal education. Such attempts might be dismissed as a comparatively harmless and not unusual professional baiting of the schools except that law deans and professors are acutely attuned to professional criticism and hence may be led to waste their substance in doing what they cannot do effectively and what if they could would not be pedagogically worth while. . . . I shall argue that law school training is now effectively efficient . . . . that there is no real basis for the criticism implicit in this pressure for practical training; that the latter is limited, partial and fragmentary at best; and that the present-day legal education in problem analysis and exposition and in thorough documentation of sources is much more important and valuable, as well as more within the practical competence of the schools.
\end{quote}

Charles E. Clark, “Practical” Legal Training an Illusion, 3 J. LEGAL EDUC. 423, 423 (1951); see also Robert S. Summers, Fuller on Legal Education, 34 J. LEGAL EDUC. 8, 9 (1984) (“Some legal educators today hold that training of practitioners is at most a secondary function of law schools. Rather, the primary function is to create and disseminate new knowledge and understanding about law.”).

\textsuperscript{26} See Thomas Swan, Report of Thomas W. Swan, Dean, to the President and Fellows of Yale University, in REPORTS OF THE PRESIDENT, ACTING PROVOST AND SECRETARY OF YALE UNIVERSITY AND OF THE DEANS AND DIRECTORS OF ITS SEVERAL SCHOOLS AND DEPARTMENTS FOR THE ACADEMIC YEAR, 1919–1920, at 393 (1920), available at http://books.google.com/books/reader?id=RrNJAAAAYAAJ&printsec=frontcover&output=reader&pg=GBS.PP7 (“A university law school has two functions. It aims by the case method of instruction to train its students so that they may become successful practitioners in their chosen profession. It aims also or at least it should aim, though too few schools have recognized this obligation, to aid in improving the law . . . .”); see also LON L. FULLER ET AL., PRELIMINARY STATEMENT OF THE COMMITTEE ON LEGAL EDUCATION OF THE HARVARD LAW SCHOOL 4 (1947) (noting that the two purposes of training lawyers and promoting understanding and improvement of the law “are so closely related that it is unrewarding to discuss which is primary and which is secondary”).
centre [sic] of research, criticism, and contribution to the better understanding of the laws” with the goal of improving the law.27 It is for each school to determine the extent to which it is committed to the latter goal; however, it should be the aim of all to be effective at achieving the former.28

Unfortunately, the law school of today is not optimally designed to prepare students for practice. The focus across most of the three years of law school is on teaching legal doctrine,29 using principally a method of limited effectiveness,30 with too few students being thoroughly instructed in the practice skills and core competencies needed to be a successful lawyer.31 These

---


29. See Carnegie Report, supra note 22, at 3 (noting the concentration of courses on legal doctrine throughout all three years of law school).

30. Critiques of the Socratic case method abound. See, e.g., Roy Stuckey et al., Best Practices for Legal Education 112 (2007) (“[T]he Socratic dialogue and case method leaves students confused, teachers often use it poorly, and it contributes to a hostile, competitive classroom environment that is psychologically harmful to a significant percentage of students.”); id. at 97–104 (cataloging criticisms of the Socratic method); Harno, supra note 23, at 138–39 (“One of [the] limitations of the Socratic method is the fact that after the first year of law study, there is a distinct lag of interest in the reading of cases on the part of law students. The method by that time has lost much of its sparkle.”). The Socratic case-dialogue method will be discussed in greater detail in Part III.B below.

31. See, e.g., David E. Van Zandt, Foundational Competencies, 61 Rutgers L. Rev. 1127, 1136 (2009) (“[N]o school has addressed the core competencies that it takes to be an effective lawyer in a variety of organizations over a multi-job career.”). In fact, the practical skills training and experiential learning opportunities that many law schools offer tend to be elective and are taken up in earnest by a minority of law students overall, when compared with the extensive doctrinal instruction all law students receive. NALP Found., 2010 Survey of Law School Experiential Learning Opportunities and Benefits 6–7 (2011) [hereinafter NALP, 2010 Survey of Law School Experiential Learning Opportunities and Benefits], available at http://www.nalp.org/may2011research_exp_learning?s=2010surveyoflawschoolexperientiallearning
defects have their origins in the innovations of Christopher Columbus Langdell, the late nineteenth century Dean of the Harvard Law School who envisioned law as a scientific discipline that should be taught by full-time academics using the case-dialog method more so than as a craft to be learned by apprentices studying at the feet of experienced practitioners.\(^{32}\) Much has changed in law schools since Langdell’s time, but to a remarkable degree, much is still fundamentally the same. Appreciating both the legal education system to which Langdell was responding and the nature and rationale of Langdell’s various reforms is useful in permitting us to think critically about the basic design of law schools today, especially as we consider how that model should change in response to contemporary challenges.

But why should we care? Members of the bar are rightly giving their attention to this issue, but there are likely some academics who would scoff at the notion that one should concern oneself with this question or that any real problem exists. Whether there is a problem, no one should seriously doubt.\(^{33}\) Nor should anyone doubt whether it is for legal academics to devote our attention to this matter; maintaining the quality of the legal education system is vital to any effort to improve the law and the administration of justice.\(^{34}\) Furthermore, in an environment opportunities and benefits (showing 30.2% of survey respondents had participated in at least one legal clinic, 36.2% of respondents had taken part in an externship, and 40% of respondents had taken three or more practice skills courses).

\(^{32}\) See infra Part II.E (discussing Dean Langdell’s teaching philosophy towards the law).

\(^{33}\) Of course, there are those who doubt there is a problem. See, e.g., Steven Harper, The Law School Quandary, AMLAW DAILY (Jan. 20, 2012), http://amlawdaily.typepad.com/amlawdaily/2012/01/the-law-school-quandary.html (last visited Sept. 9, 2012) (“If the vast majority of students are happy with the law school experience and changing it won’t improve their job prospects, perhaps the legal academy and its critics should consider focusing attention elsewhere.”) (on file with the Washington and Lee Law Review).

\(^{34}\) William G. Hammond et. al., Report of the Committee on Legal Education, 14 AM. B. ASS’N REP. 320–21 (1891) [hereinafter Hammond et. al., 14 AM. BAR ASS’N REP.]

No lesson has been more clearly taught by the history of our science from the beginning than that, wherever the law has been best
where legal jobs are more difficult to attain and the costs and
debt associated with legal education are on the rise, allowing the
quality of legal education to erode through indifference is a
disservice to the students we serve.

Taking the need for reform of some kind to be necessary, the
first step in the reform process should be a thorough
consideration of what brought us to this point and why our
schools take the approach to legal training that they do. Why is
law school designed the way that it is today, as a three-year
program led by scholars removed from practice focusing on
doctrinal legal instruction? What efforts have already been made
to move legal education away from that model and why have
those efforts failed at achieving any fundamental alteration of
how we deliver legal education? What criticisms have been lodged
against legal education in the past and what insights might we
gain from them for our time? And what suggestions for reform
have been made but ignored that could offer ideas whose time
may now be ripe? Part II of this Article will plumb some of the
history of legal education in this country, while Part III traverses
the long line of critiques that have been leveled at legal
education. Part IV explores the state of legal education today,
with Part V featuring a discussion of the lessons to be learned
from this history for contemporary reform efforts.

II. From Blackstone to Langdell

The system of legal education that we have today is the
product of an evolutionary process in which subsequent
approaches reflect efforts to build on and improve what has come
before. This means that law school, as it exists today, is an
artifact of its past, with a structure and tradition that is rooted in
history more so than being founded on rational design. As a
result, although many innovations characterize the modern
approach to law school, these adjustments tend to be more

administered and most truly worthy of its high mission, its votaries
have been most careful of the education of students and in the
statement of its principles, so that such students could thoroughly
master them.
superstructure than substitute, supplementing traditional law school education rather than supplanting it.\textsuperscript{35} As I will argue, however, real and lasting change in legal education requires fundamental rethinking rather than accretive reform. Getting beyond evolution to a revolution in how we educate lawyers, however, requires that we first understand the history that has brought us to where we are. Thus, in this Part and the next, we will review the historical development of formal legal education to its present form to understand the “why” behind the current system, to identify some of the roots of its shortcomings, and to inform our current critique with some of the wisdom and insights from the past.

\textbf{A. Private Reading and Office Apprenticeships}

“Why disgust and discourage a young man by telling him he must break into his profession through such a wall as this?”

—Daniel Webster, 1858\textsuperscript{36}

Pre-Revolutionary lawyers in America imported the English common law and with it the rudiments of the English approach to training aspiring lawyers for practice.\textsuperscript{37} This consisted of the office apprenticeship, in which an aspirant was assigned the reading of classic common law texts of varying utility,\textsuperscript{38} and

\textsuperscript{35} See Carnegie Report, supra note 22, at 76 (“[T]oday’s trend is to supplement rather than replace the inherited reliance on this venerable case-dialogue teaching in the first phase of doctrinal instruction.”).

\textsuperscript{36} 1 Daniel Webster, Biographical Memoir, in The Works of Daniel Webster xiii, xxvii (11th ed. 1858) (reflecting on his legal training as an office apprentice, during which he was tasked with reading obscure commentaries on the English common law by Sir Edward Coke).

\textsuperscript{37} See Harno, supra note 23, at 27 (“In America, . . . legal education until well past the middle of the 1800’s followed the English pattern.”).

\textsuperscript{38} Joseph Story has commented on this phase of his legal preparation, writing that he was told to read Coke on Littleton during his apprenticeship, which he described as “intricate, crabbed, and obsolete learning.” Joseph Story, Autobiography, in Miscellaneous Writings of Joseph Story 1, 19 (1852). Story went on to state, “I took [Coke on Littleton] up, and after trying it day after day with very little success, I sat myself down and wept bitterly” but that “[w]hen I had completed the reading of this most formidable work, I felt that I breathed a purer air, and that I had acquired a new power.” Id. at 20. Daniel
placed “at the desk of some skilful [sic] attorney[,] in order to initiate them early in all the depths of practice, and render them more dexterous in the mechanical part of business.” Such preparation was sufficient to become an attorney given the weakness of—or even the absence of—regulations setting forth requirements to practice law during late seventeenth and early eighteenth centuries. For example, James Flint, an Indiana attorney, wrote in 1819, “Blackstone’s Commentaries are considered the great medium of instruction. The young man who has carefully read these, and who has for a short time written for a practicing attorney, is admitted to the bar.”

The apprenticeship approach to legal education left much to be desired. As Blackstone pointed out in his famous Commentaries in remarking on the office apprenticeship:

[A] lawyer thus educated to the bar, in subservience to attorneys and solicitors, will find he has begun at the wrong end. If practice be the whole he is taught, practice must also be the whole he will ever know: if he be uninstructed in the

Webster, who similarly was assigned Coke on Littleton as an apprentice, remarked, “A boy of twenty, with no previous knowledge of such subjects, cannot understand Coke. It is folly to set him upon such an author.” WEBSTER, supra note 36, at xxvii.

39. 1 WILLIAM BLACKSTONE, COMMENTARIES *32; see also Roscoe Pound, The Law School and the Professional Tradition, 24 MICH. L. REV. 156, 158 (1925) (“[T]raditional ideals of the common-law books and the professional tradition brought over from England by pre-Revolutionary lawyers [were] handed down from lawyer to lawyer in the apprentice training of the old-time law office.”).

40. Harno describes the early practice regulations as requiring a number of years of preparation, which did not mean formal legal education at the time, with New Jersey requiring an examination. See Harno, supra note 23, at 33.


42. The MacCrate Report aptly describes the deficits of the system when it states:

The experienced attorney received cheap labor in exchange for the use of his library, but provided the apprentice with very little in the way of actual legal training. Often, experienced attorneys were too busy practicing law to spend time with their apprentices. There was no guarantee that a skilled practitioner was an adequate teacher. Many apprentices spent their time tediously copying documents for their masters, not studying legal tenets.

MACCRATE REPORT, supra note 22, at 104 (citations omitted).
elements and first principles upon which the rule of practice is founded, the least variation from established precedents will totally distract and bewilder him.43

To that critique, Blackstone added that the apprenticeship model would fail to yield lawyers who could comprehend “arguments drawn a priori, from the spirit of the laws and the natural foundations of justice,” and would never attract “a gentleman of distinction or learning” to the bar, leaving it “wholly in[] the hands of obscure or illiterate men.”44 Much later, Roscoe Pound would criticize the apprenticeship model on different grounds: that it was ill suited for preparing lawyers in an increasingly urbanized and industrialized America.45 In Pound’s view, the growing complexity of the economic and social structure that came with industrialization “call[ed] for a deeper and wider training of lawyers than the training in rules of thumb and in procedure which was afforded by the law office.”46 As another twentieth century commentator noted, “That the study of law through mediums of that sort would produce lawyers of limited knowledge and perspective is not a subject for wonder.”47

This is not to say that the office apprenticeship system lacked its advantages. Pound identified certain benefits that were lost with the passage of the apprenticeship approach to academic instruction, including the “handing down of professional traditions from lawyer to lawyer” and “that the law student in his formative days came in contact immediately with the leaders of the bar”:

By daily contact he absorbed from them certain traditions, certain ideals of the things that are done and are not done by good lawyers, and a certain feeling as to what was incumbent on him as a member of the profession. We cannot transmit

43. BLACKSTONE, supra note 39, at *32.
44. Id.
45. See Pound, supra note 39, at 159 (describing how industrialization of society shifted the role of lawyers from “the great trial lawyer” to “the steward of the leaders of industry” and thus necessitated knowledge of business as well as law, something that could not be provided solely through the apprenticeship model).
46. Id.
47. HARNO, supra note 23, at 19–20.
Ultimately, however, this approach was not enough; it focused on lawyering as a craft with undue attention to the need for lawyers to have a fuller understanding of the law and the ability to engage in more sophisticated legal analysis in a society of increasing legal complexity.

B. Professorships in Law

Returning to Blackstone’s time, his proposed solution to the deficits of apprenticeship training was to make academic instruction in the law a prerequisite to office-based training. In his words, “The inconveniences [associated with the apprenticeship model] can never be effectually prevented, but by making academical [sic] education a previous step to the profession of the common law, and at the same time making the rudiments of the law a part of academical [sic] education.”49 The creation of professorships in law early in the history of our country within universities was an effort to respond to Blackstone’s suggestion and to his example as a professor in law at Oxford.50 The first such professorship was established in 1779

48. Pound, supra note 39, at 160; see also Michael Burrage, From Practice to School-Based Professional Education: Patterns of Conflict and Accommodation in England, France, and the United States, in The European and American University Since 1800, at 142 (Sheldon Rothblatt & Bjorn Wittrock eds., 1993) (“[P]rofessional schools often displaced or discredited alternative practice-based forms of professional education. There are, therefore, opportunity costs and another side of its history, the side of the losers, of the viable, traditional institutions directly under the control of practising [sic] professionals.”).

49. BLACKSTONE, supra note 39, at *32.

50. See Harno, supra note 23, at 23 (“One immediate effect traceable to Blackstone’s influence was the establishment of chairs of law in several American colleges and universities.”); see also Pound, supra note 39, at 160 (“One might say with truth, even if somewhat paradoxically, that American legal education begins with Blackstone’s professorship at Oxford.”); James Thayer, The Teaching of English Law at Universities, 9 Harv. L. Rev. 169, 170 (1895) (describing how “Blackstone’s example was immediately followed here [in the United States]” at William and Mary College, Harvard College, and Litchfield Law School).
at William and Mary and filled by George Wythe.51 Aided greatly by Blackstone’s *Commentaries*—which taught early American lawyers “the continuity, the unity, and the reason of the Common Law”52—figures such as St. George Tucker at William and Mary and James Kent at Columbia College taught in the Blackstone tradition and were pioneers in matters of legal scholarship.53 These professors were expositors and systematizers of the law, who lectured on legal history, the broad principles that served as the foundation for the law, and the jurisprudence of the day in various subject areas.54 Several of these law professorships were established at various American colleges during the late eighteenth and early nineteenth centuries.55 However, legal

51. HARNO, supra note 23, at 23.
52. CHARLES WARREN, A HISTORY OF THE AMERICAN BAR 177 (1911) [hereinafter WARREN, A HISTORY OF THE AMERICAN BAR].
53. See HARNO, supra note 23, at 23–26. St. George Tucker based his lectures on Blackstone’s *Commentaries*, supplemented with his own notes on American law. Id. at 23. His notes became the basis for an annotated edition of Blackstone. Id. James Kent also based his lectures on Blackstone and used his lectures as the foundation for his own *Commentaries on American Law*. Id. at 25.
54. To receive the Royall Professorship in law at Harvard College, the incumbent was required to deliver certain lectures, including:

[T]he theory of law in its most comprehensive sense; the principles and practical operation of the Constitution and Government of the United States . . . ; an explanation of the principles of the Common Law of England, the mode of its introduction into this country, and the sources and reasons of its obligation therein; also the various modifications by usage, judicial decision, and Statute; and, generally, those topics connected with law as a science which will best lead the minds of students to such inquiries and researches as will qualify them to become useful and distinguished supporters of our free system of government, as well as able and honorable advocates of the rights of the citizen.

1 CHARLES WARREN, HISTORY OF THE HARVARD LAW SCHOOL AND OF EARLY LEGAL CONDITIONS IN AMERICA 298 (1908) [hereinafter 1 WARREN, HISTORY OF THE HARVARD LAW SCHOOL] (quoting the statutes of the Royall Professorship); see also id. at 301 (quoting Isaac Parker as saying that “law is a comprehensive system of human wisdom” and, thus, a “science . . . worthy to be taught”).
55. Other professorships established during this period include the Royall Professorship of law filled in 1815 at Harvard College, a professorship of law at the College of Philadelphia in 1790, a professorship of law at Yale College in 1801, a professorship of law at the University of Maryland in 1816, and a chair of Law and Politics at the University of Virginia in 1826. HARNO, supra note 23,
education in universities under professors of law was conceived of as “a broad foundation for the further education of prospective lawyers,” not as a substitute for training to become a practicing lawyer. Thus, aspirants to a career as a lawyer remained obliged to pursue an office apprenticeship to be admitted to the bar, establishing a distinction between academic legal education in colleges and training for practice in law offices.

C. The Litchfield Law School

These law professorships did not immediately lead to the development of schools of law within universities as had been hoped. Indeed, the first established law school, the Litchfield

---

56. Id. at 27.
57. See MARIAN CECILIA MCKENNA, TAPPING REEVE AND THE LITCHFIELD LAW SCHOOL 59 (Oceana 1986) (“None of these professorships, as the record clearly shows, attempted to offer a complete technical education for law students.”); see also Pound, supra note 39, at 160 (“But their lectures were not and were not meant to be professional training in law.”).
58. See HAMMOND, supra note 23, at 27 (“The educational route to the practice of law was then, as it was to be for some time yet to come, through office apprenticeships, and those who attended these lectures were not relieved from traveling that route.”); see also REED REPORT, supra note 1, at 45 (“The special obstacle in northern states [to university based legal education], during these early years, was the still prevailing requirement of a period of clerkship.”). Admissions requirements in Virginia during this time were reported to have been weaker than those found in the northern states. See id. at 44 (“In [Virginia] ... the requirements for admission to the bar were already so weak that William and Mary's law department had no difficulty in securing its start at once.”).
59. This distinction held for the training of barristers in England at this time, as law professors at Oxford and Cambridge had no hand in their training, which was managed by the Inns of Court—the practitioner “clubs” that oversaw student barristers during their pupillage. See Burrage, supra note 48, at 144–45 (“Oxford and Cambridge professors of law never sought to challenge the Inns' control of the education of the common lawyers. They gave lectures on Roman law, international law, jurisprudence, and constitutional history but hardly referred to the law used by practising [sic] English lawyers.”).
60. See JAMES BARR AMES, THE VOCATION OF THE LAW PROFESSOR, in LECTURES ON LEGAL HISTORY 354, 359 (1913) (“The hopes that may have been entertained of developing schools of law out of these professorships were in the main doomed to disappointment.”).
Law School in Connecticut, was independent of the university system and arose as an expansion from office apprenticeships in the law office of Tapping Reeve. This became possible once state bar admissions authorities abandoned efforts to limit the number of apprentices that could study under an attorney at any given time. The course of instruction offered by Reeve at Litchfield covered all of Anglo-American private law with no special attention given to the law of any one state or to areas of public law. The weekly Saturday examinations were three hours in length and mainly oral, consisting of a thorough investigation of the principles of each rule of law, and not merely of such questions as can be answered from memory without any exercise of the judgement . . . .

An alumnus of the school, James Gould, would later become an associate of Reeve and subsequently shared in the lecturing responsibilities at the school. He restructured the curriculum to include topics such as master and servant, bailments, and real property. It is worth noting that Gould discouraged the reading of cases as the primary means of legal instruction, remarking:

I always dissuade them from reading reports in course, until they have acquired a pretty thorough knowledge of the outline

---

62. See Reed Report, supra note 1, at 45 (“What saved the situation . . . was the fact that attempts to limit the number of clerks who might study under one attorney were soon abandoned. This paved the way for the thoroughly natural development of a private attorney's law office into a private class or school . . . .”)
63. McKenna, supra note 57, at 63, 65–66.
64. Id. at 85 (quoting Catalogue of the Litchfield Law School 5 (photo. reprint) (1900)).
65. Id. at 93.
66. Id. at 64, 108. Other subjects included municipal law, baron and feme (husband and wife), parent and child, guardian and ward, executors and administrators, sheriffs and gaolers (jailors), contracts, fraudulent conveyances, inns and innkeepers, covenant broken, action of debt, action of account, notice of request, assumpsit, defenses to actions, private wrongs, evidence, systems of pleading, new trials, bills of exceptions, writs of error, practice in Connecticut, bills of exchange and promissory notes, insurance, charter parties, joint owners of vessels, partnership, stoppage in transitu, sailors' contracts, powers of chancery, criminal law, estates upon condition, modes of acquiring estates, devises, title by deed, and actions for injuries to things real. Id. at 64.
of the science by studying each principal [subject] separately; being fully convinced that reading in the former mode is of little comparative profit in an early stage of legal studies.67

Although the life of this school was only from 1784 to 1833, it had a remarkable record of producing illustrious graduates.68 Its other legacy was to take legal education some degree beyond the office apprenticeship with a more systematic approach to instruction in law, though still occurring nominally in a law office setting.69 Other independent schools, though less notable than Litchfield, were in existence during this time as well.70 In these schools, we see the attempt to wed academic education and practical training in one setting—the law office rather than the university. This approach, however, would not last.

D. University Law Schools

Some of the early holders of law professorships in universities were able to achieve the development of entire law schools within their institutions. Chief Justice Isaac Parker, the first holder of the Royall Professorship in law at Harvard and the leader of the Massachusetts Supreme Judicial Court, was keen for Harvard to found a graduate school to instruct aspiring lawyers in the law before entering “into the office of a counselor to obtain a knowledge of practice.”71 Harvard Law School was
founded in 1817, adding Asahel Stearns as a professor.\textsuperscript{72} Although Yale's first professor of law, Elizur Goodrich, ended his nine-year tenure in 1810 without founding a law school there, the head of a private law school in New Haven, Judge David Daggett, was appointed to succeed Goodrich in 1826, bringing his private law school into the College as Yale Law School.\textsuperscript{73} Several other universities soon followed and established law schools, but Harno reports that “[t]hese were not distinguished schools” and “not one that was in operation contemporaneously with the Litchfield School enjoyed at that time the favorable reputation of that School.”\textsuperscript{74}

More university law schools were founded throughout the nineteenth century, bringing the number to thirty-one by 1870.\textsuperscript{75} The schools were not offering the three-year law degree of today but rather offered one-to two-year courses of study consisting of lectures and readings of treatises in the areas of law considered important at the time: constitutional law, American jurisprudence, English common law, equity, pleading, evidence, bailments, insurance, bills and notes, partnerships, domestic relations, conflict of laws, sales, and real property.\textsuperscript{76} Liberal education in other topics such as history, philosophy, ethics, and the law of nature, although viewed as important in the effort to produce lawyers of professional excellence,\textsuperscript{77} was presumed to have been obtained elsewhere.\textsuperscript{78} As Harno points out, this presumption was a fallacious one, given that none of the law

\begin{thebibliography}{9}
\bibitem{72} Id. at 307.
\bibitem{73} See Warren, A History of the American Bar, supra note 52, at 364; see also Harno, supra note 23, at 38.
\bibitem{74} Harno, supra note 23, at 38.
\bibitem{75} Id. at 51.
\bibitem{76} 2 Warren, History of the Harvard Law School, supra note 54, at 218 (1908).
\bibitem{77} See Joseph Story, The Value and Importance of Legal Studies, in Miscellaneous Writings of Joseph Story 503, 529 (1852) (“It is by such studies, and such accomplishments, that the means are to be prepared for excellence in the highest order of the profession.”).
\bibitem{78} See id. at 536 (“[T]he course of the academical [sic] instruction in this University already provided for . . . the subjects of ethics, natural law, and theology, [which were] assigned to other professors.”).
\end{thebibliography}
schools established by 1870 required a college education as a prerequisite to admission.\textsuperscript{79}

Notwithstanding the development of formalized legal education within the university, the office apprenticeship remained the common method of preparing for a career as a lawyer.\textsuperscript{80} Indeed, Justice Story—upon being inaugurated into the Dane Professorship at Harvard Law School in 1829—confessed, after describing all that would be required to prepare a lawyer for practice, that “[l]ittle, indeed, of what has been sketched out in this discourse, can be attained by any academical [sic] instruction during the usual period assigned for the preparatory studies for the bar . . . . What we propose is no more than . . . something to assist the student in the first steps of his studies . . . .”\textsuperscript{81} This reflected the continuing sense that academic legal instruction was only preparatory for and supplemental to apprenticeship learning of the legal craft in law offices.\textsuperscript{82} Further, such limited training was sufficient preparation for the small-scale, local practices that were typical of the profession at the time.\textsuperscript{83}

The advances in the formalization of legal education described above were not accompanied by similar advances in the standards for bar admission. To the contrary, as university-based law school education enhanced the intellectual abilities of aspirants to law office apprenticeships, bar admissions standards

\textsuperscript{79} See Harno, supra note 23, at 51. Harvard Law School did not require a previous college degree as a prerequisite to admission until 1894, when it announced such a requirement for the class that would enter in 1896. \textit{Id.} at 82–83.

\textsuperscript{80} See id. at 39, 52; Lawrence M. Friedman, A History of American Law 241 (3d ed. 2005) (noting that by the 1840s, “[f]or most people, the path to practice still went through a clerkship at a lawyer’s office”).

\textsuperscript{81} Story, The Value and Importance of Legal Studies, supra note 77, at 532.

\textsuperscript{82} See Burrage, supra note 48, at 152 (“American universities . . . provided law lectures merely as a part of a liberal education as part of the education of a gentleman.”).

languished and waned under the influence of Jacksonian democratic ideology.\(^{84}\) Under this philosophy, screening mechanisms and educational prerequisites were distrusted as obstacles that made law practice accessible only to elites.\(^{85}\) The practicing bar thus remained hostile to formal legal education, declining to refine bar admissions requirements to include such education as a prerequisite to being licensed to practice.\(^{86}\) This is a skepticism that endured until the late nineteenth century, when the ABA first adopted a resolution calling for each state to require three years of study in law school before applying to sit for the bar examination.\(^{87}\) Still, in the early twentieth century, one finds bar examinations that bore little connection to the

---

\(^{84}\) See Richard P. Cole, *Orthodox and Heresy: The Nineteenth Century History of the Rule of Law Reconsidered*, 32 Ind. L. Rev. 1335, 1365 (1999) ("Within the context of the democratic and anti-monopolistic ideology of the Jacksonian period . . . professional legal education markedly declined. Local bar associations no longer set standards of admission for law practice and legislative statutes seemed to make it easier to become a lawyer." (citing David Ray Pake, *Heretics in the Temple: Americans Who Reject the Nation's Legal Faith* (1998))). Jacksonian philosophy also impacted the process by which judges were selected, leading many states to move toward an elected rather than appointed judiciary. See Larry C. Berkson, *Judicial Selection in the United States: A Special Report*, 64 Judicature 176, 176 (1980) ("People resented the fact that property owners controlled the judiciary. They were determined to end this privilege of the upper class and to ensure the popular sovereignty we describe as Jacksonian Democracy.").

\(^{85}\) See Harro, *supra* note 23, at 39 ("It was a period in American history that exalted the rights and powers of the common man . . . [T]his creed seemed to hold that all male citizens had the inherent right to practice law.").

\(^{86}\) See id. at 40, 78–79 ("A substantial part of the practicing bar was unconvinced, if not distrustful, of the benefits that might flow to a lawyer through either a university or law school training. This unbelief . . . stifled efforts to establish educational requirements for admission to the schools and to the bar.").

\(^{87}\) *Transactions of the Twentieth Annual Meeting of the American Bar Association*, 20 Am. B. Ass’n Rep. 31–33 (1897). The language of the adopted resolution was as follows:

*Resolved*, That the American Bar Association approves the lengthening of the course of instruction in law schools to a period of three years, and that it expresses the hope that as soon as practicable a rule may be adopted in each State which will require candidates for admission to the bar to study law for three years before applying for admission.

Id. at 31.
lessons from law school or the needed abilities of a practicing attorney.88

Thus, from the time of Blackstone through the establishment of university law schools, we can see that formalized legal education was focused on instruction in the English common law and modern American legal doctrine as a foundation for subsequent training in practical skills after law school under the tutelage of practicing attorneys.89 No effort was made to attend to the preparatory education students needed before taking up legal studies in universities, nor was any effort made at the university level to train students in practical legal skills.90 Academic legal instruction was not developed or designed to serve as a substitute for apprenticeship training in the practical skills of legal practice.91 Rather, it was a response to the sentiment reflected by Blackstone that an office apprenticeship by itself was not enough.92 In this context, academic legal education and subsequent practical training through apprenticeships were necessary partners in the effort to prepare well-qualified lawyers for practice. It makes sense, then, that with the background understanding that academic and practical training worked in tandem, academic legal education gave no attention to practical skills, focusing purely on instruction in legal principles and doctrines. What remained to be refined was the method to be used for such instruction and the character of those who would be principally called to deliver it.

89. See supra notes 60–64 and accompanying text.
90. See supra notes 74–79 and accompanying text.
91. See supra note 57 and accompanying text.
92. See supra note 44 and accompanying text.
E. Langdell and the Modern American Law School

“I entered upon the duties of my present position . . . with a settled conviction that law could only be taught or learned effectively by means of cases in some form.”

—Christopher Columbus Langdell, 1871

Christopher Columbus Langdell, who became the Dane Professor and Dean at the Harvard Law School in 1870, was a pivotal figure in the history of legal education in this country. During his time at Harvard Law School, he ushered in several innovations that characterize the modern American law school today.

1. The Case Method

The most well-known and enduring innovation Langdell introduced was the instruction of students in legal doctrine through the study of written opinions in decided judicial cases—the case method. Up to this point, the method of legal instruction in law schools was a combination of the lecture method and the text method, meaning students read texts that related and summarized particular bodies of law, and professors lectured on that material in class. As Theodore W. Dwight, Dean of the Columbia College Law School in New York, explained the method, “The student is assigned daily a certain portion of an approved text-book for his reading prior to listening to expositions of the subject involved . . . . [H]e is asked questions upon the topic . . . . Pertinent illustrations are resorted to . . . .”

94. 1 WARREN, HISTORY OF THE HARVARD LAW SCHOOL, supra note 54, at 359–60. Warren described Langdell as “a young man of no legal reputation . . . ., a man of no national fame, and a lawyer who had had substantially no court practice.” Id. at 360.
95. JOSEF REDLICH, THE COMMON LAW AND THE CASE METHOD IN AMERICAN UNIVERSITY LAW SCHOOLS 7–8, 12 (1914).
96. Theodore W. Dwight, Columbia College Law School, New York, 1 GREEN BAG 141, 145 (1889).
Such an approach did not jibe with Langdell’s view of the law; he firmly believed that law was an inductive science and that the best preparation for a career as a lawyer was to study and master its fundamental principles, not through second-hand accounts but right from the source:

Law, considered as a science, consists of certain principles or doctrines. To have such a mastery of these as to be able to apply them with constant facility and certainty to the ever-tangled skein of human affairs, is what constitutes a true lawyer; and hence to acquire that mastery should be the business of every earnest student of law. Each of these doctrines has arrived at its present state by slow degrees . . . in many cases through centuries. This growth is to be traced in the main through a series of cases; and much the shortest and best, if not the only way of mastering doctrine effectually is by studying the cases in which it is embodied. 97

In these remarks, one can see two important strains of thought that would have a tremendous impact on the development of law school education. First, Langdell’s understanding of what it means to be a lawyer—reflected in his comment that “mastery” of “certain principles or doctrines . . . is what constitutes a true lawyer”98—gives rise to his view that the best preparation for practice was to study legal rules—principles and doctrines—rather than learning the skills of legal practice—law as mere “handicraft.”99 Second, Langdell’s understanding of the law—that it was a science whose study would be most effective when focused on examining the sources of legal principles and doctrines—leads to his determination that reported cases must be studied rather than summary expositions of the law contained in texts and treatises. 100 In other words,

97. LANGDELL, supra note 93, at vi.
98. Id.
100. REDLICH, supra note 95, at 15; see also 1 WARREN, HISTORY OF THE HARVARD LAW SCHOOL, supra note 54, at 423 (“The Case System then proceeds on the theory that law is a science and as a science should be studied in the original sources, and that the original sources are the adjudged cases, and not the opinions of text writers based on the adjudged cases.”); id. at 421 (“[T]he student must look upon law as a science consisting of a body of principles to be
Langdell believed that law was a form of natural science in that it consisted of a coherent system of rules derived from general principles that could only be discerned through the study of observable phenomena—the judicial opinions in which the principles were manifested. These views of the law and of lawyering supported and furthered the move away from practitioner-based apprenticeship in favor of formal instruction by full-time, legal scholars.

2. The Cloistered Law Professor

Indeed, the conversion of the professoriate into cloistered academics, rather than experienced practitioners, was Langdell’s next innovation. With law school being characterized by a curriculum focused on doctrinal study, Langdell remarked that “[w]hat qualifies a person, therefore, to teach law is not experience in the work of a lawyer’s office, not experience in dealing with men, not experience in the trial or argument of causes—not experience, in short, in using law, but experience in learning law . . . .” As a result, Langdell took the bold step in the fall of 1873 of hiring an assistant professor who had only graduated from law school two years prior and whose experience was limited to serving as an instructor in French, German, and

---

101. See Edward Rubin, What’s Wrong with Langdell’s Method and What to Do About It, 60 Vand. L. Rev. 609, 632–33 (2007) (“Since Langdell saw this methodology as a means of discerning general, objectively identifiable principles of law, he really seemed to believe that his approach was natural science itself.”).

102. See Carnegie Report, supra note 22, at 5 (“[L]egal positivists . . . viewed the law as . . . a set of rules and techniques rather than a craft of interpretation and adaptation . . . . All this spelled the eclipse of traditional forms of practitioner-directed apprenticeship by academic instruction given by scholar-teachers.”).

103. 1 Warren, History of the Harvard Law School, supra note 54, at 361 (quoting a speech made by Dean Langdell at the dinner of the Harvard Law School Association on November 5, 1886).
history at Harvard College—James Barr Ames. Prior to that time, Harvard’s law professors had been drawn primarily from the practicing bar. As Harvard President Eliot explained, however, the idea of hiring young, inexperienced persons to teach law was borrowed from the law schools of Continental Europe:

Those schools have selected young men of mark who have shown a genius for law and a desire for the life of a teacher, and . . . made them Professors, at an age so early that the whole vigor of their youth and prime could be thrown into teaching and authorship.

President Eliot went on to predict, quite presciently, that this innovation would take hold and fundamentally change the legal academic profession:

Professor Langdell early advocated the appointment, as teachers of law, of young men who had had no experience whatever in the active profession . . . . Now that experiment . . . has not only been extended in our own Law School with perfect success but it has been adopted by various other law schools throughout the country . . . . In due course . . . there will be produced in this country a body of men learned in the law who have never been on the bench or at the Bar, but who nevertheless hold position of great weight and influence as teachers of law, as expounders, systematizers and historians. This, I venture to predict, is one of the most far reaching changes in the organization of the profession that has ever been made in our country.

Eliot rejected the analogy to medical school training, in which practitioners are acknowledged to be the best instructors of clinical training, by insisting—as did Langdell—that “law is to be learned almost exclusively from the books in which its principles and precedents are recorded, digested, and explained,” and, thus, “[t]he law library, and not the court or the law office, is the real analogue of the hospital.” With this vision, the career, legal

104. *Id.* at 388 (stating that the appointment was an “experiment” and an “important innovation” for the law school).

105. See *id.*

106. *Id.* at 388–89.

107. *Id.* at 389.

108. *Id.* at 391–92.
professoriate was born, purely academic in character and divorced from the practicing bar.

3. The Socratic Method

A critical companion to the case method was another Langdellian innovation: the instructional approach used to discuss the cases with students in class known as the Socratic method. Langdell conducted his own courses using this method, having a student analyze the facts and the law in each case, followed by a series of questions designed to reveal the legal principles to be found therein. In addition to learning legal doctrine, Langdell’s disciples and subsequent adherents have rightly noted that the case-dialogue method was also designed to train students to develop the analytical abilities of a lawyer: “The student is challenged to reconcile discrepancies, to explain conflicts, to pick up the tangled threads of thought where they are left by the decisions and put them in order.” By employing the dialogue approach to the exploration of cases, the idea was that the case method would not only facilitate the scientific discovery of legal principles from their sources, but it also would contribute to the development of students’ ability to engage in legal reasoning, analysis, and synthesis.

109. Redlich, supra note 95, at 12. Redlich described the method as:
Teacher and pupils... work[ing] together unremittingly to extract from the single cases and from the combination or contrasting of cases their entire legal content, so that in the end those principles of that particular branch of the law which control the entire mass of related cases are made clear.

Id.

110. Harno, supra note 23, at 137–38 (quoting the Preliminary Statement of the Committee on Legal Education of the Harvard Law School); see also 1 Warren, History of the Harvard Law School, supra note 54, at 421 (“The student is required to analyze each case, discriminate between the relevant and irrelevant, between the actual and possible grounds of decision. And after having thus discussed a case, he is prepared and required to deal with it in its relation to other cases.” (quoting William A. Keener, A Selection of Cases on the Law of Quasi Contracts, Preface (1888))).

111. See William A. Keener, The Inductive Method in Legal Education, 17 Am. B. Ass’n Rep. 473, 482 (1894) (“While the student’s reasoning powers are being thus constantly developed, and while he is gaining the power of analysis
The case-dialogue method of instruction was seen as an application of the scientific method to the study of law at a time when American colleges were moving away from a classical approach to higher education toward methods reflected in the modern study of science. \(^{112}\) Langdell recognized this trend and the need to conform the study of law to it in order to retain its place within the university: “If law be not a science, a university will best consult its own dignity in declining to teach it. If it be not a science, it is a species of handicraft, and may best be learned by serving an apprenticeship to one who practices it.” \(^{113}\) Langdell’s contemporary, Harvard Law School Professor William Keener, was more explicit in drawing the connection when he wrote, “[I]f [law] is not a science it has no place in the curriculum of a university.” \(^{114}\)

4. Other Reforms

Langdell was also responsible for several other major reforms that transformed university-based legal education to the model we know today. In 1871, the course of study at the Harvard Law School was extended from eighteen months to two years, a change prompted by the realization that the case method required a longer period of time to be successful than eighteen months could allow. \(^{115}\) Strict examinations were introduced as prerequisites to proceeding to the next year of study and to receiving the degree. \(^{116}\) These changes, along with the introduction of the case

\(^{112}\). See REDLICH, supra note 95, at 16 (describing how “the old, strictly classical ideal of college instruction in America” was “yielding in favor of a new theory of education” that was based predominantly on the study of the natural sciences).

\(^{113}\). 1 WARREN, HISTORY OF THE HARVARD LAW SCHOOL, supra note 54, at 374.

\(^{114}\). Keener, supra note 111, at 475.

\(^{115}\). See 1 WARREN, HISTORY OF THE HARVARD LAW SCHOOL, supra note 54, at 379 (“Firm in his belief that law was a science, to be taught and learned as such, Langdell was equally convinced that no student could receive a proper legal education in so short a period as eighteen months.”).

\(^{116}\). See id. at 380, 397 (noting that students “shall have passed satisfactory
method, were initially controversial among the practicing bar, leading to a decline in enrollments compared with the previous decade. In 1875, an admissions examination was instituted and required of all applicants, beginning in 1877, who lacked prior collegiate education. Finally, for the 1877–1878 academic year, Langdell extended the course of study to three years. Rather than result in a reduction of enrollment, however, these latter two changes were followed by an increase in both the size and quality of the law school’s student body.

5. Spreading the Langdellian Model

By the early part of the twentieth century, formalized university-based legal education was well established and highly developed compared with the system of legal education in other common law countries. These law schools were increasingly Langdellian; although Langdell ended his tenure as Dean of the Harvard Law School in 1895, this case method approach to legal examinations at the end of each year” and that “the whole tone of the School changed from laxity to strictness”).

117. See id. at 382, 386 (“[T]here was much criticism and even bitter opposition among lawyers . . . over the new Case System employed by Professor Langdell. Undoubtedly this kept many students from entering the School.”).

118. See id. at 394–95, 399 (noting that in 1875, the school established an admission exam applicable “to all candidates for the of the School who are not already Bachelors of Arts, Science, or Philosophy”).

119. Id. at 399.

120. See id. at 400 (“[T]he prediction . . . that the establishment of an examination, together with an extension of the course of study from two years to three years, would cause an immediate and material diminution of our numbers, had not thus far been verified.”).

121. REDLICH, supra note 95, 5–6. In comparing the advancement of the American legal education system to other common law countries, Redlich noted:

[T]he existing American system of legal education has hardly a rival worth mentioning in the entire great jurisdiction of the English common law. Neither in England itself, nor in the great English colonies, has systematic instruction in law achieved a development so intensive and at the same time so comprehensive as in the United States.

Id. Redlich also reports that in 1914 there were more than 150 American law schools, with over 20,000 students. Id. at 7 n.1.
education was continued by his successor, James Barr Ames.\textsuperscript{122} When William Keener—who had been a Harvard professor under Langdell—joined Columbia’s faculty in 1890, he brought the case method with him.\textsuperscript{123} By 1906, the Langdell method had spread to law schools at Chicago, Northwestern, Cincinnati, Stanford, Wisconsin, Hastings, and Tulane—among others.\textsuperscript{124} Yale had been an early holdout, preferring a method of instruction that involved lectures and recitations.\textsuperscript{125} By 1913, however, the Langdell method dominated the classes in all three years at Yale; in 1916 a Harvard Law School graduate—Thomas Swan—was appointed as Yale’s dean.\textsuperscript{126} While much of this expansion was due to the direct transmission of the method by Harvard professors and graduates, who were its adherents, this “Harvardization” of American law schools was also motivated by the desire of prestige: “[O]nce elite law schools had decided to approve of the system, those aspiring to be considered elite rapidly followed.”\textsuperscript{127} No doubt the financial benefits of the method were an attraction as well; because it could be used with classes of large size, larger tuition-paying student bodies could be supported with relatively few professors.\textsuperscript{128} Regardless of the

\textsuperscript{122} See Proceedings of the Seventh Annual Meeting of the Association of American Law Schools, 31 AM. B. ASS’N REP. 1010, 1025 (1907) (“[W]e believe that men who are trained, by examining the opinions of the greatest judges . . . are in a better position to know what legal reasoning is and are more likely to possess the power of solving legal problems . . . .”).

\textsuperscript{123} Robert Stevens, Law School: Legal Education in America from the 1850s to the 1980s, at 60 (1983). Stevens reports that other faculty members, including Theodore Dwight, opposed the case method and resigned beginning in 1891. Id. This defection gave birth to The New York Law School. Id. at 75 n.21.

\textsuperscript{124} See id. at 60–61, 64 (“[B]y 1907, that number [of schools that had accepted the case method] had risen to over thirty.”).

\textsuperscript{125} See Frederick C. Hicks, Yale Law School, 1869–1894, at 28–35 (1937) (“The traditional scheme of instruction was adhered to. It consisted chiefly of recitations on required reading supplemented by lectures and moot court work.”).

\textsuperscript{126} Stevens, supra note 123, at 62.

\textsuperscript{127} Id. at 63.

\textsuperscript{128} See id. (“The vast success of Langdell’s method enabled the establishment of the large-size class . . . . Any educational program or innovation that allowed one man to teach even more students was not unwelcome to university administrators. The ‘Harvard method of instruction’ meant that law schools could be self-supporting.”).
reasons, the Langdellian model of legal education ultimately took hold and spread through law schools across the nation.\textsuperscript{129}

It is worth noting that the rise of the Langdellian law school coincided with—and was doubtless aided by—the rise of the organized bar. Up through 1870, law practice varied greatly across the country but was largely characterized by practitioners who were not formally organized as a group with clear standards for preparation and admission.\textsuperscript{130} Local bar groups began organizing in cities such as New York and Chicago during the 1870s; in 1878, the American Bar Association (ABA) was founded for a mixture of reasons, including improving legal education and standards for admission to the bar.\textsuperscript{131} As the twentieth century arrived and progressed, the rise of corporations, administrative law, and business and banking regulation certainly made the practice of law more complex, changes that challenged the traditional, informal approach of reading for the bar.\textsuperscript{132} It is in this context that formalized legal education and the ABA’s push for the same must, at least in part, be understood.

\textsuperscript{129} The model of university-based legal education did not catch on, however, in England. The University of London repeatedly tried to wrest the legal education of barristers from the Inns of Court during the late nineteenth and early twentieth centuries, to no avail. See Burrage, supra note 48, at 145 (noting that between 1884 and 1904 the University of London “made four separate proposals to establish a school of law jointly with the Inns” but that “the Inns remained unconvinced and rejected all proposals”). This resistance of the practicing bar to formalized, university-based legal education held up until the mid-twentieth century. Id. at 145–46. There was more success in moving the education of English solicitors into the University in the first half of the twentieth century, but that ended with a severance between the universities and the solicitor’s professional association, the Law Society, in 1952, when all solicitors-in-training were withdrawn from the universities. See id. at 147 (“The universities paid little heed to the [Law Society’s] criticisms, and in 1952 the Law Society therefore carried out its oft-repeated threat and withdrew all of its students from the universities.”).


\textsuperscript{131} Id. at 286–87.

\textsuperscript{132} See id. at 308 (“Before [the 1870s] lawyers could educate one another to a large extent . . . . Before then the intricate web of administrative regulation had not been woven. After the ’70s, the printed sources of the law became a flood . . . . and the reach of government vastly expanded.”).
III. Over a Century of Critique and Reform

“Our law schools, as usually conducted, offer nothing. Most of
them do not, in their plan of study, seem ever to recognize the
need.”

—ABA Committee on Legal Education, 1890

Although the Langdellian model of legal education became
dominant and widely accepted in the early part of the twentieth
century, it was not without its contemporary critics. Indeed, since
the advent of the Langdellian law school, there has been a steady
stream of critiques of legal education, each of which have had
varying degrees of success in encouraging legal education reform.
Sadly, however, when one canvases the various assessments of
formal legal education over the past 130 years, it is remarkable
how consistent the criticisms are and how persistent the
Langdellian model has been in the face of these critiques.

A. Early ABA Reports

An early appraisal of the American law schools, at the time of
Langdell, came from the ABA’s Committee on Legal Education
and Admission to the Bar in 1879:

But it is difficult to deny that there are American [law schools]
not deserving of commendation. Institutions where the course
is unjustifiably limited and circumscribed; where the term of
study is evidently too brief for useful purposes; where students
continue to be invited, when they are unfit by reason of
deficient education and want of contact with liberal studies, to
wrestle with the difficulties of the law; where, in a way
unworthy of the cause of legal learning, a spirit of competition
to attract greater numbers than are to be found in other
establishments, is allowed to obtain control; where
examinations, which are such only in name, take the place of a
searching scrutiny of the student’s acquirements; where there
is no connection with any influence, except that of a faculty
insufficient to meet the demands of a progressive time; where

133. William G. Hammond et al., Report on the Committee of Legal
Education, 13 Am. B. Ass’n Rep. 330 (1890) [hereinafter Hammond et. al., 13
Am. Bar Ass’n Rep.].
there are no exercises sufficiently serious to try and develop the abilities the student may have; and where degrees are thrown away on the undeserving and the ignorant.134

Although this critique was contemporaneous with Langdell’s reforms at Harvard, they cannot truly be said to reflect on Langdell’s Harvard because at that point in time, Langdell’s model had yet to spread to other law schools.135 Thus, the law schools being critiqued were largely those of the kind that preceded Langdell, including the kind that Harvard had been prior to the implementation of his reforms. Indeed, one can view Langdell’s reforms as responsive in substance to many of the charges of the 1879 report: Langdell lengthened the period of study from eighteen months to three years, he implemented an undergraduate education requirement for admitted law students, he strengthened the examinations used to determine students’ qualifications to advance and obtain the degree, and he sought to strengthen the faculty by moving towards full-time academics rather than practitioners who had less time to dedicate to the job of teaching law students.136

In the eyes of the ABA Committee on Legal Education, however, little had changed in law schools generally by 1890, when it identified the defects of law schools as including the “constant temptation to attract more students and to graduate a large class by allowing all to get through who can ‘cram up’ for a not very rigid examination, which tests nothing but a mere recollection of the written letter of the law.”137 The Committee went on to observe that “all [law schools] suffer from the want of a proper standard of true legal education, a definite plan of the

134. Carleton Hunt, Report of the Committee on Legal Education and Admissions to the Bar, 2 AM. B. ASS'N REP. 219 (1879) [hereinafter Hunt, 2 AM. BAR ASS'N REP.]. It should also be noted that the Committee was critical of the failure of law schools to give instruction in Roman Civil Law, the law of nations, admiralty and maritime law, and comparative jurisprudence. Id. at 220–28.

135. See STEVENS, supra note 123, at 60–64 (discussing the growth of Langdell’s method of instruction in the late nineteenth century).

136. See supra Part II.E (providing an overview of Langdell’s reforms to the legal education system of America).

137. Hammond et. al., 13 AM. BAR ASS'N REP., supra note 133, at 329.
entire course.”\textsuperscript{138} In its most damning critique, the ABA committee concluded its observations with the following analysis:

The defects of the present method may be summed up, we think, in one very familiar antithesis: they do not educate, they only instruct. They aim only to heap up in the student’s mind a great mass of legal “points”—rules, definitions, etc.—but they do not fashion these into a system, nor even do they give him the faculty of constructing for himself such a system. The mutual influences of different rules, the construction of legal relations and institutions, the processes by which the law is constantly developing and assuming new phases, are neglected, or, rather, positively ignored. He is supplied with an abundance of crude material, but not taught to use it. In office study, the daily participation in actual business gave the student at least some empiric training. He learned to use his acquisitions as an apprentice learns to use the tools of his trade… The process was a rude and imperfect one, very uncertain in its results and exceedingly wasteful of time and labor, but for two or three centuries it has been the way in which English and American lawyers have been instituted, and it will not, perhaps cannot, be abandoned without something better is [sic] offered in its place. Our law schools, as usually conducted, offer nothing. Most of them do not, in their plan of study, seem ever to recognize the need. It is fortunate for them and for their pupils alike that the training thus omitted may be supplied in the early years of practice, at least to a very considerable extent.\textsuperscript{139}

We see in this critique charges that could be leveled against law school education today. An instruction in legal “rules” dominates; a comprehensive legal education that synthesizes different areas of the law, teaches how to establish and maintain legal relationships or entities, and studies the processes—legislative, administrative, collaborative, and judicial—by which the law and legal relations are shaped is still “neglected” or “ignored” in law schools today. The charge that the student “is supplied with an abundance of crude material, but not taught to use it” is as true today as it was then at those schools that do not require extensive practical skills training or experience before graduation. Indeed, it can still be said of some law faculty that

\textsuperscript{138} \textit{Id.} at 330.
\textsuperscript{139} \textit{Id.}
they do not “seem ever to recognize the need” to offer training that approximates what students miss by not going through an apprenticeship experience. The key difference between now and then is that the saving grace for the Committee—“that the training thus omitted may be supplied in the early years of practice”—no longer accurately characterizes the opportunities facing most law school graduates today, given the increasing unwillingness of legal employers to foot the costs of basic training for new lawyers and the reality that many law graduates do not obtain work with employers who have the time, ability, or resources to support such training.

In the year following this report, 1891, the Committee again issued a report highly critical of the program of legal education at most law schools, this time focusing on the schools' emphasis on rules over principles and on rote memorization rather than reflection and analytical thinking:

The only method of teaching that deserves entire reprobation is that which encourages or even permits the student to make the entire course a mere exercise of memory, without reflection or judgment; not beginning with the fundamental notions or principles by which all his reasoning is to be conducted in actual practice, but laying up rule after rule, decision after decision, as if they were to constitute the fund of knowledge upon which he had only to draw during his after life.

Absurd as this method is to anyone who knows the daily work of an active practitioner, there is a fatal tendency toward it in much of the school routine . . . . No amount of text-book learning, no familiarity even with the cases will avail him, if he cannot reason from one set of facts to another by the use of the exact terms in which the law sums up its principles.\textsuperscript{140}

This would appear to be a more direct attack on the Langdellian case method,\textsuperscript{141} though the larger point of the 1891 and 1890 critiques was the overemphasis of doctrine at the expense of

\textsuperscript{140} Hammond et. al., 14 AM. BAR ASS'N REP., supra note 34, at 332–33.

\textsuperscript{141} Such direct attacks continued in the 1892 report of the Committee, in which it faulted the case method for focusing on disputes rather than settled doctrines, training a student to be a “gladiator” more so than a counselor who can “advise a client when he is safe from litigation.” William G. Hammond et al., Report of the Committee of Legal Education, 15 AM. B. ASS'N REP. 341 (1892).
conveying fundamental and enduring legal principles, an understanding of the larger system formed by the law and the reason underlying it, and the ability to use such knowledge in practice.

Another common thread running through each of these critiques was the insufficiency of the modes of assessment, which simply required the memorization and regurgitation of rules. Not surprisingly, bar examination questions of the early twentieth century—not long after these ABA appraisals of law schools—were similarly narrow in their approach to assessment, posing such questions as “How must a partnership exist?” and “Name the twelve maxims of Equity.” Today, law schools remain focused on doctrinal instruction and summative assessments that test substantive knowledge and analytical abilities rather than the full range of skills needed by the active practitioner. Bar exams similarly remain oriented primarily towards doctrinal knowledge and analytical ability rather than practical competence.

Through the remainder of the nineteenth century and into the early twentieth century, the Committee on Legal Education continued to monitor and recommend improvements to legal education. By 1912, the Committee’s view of law school education had become more favorable, highlighting what it viewed as

142. Wormser, supra note 88, at 34 (providing a critique of bar examination questions of the early twentieth century). Wormser, in his critique, questions:

What kind of nonsense, what order of foolishness is it, that impels us lawyers to agree that the vital necessity for the law-student is his acquisition of the power of logical analysis and thoughtful discrimination in handling legal propositions, and at the same time, leads us to furnish many a bar examination paper calculated to test little more than his memory . . .?

Id.; see also Gordon Hickey, After 100 Years, Bar Examiners Still Protecting the Public, 60 Va. Law. 18, 18 (2012) (reporting questions from the first Virginia bar exam, including “What is a freehold estate?” and “What is the distinction between a vested and a contingent remainder?”).

143. See, e.g., Va. Bd. of B. Examiners, Rules—Section I: Examinations (Aug. 2012), http://www.vbbe.state.va.us/pdf/VBBERules.pdf (noting that each examination consists of two parts, an essay part that tests the applicant’s knowledge in twenty-two substantive law areas, and a multistate exam that tests knowledge in six substantive law areas).
several advances in legal education that had occurred up to that time:

1. The recognition of the superiority of the law school over the office preparation for the Bar.

2. The recognition of a definite period of legal study upon the completion of which, and not before, the applicant can apply for admission to the Bar.

3. The lengthening of the law school course of study to three years.

4. The changed method of law instruction which has substituted in so many of the law schools of the country the study of law through cases, either as an exclusive system, or in combination with the use of text-books, in lieu of the old system of lectures, or of lectures and text-books.

5. The development of a class of law teachers who are withdrawn from law practice, and whose vocation it is to teach law.144

Here we see the lauding of the case method of instruction as well as an approval of a full-time professoriate removed from the practice of law.145 Not wanting to rest on these laurels, however, the Committee, in 1913, asked the Carnegie Foundation for the Advancement of Teaching to undertake an investigation of legal education in the United States, one that would be similar to the Foundation’s *Flexner Report* on medical education146 that devised

---


145. It should be noted that this statement did not reflect an endorsement of a law professoriate wholly lacking in any practice experience. To the contrary, the Committee expressly indicated that “[i]t is desirable that law teachers, while withdrawn from practice, should have had actual experience at the Bar.” *Id.* at 604.

and recommended major changes to medical training that were adopted and characterize medical education today. The Carnegie Foundation obliged, and the products of that request were the 1914 Redlich Report and the 1921 Reed Report.

### B. The Redlich and Reed Reports on Legal Education

For nearly a century, the Carnegie Foundation for the Advancement of Teaching has studied legal education and issued various reports and bulletins on the topic. The first of these, a report on the case method in American law schools, was published in 1914. This topic was chosen because of the controversy surrounding the method; although legal instruction at Harvard Law School had come to be characterized by the case-dialogue method under Langdell, the law professoriate and the practicing bar was not universally convinced of its efficacy. The Carnegie Foundation sought to shed light on this debate by enlisting a law professor from Austria—Josef Redlich—to visit American law schools and study the method from an impartial point of view.

After spending two months visiting ten law schools—four of which did not employ the case method—Redlich generally concluded that the case method was meritorious, with a few caveats. Redlich quibbled with Langdell’s claim that law was
an inductive science for which the study of cases, as original sources, was well suited by pointing out that scientific inquiry is certainly not limited to inductive methods\textsuperscript{153} and that legal reasoning in any given case is actually deductive rather than inductive:

The judge who, in the individual case, decides according to the common law, applies . . . to the state of facts then before him, one of these already existing norms . . . and pronounces . . . only the rule or norm applicable to the specific case. His intellectual activity in this is, therefore, essentially deductive; for by deduction we mean the application of an already existing general rule to the particular case.\textsuperscript{154}

Redlich also distinguished legal science from the physical sciences by noting that the law does not consist of naturally observable facts, but rather of the products of human will directed at ordering and guiding human behavior, making law more of an "intellectual science" or "normative science."\textsuperscript{155} That said, Redlich did find that the case method was valuable in that it exposed students to mastery both of the content of the law and of the analytical methods lawyers must employ to make use of the law as practicing attorneys.\textsuperscript{156}
Redlich’s general endorsement of the method was coupled with a caveat and a critique. The caveat was that the case method was appropriate for the study of Anglo-American law because such law—at that time—was largely unwritten, common law. With common law being case law, studying cases made sense, compared with the task of judges in civil law countries of finding the applicable rules and statutes that govern the facts at hand. Although the common law may have been prevalent during Langdell’s time in the late nineteenth century, in the twenty-first century other sources of law—constitutional, statutory, regulatory, and negotiated—dominate the legal landscape, making it relevant to wonder whether the common law-based case method legitimately retains its purchase.

Redlich’s critique of the method was as follows:

It is characteristic of the case method that where it has thoroughly established itself, legal education has assumed the form of instruction almost exclusively through analysis of separate cases. The result of this is that the students never obtain a general picture of the law as a whole....

His prescription for remedying this defect was to create an introductory law school course:

[I]n American university law schools the students ought to be given an introductory lecture course, which should present, so to speak, ‘Institutes’ of the common law. Every department into which the American law is divided, whether as common

---

157. See id. at 35 (“[T]he fundamental reason for this success [of the case method] is to be found in the present condition of American law, and within this especially in the unshaken authority of the common law.”).

158. See id. at 36–37

To the Englishman and the American... the law appears rather as the single case of law... conducted by the regular judge, and depending only upon his ‘finding of the law.’ The task of the European judge is to find which of the rules and principles of law that are contained in the system of law governs the state of facts in question.

159. See Rubin, supra note 101, at 619 (“To the extent that a national law school wants to teach the consensus view among state jurisdictions, it should be teaching the typical or model provisions of state statutes and regulations, not the common law.”).

160. REDLICH, supra note 95, at 41.
law or equity, employs certain common elementary ideas and fundamental legal concepts, which the student ought to be made to understand before he is introduced into the difficult analysis of cases.\textsuperscript{161}

This is sound advice even today, given the fact that pre-legal education does not offer such an overview.\textsuperscript{162} Pre-legal education is not formally connected with law school training in any way,\textsuperscript{163} meaning that students learn little about the legal profession and law school before deciding to become a lawyer,\textsuperscript{164} and are not guaranteed to have had any training or learning preparatory for the study of law.\textsuperscript{165} Indeed, American law schools typically offer a

\begin{footnotesize}
\begin{enumerate}
\item Id. at 42.
\item See HARNO, supra note 23, at 127 ("Many of the problems of legal education owe their being to deficiencies in the pre-legal period.").
\item See Preparing for Law School, AM. B. ASS'N, http://www.americanbar.org/groups/legal_education/resources/pre_law.html (last visited Nov. 14, 2012) ("The ABA does not recommend any undergraduate majors or group of courses to prepare for a legal education. Students are admitted to law school from almost every academic discipline.") (on file with Washington and Lee Law Review).
\item See MacCrAte REpORt, supra note 22, at 228 ("Prospective law students generally are not knowledgeable about the profession . . . .").
\end{enumerate}
\end{footnotesize}
course of the kind proposed by Redlich to their foreign LL.M. students only, presumably under the assumption that American law students pursuing a J.D. would not benefit from this type of overview.

The Carnegie Foundation followed up the Redlich Report with a more comprehensive report on legal education authored by Alfred Zantzinger Reed and published in 1921. Right from the outset, the tenor of the Reed Report was less laudatory of law schools, as it began by bluntly proclaiming, “Our contemporary American system of legal education . . . is generally recognized to be defective in many respects.” As Reed traced the movement of legal education from the office to the school, he lamented the concomitant loss of practical training:

[W]e are in a fair way of losing entirely the practical training secured under a practitioner, that was once assumed to be the only logical means of preparing students in Anglo-American law. Even its remnants are not usually regarded by the law schools as worth preserving, now that they have virtually preempted the entire field of legal education. Moot courts . . . and “practice courses” are among the devices by which they


167. Reed Report, supra note 1. See also Stevens, supra note 123, at 112–13 (describing the Carnegie Foundation funding both the Reed and Redlich reports). Reed published a further report on legal education funded by the Carnegie Foundation in 1928. See Alfred Zantzinger Reed, Present-Day Law Schools in the United States and Canada, Bulletin No. 21 (1928) [hereinafter Reed Report II].

168. Reed Report, supra note 1, at 3.
conceal from themselves and others the necessarily theoretical character of their instruction.\textsuperscript{169}

This neglect of the practical was, for Reed, unconscionable: “The failure of modern American law school to make any adequate provision in its curriculum for practical training constitutes a remarkable educational anomaly.”\textsuperscript{170}

Reed’s concern over the absence of practical training came in the context of a declining presence of law office apprenticeship training as supplementary to formal academic legal training. As early as the late nineteenth century, the ABA was urging that law school be accepted as a substitute for law office training rather than its mere predicate.\textsuperscript{171} By 1895, law school study and clerkship training had come to be regarded as interchangeable, rather than indispensable, in legal education, though no state bar admissions authorities were requiring law school at that time.\textsuperscript{172} And by Reed’s time, four states were requiring formal legal education at law school, and in almost all of the remaining jurisdictions, “law school and law office training had become alternatives.”\textsuperscript{173} In a context in which apprenticeship training had become optional, it is little wonder that Reed found fault with law school’s failure to attend to practical instruction.

Another major strain of Reed’s critique was of what he referred to as “the theory of a unitary bar, whose attainments are to be tested by uniform examinations.”\textsuperscript{174} This theory held that rather than having different kinds of lawyers, who were differently qualified to handle different matters, our political

\textsuperscript{169}. \textit{Id.} at 48.  
\textsuperscript{170}. \textit{Id.} at 281.  
\textsuperscript{171}. \textit{See Transactions of the Forty-Second Annual Meeting of the American Bar Association,} 42 AM. BAR ASS’N REP. 28–30 (1919) (adopting a resolution that “time spent in any chartered and properly conducted law school ought to be counted in any state as equivalent to the same time in an attorney’s office in such state . . . .”).  
\textsuperscript{172}. \textit{See STEVENS, supra note 123,} at 95 (“T]ime spent in law school came to be counted as time spent in a law office, as states gradually increased (or restored) the requirement for some kind of apprenticeship or clerkship. Yet no state required attendance at law school . . . .”).  
\textsuperscript{173}. \textit{See id.} at 174 (noting that by 1930 only West Virginia, Colorado, Kentucky, and Wyoming required at least one or more years of legal education).  
\textsuperscript{174}. \textit{REED REPORT, supra note 1,} at 57.
philosophy abjured distinctions and ranks among lawyers, resulting in the “American ‘attorney and counselor-at-law,’ privileged to practice all branches of his profession in all courts equally.” The unitary bar theory, in Reed's view, obscured the real distinctions between the law schools, mainly the divide between the full-time national law schools and the local law or part-time law schools, which “differ[ed] vastly from one another in type of law studied, in methods of instruction, and in the amount of actual work represented by their degree.” The problem here was that the unitary ideal forced each type of law school to compete with the other, to the detriment of each: “Coming, however, into direct competition with one another under the accepted dogma of a unitary bar, each affects injuriously the other's development.” Further, the notion of a unitary bar was at odds with the needs of the increasingly pluralistic society that was 1920s America.

In Reed's estimation, the three types of emerging law schools—full-time national, full-time local, and part-time evening—represented different approaches to legal education that should be encouraged and strengthened rather than homogenized or suppressed:

Each has seen clearly that if all American lawyers were educated in accordance with the other's plans, we should be in a bad way . . . . If [some] thought . . . had been expended upon

175. See id. at 39–40 (describing the transformation of the legal profession as moving “as a whole into an officially undifferentiated and, as it were, flattened-out profession”).

176. Id. at 63.

177. Id. at 64. Reed posited that “the night schools are damaged by the obligation placed upon them to cover the same curriculum as the day schools.” Id. at 57. The full-time schools “are so fearful of losing students to [the night schools], that they hesitate to raise their own entrance requirements to the level that they really believe in.” Id. at 58. Bar examiners, eager not to favor graduates of one type of school over the other, “are obliged to limit the scope of their enquiries to those elements that any law school, offering the standardized curriculum, must provide—something that even the more poorly trained of the night school students can pass, at least with the aid of special coaching . . . .” Id. at 59.

178. See STEVENS, supra note 123, at 113 (describing Reed’s view of a pluralistic 1920s America as “standing out sharply against the formal theory of a unified bar”).
the problem of dividing the bar along lines that can be justified... we should hardly have been in so bad a situation as we are now, when the schools are... wrangling for possession of a field too large for any of them to cover in its entirety. The scholarly law school dean properly seeks to build up a “nursery for judges” that will make American law what American law ought to be. The practitioner bar examiner, with his satellite schools, properly seeks to prepare students for the immediate practice of the law as it is. The night school authorities, finally, see most clearly that the interests not only of the individual but of the community demand that participation in the making and administration of the law shall be kept accessible to Lincoln’s plain people. All these are worthy ideals. Taken together, they roughly embrace the service that the public expects from its law schools as a whole. But no single institution, pursuing its special aim, can attain both the others as well.179

By permitting this differentiation to thrive, Reed estimated that schools would be able to focus on what they each did best, producing graduates who were well-suited for the kind of lawyering that was the focus of their respective schools.180

The ABA Section of Legal Education and Admissions to the Bar, which in 1919 superseded the abolished Committee on Legal Education and Admissions to the Bar,181 received an early draft of the Reed Report in 1920 and did not like what it saw.182 It promptly formed a special committee chaired by Elihu Root, which met in March and May of 1921 and was tasked with developing its own recommendations for improving legal education.183 Regarding Reed’s suggestion that the diversity in

179. Reed Report, supra note 1, at 417–18.
180. See id. at 417–19 (“[T]he independent development of each type of law school will naturally result in a considerable variation in the kind of professional work for which its products are especially fitted.”).
182. See Stevens, supra note 123, at 115–16 (“Reed’s suggestion in his first report that the most democratic, egalitarian, and American solution would be a differentiated bar, with differing types of law schools... managed to offend both his ABA and AALS audiences.”).
183. See Report of the Special Committee to the Section of Legal Education and Admissions to the Bar of the American Bar Association, 44 Am. B. Ass’n Rep. 679–88 (1921) [hereinafter Special Committee Report, 44 Am. Bar Ass’n Rep.]
types of law schools should be encouraged, the Root Report stated as follows:

Turning first, then, to a consideration of what a lawyer’s training should be, we meet the suggestion that there must be different kinds of training to produce different kinds of lawyers.

With this position we do not agree. In spite of the diversity of human relations with respect to which the work of lawyers is done, the intellectual requisites are in all cases substantially the same.\footnote{Id. at 681.}

The Root committee’s view was that differentiating among law schools meant tolerating law schools of lesser quality: “Nor can there be tolerated a recognized distinction between good and poor legal education. There should be no distinction . . . . [W]e cannot favor the continuation of a class of incompetent practitioners.”\footnote{Id. at 681–82.} Uniformity was to be preferred, which the special committee went about pursuing by recommending that all bar applicants be required to have graduated from a law school complying with prescribed standards, that they be required to have completed at least two years of study in college, and that the course of full-time study of law be three years.\footnote{See id. at 683–85 (spelling out the special committee’s recommendations for uniformity).}

The Root committee acknowledged the need for afternoon and evening law schools to accommodate the student who had to work to support himself and his family; however, it rejected the idea that such schools should be operated with what it viewed as lower educational standards.\footnote{Id. at 685.} Specifically, the special committee noted:

But in recognizing the necessity for afternoon and evening schools we do not recognize the propriety of permitting such schools to operate with low educational standards . . . . On the contrary, the democratic necessity for afternoon and evening schools compels a lifting of these schools to the highest standards which they can be expected to reach.

\textit{Id.}
recommended that “a law school shall not be deemed competent to educate students for the bar unless it provides an adequate library available for their use” and that a school “have among its teachers a sufficient number giving their entire time to the school.” These requirements, of course, would prove to be more burdensome to night schools and local, practice-connected schools that might lack sufficient funds to maintain a complete law library and whose instructors were principally practitioners who devoted most of their time to law practice during the day. These requirements all meant to homogenize law schools around standards that tilted in favor of the national, full-time law schools, to the detriment of night schools and other alternative types of law schools that might have otherwise been able to develop, the latter being schools that non-elites and working class individuals were more likely to be able to attend.

One can hardly quarrel with the notion that there should not be law schools that expressly aspire to lower educational standards or knowingly produce a class of incompetent practitioners. But that is a straw man and as such was not the proposition furthered by Reed in his report. Rather, the idea was that the variation among types of law schools should be recognized in a way that permits each to focus on their respective strengths in service of various segments of the public and bar. Reed did not know what such a regime would look like in its

188. Id. at 685–86.
189. The Association of American Law Schools (AALS), whose membership consisted of the elite law schools, similarly opposed the idea of a differentiated bar or the perpetuation of a separate, lower class of law schools. See Stevens, supra note 123, at 116 (describing the membership of the AALS and referring to the 1921 presidential address by Yale’s Arthur Corbin to the AALS that “reject[ed] any concept of a differentiated bar”).
190. See Reed Report, supra note 1, at 55–56 (describing the development of different institutions to serve those who can devote all their time to legal education while addressing “the democratic desire to extend the privileges of education to the many”).
191. See id. at 56–57 (“If law school graduates enjoyed different privileges in the practice of law, corresponding to the differences in educational effort between full-time and part-time work, the two types instead of rivaling would supplement one another. Each could develop independently along the lines indicated by its own special aim.”).
particulars, but it is likely that he would endorse moves from some quarters today to get the ABA to amend its approval standards to allow creative flexibility among the law schools; that is, efforts such as those that relax the tenure requirement in favor of alternative means of giving faculty security of position might permit some schools, for example, to build a faculty of practitioners who practice during the day and teach only at night. Relaxing other standards—such as the full-time faculty and library requirements or the limitations on distance

192. See id. at 419

The more scholarly institutions may in time be glad to lighten their own burden by throwing upon schools of other types responsibility for certain portions of this broad field. Conveyancing, probate practice, criminal law and trial work are examples of topics that seem particularly appropriate for the relatively superficial schools. All this is mere guesswork, however. It is not even certain that a rigorous functional division of the bar will ever develop. The dividing line between the different types of lawyers may be determined by the economic status of the client rather than by the nature of the professional service rendered. The general principle of a differentiated profession is something that we already have, and could not abolish if we would. The particular principle of a functionally divided bar is something that we may or may not be able eventually to introduce, as one means for making the general principle work better.


education—might similarly free law schools to deliver an education sufficient to serve a different segment of the bar than might a graduate of a national full-time law school.

The challenge would be on the licensing side; state licensing authorities would be responsible for determining whether graduates of all law schools were prepared to practice, something most bar exams, as structured at the time, largely failed to do. Beyond making bar exams a better judge of practice-readiness and legal competence, one could imagine a differentiated bar with various levels of licensure. Such a system would not have to mimic the barrister–solicitor distinction of England precisely, but that model at least could be an example of where to start the thought process. Alternatively, phased licensure could be

195. Current ABA Standards limit the credit that students may receive through distance education to four hours in any term and twelve hours overall. 2012–2013, ABA STANDARDS AND RULES OF PROCEDURE FOR APPROVAL OF LAW SCHOOLS, supra note 28, Standard 306(d). The Standards Review Committee considered revising this standard, but only to permit fifteen hours total of distance education to count towards the J.D. degree. See AM. B. ASS’N SECTION OF LEGAL EDUCATION AND ADMISSIONS TO THE BAR, Standard 311(f), (Proposed Draft 2011) [hereinafter 2011 Proposed Draft], available at http://www.americanbar.org/content/dam/aba/migrated/2011_build/legal_education/committees/standards_review_documents/jan2012/20111222_standards_chapters_1_to_7_post_nov11.authcheckdam.pdf. Currently, there is one completely online law school, The Concord School of Law, which is ineligible for ABA approval due to Standard 306. However, one of its graduates did successfully petition the Massachusetts Supreme Judicial Court for permission to sit for that state’s bar exam, which he passed. See Kristina Horton Flaherty, Court Win for Online Law School Grad, CAL. B. J. (Jan. 2009), http://archive.calbar.ca.gov/Archive.aspx?articleId=94802&categoryId=94651&month=1&year=2009 (last visited Aug. 31, 2012) (on file with the Washington and Lee Law Review).

196. See REED REPORT, supra note 1, at 59 (“The actual situation is that neither the tests of the state nor those of the law schools serve to prevent incompetents from flooding the profession.”).

197. See, e.g., ALL-ABA CONTINUING PROF. EDUC. & ASS’N OF CONTINUING LEGAL EDUC., EQUIPPING OUR LAWYERS: THE FINAL REPORT OF THE CRITICAL ISSUES SUMMIT 8 (2010), available at http://www.equippingourlawyers.org/documents/final_report.pdf (“Regulatory authorities should consider restructuring one-time bar examinations into phased examinations over time, linked in part to attainment of legal practice skills, with some parts of the examination occurring as early as in the law school years.”).

considered, with probationary or apprenticeship periods preceding full admission, as is currently the approach in Canada. Whatever the specifics might be, revising the ABA standards in a way that permits flexibility while ensuring that basic levels of quality are maintained—in tandem with improved state bar admissions standards—seems to be the ideal that Reed was endorsing and would seem to be a way to free law schools to pursue their respective strengths without having to devote resources toward compliance with standards that do not truly speak to their core mission and the segment of the bar they serve.

But this was not to be; the Section rebuffed the Reed Report’s recommendation to permit the development of fundamentally different types of law schools in favor of a unitary set of standards that required all law schools to aspire toward and satisfy its vision of what was required to make a law school educationally sound. There was notable opposition to the

---


200. See Special Committee Report, 44 AM. Bar Ass’n Rep., supra note 183, at 683. The committee noted that:

To require a law-school education without setting standards for the law schools would be worse than useless. Any man or group of men professing to teach law to students may be said to be conducting a law school, no matter how inferior may be the abilities of the students, the nature of the instruction or the physical facilities of the plant . . . . We, therefore, propose not only that a law school education shall be required, but that such education must be obtained in a school complying with certain prescribed standards.

Id.; see also Proceedings of the Section of Legal Education and Admissions to the Bar, 46 AM. B. Ass’n Rep. 678 (1923). The Council came to the following conclusion:

[T]hat a school, to be in the approved list, must comply as to all of its students. Some schools or some institutions conduct classes both for full time and part time students. Some of these have arranged to comply with the standards of the American Bar Association or announced that they will in the near future comply with such standards as to their full time students, but have not arranged to so comply as to their part time students. The Council has concluded that such schools cannot be placed on the approved list.

Id.
standards proposed by the Section; Dean Edward T. Lee of the John Marshall Law School of Chicago argued:

The effect of the adoption of the recommendations would be to place the control of legal education through the country in the hands of the deans of a few large day law schools who have the fate of law teachers in their hands. It would close the profession of the law to all save the leisure class of youth with means sufficient to obtain college and law school training . . . .201

Nonetheless, the ABA approved the Section’s recommendations at its 1921 meeting202 and they were subsequently approved by the Conference of Bar Association Delegates held in Washington, D.C. in 1922.203

C. Twentieth Century Critiques and Reforms

“[I]s it not plain that, without giving up entirely the case-book system or the growing and valuable alliance with the so-called social sciences, the law schools should once more get in intimate contact with what clients need and with what courts and lawyers actually do?”

—Jerome Frank, 1933204

The ABA’s determination to impose its unitary vision of law schools based on the Langdell model—three years of full-time study (or its equivalent if part-time), centered around large physical libraries, and classes taught by full-time faculty members removed from the practice of law—was a victory for the full-time national law schools that furthered the Harvardization of legal education.205 It was not long, however, before another
major threat to the Langdellian model for the study of law in the first half of the twentieth century reared its head. This challenge came from what came to be referred to as the legal Realists, who challenged “the Langdellian notion of law as an exact science, based on the objectivity of black-letter rules.”206 Instead, the Realists posited that the operation of the law in the real world was less coherent and objective than Langdell’s scientific vision of the law suggested.207 With law as a value-laden process more so than a neutral and consistent body of enduring principles, Langdell’s belief in cases as the vehicle for learning legal doctrine and in the mastery of doctrine and principles as the route to becoming a good lawyer became subject to doubt.208

In 1933, Jerome Frank reflected the Realists’ view in his withering critique of the Langdellian case method:

Ostensibly, the students were to study cases. But they did not and they do not study cases. They do not even study the printed records of cases (although that would be little enough), let alone cases as living processes. Their attention is restricted to judicial opinions. But an opinion is not a decision. A decision is a specific judgment, or order or decree entered after a trial of a specific lawsuit between specific litigants. There are a multitude of factors which [sic] induce a jury to return a verdict, or a judge to enter a decree. Of those numerous factors, but few are set forth in judicial opinions. And those factors, not expressed in the opinions, frequently are the most important in the real causal explanation of the decisions.209

Harvard . . . has been followed by all the schools. Their main activities are devoted to instruction in the relatively narrow . . . field of the judge-made technical law.”); see also id. at 411–12 (describing innovations in legal education emanating from Harvard that took hold, including the focus on “national” law, introduction of the case method, and imposition of the requirement of preliminary college training before entering law school). For a fuller account of the standardization movement by the ABA and the AALS, see STEVENS, supra note 123, at 205–13.

206. STEVENS, supra note 123, at 156.

207. See id. (“The Realists went a long way toward killing the idea of ‘the system’ altogether. All legal logic came under suspicion. American law became increasingly purposive, increasingly secularized, and increasingly atomized.”).

208. See id. at 156–57 (“This change inevitably caused the predictive value of doctrine to be seriously questioned.”).

From the Realist perspective, the law in books—particularly books limited to reports of appellate opinions—was not the sum total of what constituted the law nor even the best source for understanding law or legal practice. Frank argued for the study of complete case files—not simply appellate opinions—and for study of the work of lawyers in practice and the work of the courts:

[T]he study of cases . . . should be based to a very marked extent on reading and analysis of complete records of cases—beginning with the filing of the first papers, through the trial in the trial court and to and through the upper courts. Six months properly spent on one or two elaborate court records, including the briefs (and supplemented by reading of text-books as well as upper court opinions) will teach a student more than two years spent on going through twenty of the case-books now in use.

Karl Llewellyn, during his time at Columbia Law School, raised similar points, suggesting the case approach used in business schools as a way of improving on the Langdellian case method.

Another Realist critique was of the abstract focus and dearth of practical training that characterized law schools. Fred Rodell

---

210. See Jerome Frank, What Constitutes a Good Legal Education?, 19 A.B.A. J. 723, 726 (1933) [hereinafter Frank, What Constitutes a Good Legal Education?] (noting the advantages of studying case files over court opinions). Frank noted:

The law student should learn, while in school, the art of legal practice. And to that end, the law schools should boldly, not slyly and evasively, repudiate the false dogmas of Langdell . . . . They must decide not to exclude, as did Langdell, but to include the “methods of learning law by work in the lawyer’s office and attendance upon the proceedings of courts of justice . . . .” They must repudiate the absurd notions that the heart of a law school is its library . . . .

Id. (citations omitted)

211. Frank, Why Not a Clinical Lawyer-School?, supra note 204, at 916.

212. See Karl N. Llewellyn, The Current Crisis in Legal Education, 1 J. LEGAL EDUC. 211, 215 (1948–1949) (“Consider, for example, the possibility of building up our so-called cases out beyond the judicial opinion into something resembling the completeness of the cases gathered for the Harvard Business School.”). For an exposure to the Harvard Business School case method, see The Case Method at HBS, HARVARD BUS. SCH., http://www.hbs.edu/teaching/inside-hbs/ (last visited Sept. 1, 2012) (on file with the Washington and Lee Law Review).
of Yale found fault with the focus on abstractions rather than on practical legal problem-solving when he wrote, “[T]hus, law school courses, since they are cut out of the pseudo-science of Law, inevitably focus on generalities and abstractions rather than on the solution of specific problems.” In 1928, Alfred Reed of the Carnegie Foundation commented on the need for more practical instruction in his second report on legal education in which he faulted the law schools for their failure to provide students with practical experience:

> [A]ll the activities of the school are necessarily conducted under artificial conditions. The hurly-burly of actual practice is systemized for the student in a manner that it never is when an authentic client begins to tell him his woes. This systemization is indispensable as a means of ensuring him adequate resources of information and of developed reasoning power upon which to draw. But there still remains the task of marshaling these resources, of focusing his scattered acquisitions upon the unlabeled and multifarious collection of facts and aims with which he will be called upon to deal. Skilled ability to do this is what distinguishes the lawyer from the legal scholar, the expert practitioner from the man who knows, or who is qualified to ascertain, the law.

Reed favored further experimentation with legal clinics associated with law schools as a way of ameliorating this condition. Jerome Frank suggested the same only a few years later.

These points were well taken; however, they did not lead to a wholesale overhaul of legal education away from the Langdellian model. Instead, law schools responded by supplementing their curricula with practice courses in the form of legal clinics. By

---

215. See id. at 216–21 (“If by this method lawyers who enter the legal profession through the law schools can acquire even a small amount of that practical experience which they now so sadly lack, the plan is certainly worth putting into operation.”).
216. See Frank, Why Not a Clinical Lawyer-School?, supra note 204, at 917 (“Suppose, however, that there were in each law school a legal clinic or dispensary.”).
217. See Stevens, supra note 123, at 214–16 (“Of all aspects of the renewed interest in skills, the particular interest in the skills embraced in the concept of
1951, there were twenty-eight clinics maintained by law schools, independent legal aid societies, or public defender offices.218 The movement grew stronger over the next two decades, resulting in clinical programs at nearly half of the law schools in the country by the early 1970s.219

Although the rise of clinical legal education was a major innovation in how law schools prepared their students for practice, very little change in the law school program had occurred overall.220 Indeed, criticism of the legal profession persisted in the 1970s, a critique that extended to law schools as the near exclusive route to practice.221 As the second and third years of law school had become largely elective and focused on courses not necessarily oriented towards practice, bar admissions authorities began considering imposing practice course requirements as prerequisites to admission.222 In 1971, the AALS issued the report of its Curriculum Study Project—known as the Carrington Report—which called for a curriculum that emphasized practice-oriented courses in the first year.223 A 1972
report sponsored by the Carnegie Commission on Higher Education noted the encouraging expansion in clinical legal education but cautioned that proper supervision and instruction were required to make real life experiences edifying and educationally sound, and doubted whether clinical education ultimately could be “the solution” to the host of ills that plagued legal education given its costs.224

In 1979, the ABA entered the dialogue by publishing the report of its Task Force on Lawyer Competency entitled The Role of Law Schools, in which it observed, “Law schools, have not . . . undertaken to provide such comprehensive training that individuals emerge upon graduation as fully competent ready-to-practice lawyers.”225 The Task Force then emphasized the need for improvement in the following areas:

(a) developing some of the fundamental skills underemphasized by traditional legal education; (b) shaping attitudes, values, and work habits critical to the individual’s ability to translate knowledge and relevant skills into adequate professional performance; and (c) providing integrated learning experiences focused on particular fields of lawyer practice, including but not limited to trial practice.226

Toward that end, the Task Force recommended that law schools provide instruction in fundamental skills critical to lawyer competence: effective writing, oral communication skills, fact-gathering, interviewing, counseling, and negotiation skills.227 It went on to recommend teaching these fundamental lawyer skills in small classes, that schedules be adjusted to “provide [the] opportunity for periods of intensive instruction in fundamental

Professions of the Law: 1971, Part One, Section II (1971), reprinted in Herbert L. Packer & Thomas Ehrlich, New Directions in Legal Education: A Report Prepared for the Carnegie Commission on Higher Education 93, 110–17 (1972) [hereinafter Carrington Report] (including, among others, a course entitled legal advocacy to be completed in the first year of law school where “[e]mphasis is placed on the need to develop and use professional skills”).

224. See Packer & Ehrlich, supra note 223, at 42–43, 46 (“We prefer to think that the path of improvement lies in experimentation with many modest ideas, one of which is clinical education.”).


226. Id. at 14.

227. Id. at 3.
lawyer skills," that law schools encourage more cooperative law student work, and that “law schools should make more extensive instructional use of experienced and able lawyers and judges.”\textsuperscript{228} Also noteworthy was its recommendation that law schools “develop and use more comprehensive methods of measuring law student performance than the typical end-of-the-term examination” and that “[s]tudents should be given detailed critiques of their performance.”\textsuperscript{229} Finally, with respect to the academic program, the Task Force urged law schools to “seek to achieve greater coherence in their curriculum,” noting that “the three-year program should build in a structured way.”\textsuperscript{230}

Notwithstanding these recommendations, when a proposal to require that law schools offer practical skills instruction as a condition of accreditation was presented to the ABA by the Council of the Section on Legal Education, the ABA softened the proposal by limiting it to an interpretative paragraph calling for skills training rather than an approval standard mandating such training.\textsuperscript{231}

More than a decade later, legal education still was perceived as inadequately preparing students for practice, as little had changed in how law schools educated their students.\textsuperscript{232} The ABA spoke to this concern in a 1992 report known as the \textit{MacCrate Report}.\textsuperscript{233} The \textit{MacCrate Report} outlined a series of skills and values that were deemed to be essential for competent representation, and then recommended that law schools offer courses that effectively instruct students in those areas.\textsuperscript{234} The fundamental lawyering skills outlined included problem solving,
legal analysis and reasoning, legal research, factual investigation, communication, counseling, negotiation, knowledge of litigation and ADR procedures, law-practice management, and the ability to recognize and resolve ethical dilemmas. The four values suggested by the report were the provision of competent representation, striving to promote justice, fairness, and morality, striving to improve the profession, and professional self-development. The report also explored the state of skills instruction offered in law schools at the time, finding that a majority of law graduates had only one or no professional skills courses beyond basic legal research and writing courses, trial advocacy, or moot court experiences. The report found that clinical programs were generally available to only thirty percent of law students at schools where such courses were offered, with skills training courses occupying only nine percent of the total instructional time at law schools.

An important perspective offered by the MacCrate Report was the importance of the entire legal educational continuum—pre-law, law school, bar admissions, and continuing legal education—and it addressed the need for improvements to occur at each stage. Ultimately, the report concluded that greater attention should be paid to “the linkage between the several phases of lawyers’ education” if the gaps in professional legal education are to be filled. Although the report spawned some efforts to increase skills instruction in law schools, such training has remained peripheral to legal education and no grand effort

235. Id. at 138–40.
236. Id. at 140–41.
237. Id. at 240.
238. Id. at 240–41.
239. See id. at 3–8 (“The skills and values of the competent lawyer are developed along a continuum that starts before law school, reaches its most formative and intensive stage during the law school experience, and continues throughout a lawyer’s professional career.”).
240. Id. at 320–21. The report proposed the creation of a national institute to attend to this coordination. Id.
241. See Carnegie Report, supra note 22, at 94 (“[S]kills training will continue to face an uphill battle unless it is linked with an accepted theory of lawyering that could provide a bridge between theory and practice and perhaps establish a rationale for more systemic continuing education beyond law.
to coordinate the various phases of legal education has materialized.

D. Contemporary Critiques

In 2007, a group of researchers compiled a report entitled Best Practices for Legal Education (Best Practices) that began by echoing the MacCrate Report’s charge that multiple stages along the legal education continuum had substantial room for improvement. Specifically, the report noted that law graduates “are not sufficiently competent to provide legal services to clients or even to perform the work expected of them in large firms” and that survey after survey has reported a decline in lawyer professionalism. However, the true focus of Best Practices was on how law schools could better design and deliver a curriculum that prepared students for practice. As the MacCrate Report had done a decade before, Best Practices spoke of the need for skills and values training by calling for “context-based education” as the means for giving students the problem-solving opportunities they need to become competent and effective legal practitioners. The report also dug into what is wrong with traditional law school teaching and assessment methods, proposing the use of a diversity of approaches that are connected with developing competent and thoughtful practitioners.

242. See STUCKEY ET AL., supra note 30, at 11–36 (noting a compelling need to improve legal education because, among other reasons: the licensing process is not protecting the public, and because law schools are not fully committed to preparing their students for bar exams or practice).

243. Id.

244. See id. at 8–9 (“Law schools should organize their curriculums to develop knowledge, skills, and values progressively; integrate the teaching of theory, doctrine, and practice; and teach professionalism pervasively throughout all three years of law school.”).

245. See id. at 141 (“Legal education would be more effective if law teachers used context-based education throughout the curriculum.”).

246. See id. at 207–63 (noting that “many teachers use the case method exclusively even when other methods of instruction would accomplish their educational objectives more effectively,” and that “the end-of-the-semester essay exam is an inadequate method for assessing student learning . . . [that] is
Overall, the report’s core concern was that law schools should recognize the broad array of skills and attributes that successful lawyers need to have beyond legal knowledge and analytical ability and incorporate courses and methods that attend to those competencies.247

Following closely on the heels of the Best Practices report in 2007 was another major critique of legal education, a report of the Carnegie Foundation entitled Educating Lawyers: Preparation for the Profession of Law (Carnegie Report).248 The report was the product of a two-year study that included an extensive literature review, consultation with the AALS and the Law School Admissions Council, and site visits to a representative sample of sixteen American and Canadian law schools, including public and private schools, more selective and less selective schools, freestanding schools, and schools that served certain ethnic minority groups.249 A central premise of the Carnegie Report is that the education of professionals is a complex and unique enterprise that cannot simply focus on the transmission of expert knowledge, but must also focus on instilling the specialized skills, standards, judgment, and values that define practice in a profession.250 The challenge for professional education is to weave the various components of professional training into a whole that attends to the interests of educators, practitioners, and the public to be served by the profession.251

According to the Carnegie Report, this has been the singular challenge of legal education, as its modern linkage with the research university has stood in tension with the historical

247. See id. at 7–10 (“While law schools help students acquire some of the essential skills and knowledge required for law practice, most law schools are not committed to preparing students for practice . . . . [M]ost law school graduates are not as prepared for law practice as they could be and should be.”).

248. CARNEGIE REPORT, supra note 22.

249. Id. at 15–16.

250. See id. at 3–4 (noting that professional schools “are perhaps the sole site where the professions' standards of good work set the agenda for learning”).

251. See id. at 4 (“This is a complex educational process . . . and its value depends, in large part, on how well the several aspects of professional training are understood and woven into a whole.”).
connection between the community of practitioners and young legal apprentices. The university connection has made legal education predominately academic in character, focusing on teaching and studying law as a science, “put[ting] a premium on formal knowledge, abstracted from context.” This positivist orientation—which, as we have seen, characterized the views of Langdell—led to a view that the traditions of practice—such as craft, judgment, and professional responsibility—were too subjective and uncritical, a perspective that “undermined the academic legitimacy of practical knowledge” and “spelled the eclipse of traditional forms of practitioner-directed apprenticeship by academic instruction given by scholar-teachers.”

The triumph of the academy in legal education has resulted in a focus on training in legal knowledge and legal analysis rather than on learning to practice. Complete professional education, however, must be defined by “three apprenticeships” that the Carnegie Report authors describe: the cognitive apprenticeship focused on expert knowledge and modes of thinking; the apprenticeship of practice, which trains students in “the forms of expert practice shared by competent practitioners;” and the apprenticeship of identity and purpose, which “introduces students to the purposes and attitudes that are guided by the values for which the professional community is responsible.” To be successful in their goal of educating lawyers, law schools must “initiate learners into all three apprenticeships.” Unfortunately, law schools focus heavily on the cognitive

252. See id. (“[L]aw schools are hybrid institutions. One parent is the historic community of practitioners, deeply immersed in the common law and carrying on traditions of craft, judgment, and public responsibility. The other heritage is that of the modern research university.”).
253. Id. at 4–5.
254. See supra Part II.E (summarizing Langdell’s theory of American legal education).
255. CARNEGIE REPORT, supra note 22, at 4–5.
256. See id. at 6–7 (“In its quest for academic respectability, legal education would come to emphasize legal knowledge and reasoning at the expense of attention to practice skills, while the relations of legal activity to morality and public responsibility received even less direct attention in the curriculum.”).
257. Id. at 28.
258. Id.
apprenticeship: “For many students, neither practical skills nor reflection on professional responsibility figure significantly in their legal education. The academic setting clearly tilts the balance toward the cognitive and intellectual.”

The Carnegie Report contrasts this tilt in legal education with the laboratory experience that characterizes engineering education and the extensive clinical work that is the hallmark of medical education.

The academic focus of professional legal education has resulted in a shift in teaching methods, away from “apprenticeship, with its intimate pedagogy of modeling and coaching, toward reliance on the methods of academic instruction, with its emphasis on classroom teaching and learning.”

Further, the Carnegie Report notes that the “signature pedagogy” of law school—the Socratic case-dialogue method—leads to two major weaknesses of typical law school education: “[T]he casual attention that most law schools give to teaching students how to use legal thinking in the complexity of actual law practice” and “law schools’ failure to complement the focus on skill in legal analysis with effective support for developing the ethical and social dimensions of the profession.”

These weaknesses “prolong and reinforce the habits of thinking like a student rather than an apprentice practitioner” and leave students with an

---

259. Id. at 79.
260. Although engineering education does include lab work, the Carnegie Foundation—in a separate report—has critiqued engineering programs for how labs are poorly designed and integrated into a curriculum that remains dominated by formal knowledge and analysis. Sheri D. Sheppard et al., Educating Engineers: Designing for the Future of the Field 16, 74 (2009).
261. See Carnegie Report, supra note 22, at 79–81 (“The distinguishing feature of medical training . . . is that most of it is carried out in settings of actual patient care. The consequence is to provide medicine a real advantage, compared to engineering or law, for integrating its forms of apprenticeship.”). Clinical clerkships in medical education are a legacy of the Flexner Report, a product of the Carnegie Foundation. See Cooke et al., supra note 147, at 13, 82 (“The third year [of medical school], dedicated to patient care and investigation of clinical problems presented by hospitalized patients, is a legacy of the Flexnerian reforms . . . .”).
263. Id. at 188.
264. Id.
underdeveloped sense of professional judgment and responsibility.265

Although there have been successful efforts in recent decades to expand practice-oriented education and apprenticeship learning methods in law schools, these experiences tend to be disconnected from the main curriculum, meaning that the three critical elements of legal professionalism—“conceptual knowledge, skill, and moral discernment”266—remain artificially distinct and, thus, are insufficiently transmitted to novice professionals:

[T]he threefold movement between law as doctrine and precedent (the focus of the case-dialogue classroom) to attention to performance skills (the aim of the apprenticeship of practice) and then to responsible engagement with solving clients’ legal problems—a back-and-forth cycle of action and reflection—also characterizes most legal practice. The separation of these phases into distinct areas of the curriculum, or as separate apprenticeships, is always an artificial “decomposition” of practice. The pedagogical cycle is not completed unless these segregated domains are reconnected.267

Ultimately, it is these deficits—insufficient education in the skills and values of professional legal practice, an overuse of the academic case-dialogue method of instruction, and the poor integration of practical and professionalism instruction into the dominant cognitive curriculum—that are contemporary legal education’s greatest weaknesses.

Recent studies and practitioner surveys reflect the sense that the Langdellian law school is insufficiently connected with the realities of modern practice. The NALP’s After the JD: First Results of a National Study of Legal Careers Survey reports only

265. See id. ([L]ack of attention to practice and the weakness of concern with professional responsibility—are the unintended consequences of reliance on a single, heavily academic pedagogy to provide the crucial initiation into legal education.); see also id. at 140 (“Insofar as law schools choose not to place ethical-social values within the inner circle . . . legal education may inadvertently contribute to the demoralization of the legal profession and its loss of a moral compass . . . .”).

266. Id. at 12.

267. Id. at 124.
moderate satisfaction with the level of practice preparedness that law graduates received in law school:

When [survey respondents were] asked to reflect on their legal education, however, most were not especially enthusiastic about the specific role of their law schools in the transition to practice. On the question of whether law school prepared them well for their legal careers, the median response is exactly in the middle (neither agree nor disagree). Respondents tended to agree—but not strongly—with the proposition that law school teaching is too theoretical and unconcerned with real-life practice. They also evinced a desire for more practical training in their assessment of the most helpful law school courses. Both clinical and legal writing courses received higher ratings than more conventional law school offerings. Most helpful in the transition to practice, however, was experience working during law school summers and during the year.268

Respondents to the Law School Survey of Student Engagement (LSSSE Survey) corroborate these findings, showing that “[f]orty percent of law students felt that their legal education had so far contributed only some or very little to their acquisition of job- or work-related knowledge and skills.”269

Respondents to the LSSSE Survey also indicated that the Carnegie Report’s assessment regarding the insufficient professionalism and ethical training in law schools has some validity: “[O]nly half of students reported that law school prepared them well . . . to deal with ethical dilemmas that may arise as part of law practice.”270 Further, survey results showed that clinical participation and pro bono work was correlated with a higher degree of preparation in the areas of “understanding the needs of future clients, working cooperatively with colleagues as part of a legal team, serving the public good through their profession, and understanding professional values that will serve them in their legal careers.”271 With only a third of law students

268. NALP, AFTER THE JD, supra note 5, at 79.
270. LAW SCHOOL SURVEY OF STUDENT ENGAGEMENT, 2010 ANNUAL SURVEY RESULTS, STUDENT ENGAGEMENT IN LAW SCHOOL: IN CLASS AND BEYOND 8 (2010).
271. Id.
engaging in clinical work during law school, these lessons are not being fully diffused across all matriculants. Recent surveys of corporate general counsels confirm the unsatisfactory state of lawyer professionalism; a majority of those surveyed reported being dissatisfied with outside counsel because of a failure to keep the client adequately informed, non-responsiveness to client interests, making decisions without client authorization, and the failure to give clear advice.

We see, then, that the critique of legal education has a long pedigree, traversing the twentieth century and enduring through the present day. Indeed, the nature of the critique has been remarkably consistent, focusing on the poor connection between traditional legal education and legal practice.

IV. The Current State of Legal Education

“[T]he [law] schools . . . must be brought into a closer sympathy and contact with the profession than is now to be found . . . . It is unjust to students, and a fraud on the public, to recommend them as practitioners until they reach some creditable degree at least of skill and knowledge.”

—ABA Comm. on Legal Educ. & Admission to the Bar, 1879

Has contemporary legal education moved beyond this history? Although still fundamentally consonant with the

272. See NALP, 2010 SURVEY OF LAW SCHOOL EXPERIENTIAL LEARNING OPPORTUNITIES AND BENEFITS, supra note 31, at 6 (showing 30.2% of survey respondents had participated in at least one legal clinic).


Langdellian model, law schools have reformed in many ways since Langdell’s time. Professors have varied their teaching methods in ways that build on or depart from the case method. Law schools have pursued and implemented many of the reforms suggested in the reports reviewed above, offering basic legal research and writing training in the first year, requiring upper-level extensive writing experiences in line with the current ABA Standards, and ensuring that students have some opportunity to experience small class sizes and group work with other students. The relevance of other disciplines to the study of law has been recognized and incorporated into the curriculum through the introduction of interdisciplinary subjects.

276. See, e.g., Climenko Fellowship and First-Year Legal Research and Writing Program, HARVARD LAW SCH. (Sept. 5, 2012), http://www.law.harvard.edu/academics/degrees/jd/fyrlwp/ last visited Sept. 7, 2012 (“The First-Year Legal Research and Writing Program (LRW) is a series of sequenced, interrelated exercises introducing students to the way lawyers conduct legal research, analyze and frame legal positions, and present their work in writing and in oral argument.”) (on file with the Washington and Lee Law Review).

277. See, e.g., JD Writing Requirement, COLUMBIA L. SCH., http://www.law.columbia.edu/academics/registrar/writing last visited Sept. 14, 2012 (describing the details of its J.D. writing requirement) (on file with the Washington and Lee Law Review); Writing Requirement, NYU SCH. OF LAW, http://www.law.nyu.edu/academicservices/degreerequirements/jdprogram/writingrequirements/index.htm last visited Sept. 14, 2012 (“In order to graduate, a student must produce an original analytic paper of substantial length (ordinarily at least 10,000 words in length and undergoes a comment and draft process) under the supervision of a faculty member who may augment these requirements.”) (on file with the Washington and Lee Law Review).

278. See 2012–2013 ABA STANDARDS AND RULES OF PROCEDURE FOR APPROVAL OF LAW SCHOOLS, supra note 28, Standard 302(a) (requiring “at least one rigorous writing experience in the first year and at least one additional rigorous writing experience after the first year”).

279. The Cramton Report noted that some upper-level class sizes are small “simply because they furnish instruction relevant to a less heavily chosen career alternative or are less frequently elected for other reasons” and added that “[t]he instruction can be, but typically is not, significantly different in nature and method from that furnished in large classes.” CRAMTON REPORT, supra note 22, at 23.

or the infusion of such learning into traditional law courses.\textsuperscript{281} The clinical legal training movement has successfully imported live-client practice experiences into the law school framework,\textsuperscript{282} affording students the opportunity to opt for such an experience.\textsuperscript{283} And, increasingly, law schools are offering practical skills training courses that teach students the skills they will need to have to practice law.\textsuperscript{284} Indeed, curricular reform is the order of the day, as schools rush to outdo each other in adjusting their programs in various ways to improve their ability to produce practice-capable graduates.

Although these contemporary reforms are appropriate moves in the right direction that will yield results on the margins, to this point, they have not resulted in a wholesale change in the practice-readiness of American law school graduates, a failing that was reflected and explored in the 2007 \textit{Carnegie Report} and


\textsuperscript{284} See, e.g., \textit{Dramatic Third Year Reform, supra} note 283 (“Traditional classroom instruction will be replaced by practice simulations, real-client interactions and the development of law practice skills.”).
other recent studies. Indeed, the numerous shortcomings of the American model of legal education have been documented extensively: Law school does not routinely provide training in many of the practice skill areas—such as drafting, counseling, planning, client development, and client management—needed to be a successful practitioner; only a tiny percentage of law schools require clinical training and the majority of students graduate with no clinical experience; its primary pedagogical approach (the case-dialogue method) is ineffective and demoralizing; its main approach to assessment remains the final essay exam, which reflects little about the professional competency of students and comes too late to allow self-

---

285. See Stuckey et al., supra note 30, at 11–13 (explaining that law school programs cannot fully prepare students for law practice and should be viewed as only one step in the training process).

286. See John O. Sonsteng et al., A Legal Education Renaissance: A Practical Approach for the Twenty-First Century, 34 WM. MITCHELL L. REV. 303, 335–37 (2007) (explaining that law schools often graduate unskilled students, who must develop practical skills outside of school and after graduation); see also John M. Burman, Oral Examinations as a Method of Evaluating Law Students, 51 J. LEGAL EDUC. 130, 132 (2001) (“[T]he required curriculum at many, if not most, American law schools virtually ignores at least half of the fundamental skills every lawyer should have.”).

287. See NALP, 2010 SURVEY OF LAW SCHOOL EXPERIENTIAL LEARNING OPPORTUNITIES AND BENEFITS, supra note 31, at 6 (showing 30.2% of survey respondents had participated in at least one legal clinic, 36.2% of respondents had taken part in an externship, and 40% of respondents had taken three or more practice skills courses).

288. See Stuckey et al., supra note 30, at 132–41 (stating that the case method is encumbered with significant problems as an instructional tool, as its impact on students is sporadic, it fails to provide feedback to address skill deficiencies, and it neglects portions of the cognitive process); Bernard D. Meltzer, The University of Chicago Law School: Ruminations and Reminiscences, 70 U. CHI. L. REV. 233, 241 (2003) (arguing that the Socratic method is “notoriously inefficient” at teaching black letter law).

289. See Alan A. Stone, Legal Education on the Couch, 85 HARV. L. REV. 392, 407 (1971) (“Socratic teaching has been attacked as infantilizing, demeaning, dehumanizing, sadistic, a tactic for promoting hostility and competition among students, self-serving, and destructive of positive ideological values.”).

290. See Sonsteng et al., supra note 286, at 347 (“The traditional assessment system creates an illusion of higher achievement when there may actually be a deficiency in actual lawyering skills. Professors routinely observe students excel in written exams, but then watch as they struggle with interviewing and counseling clients.”).
improvement;\textsuperscript{291} faculty incentives promote scholarship over the needs of students;\textsuperscript{292} many professors (particularly the more recent ones) have little or even no experience practicing law and lack membership in the bar;\textsuperscript{293} and law school costs so

\textsuperscript{291}. See id. at 337–38 ([T]he traditional law school model does not provide regular or relevant performance feedback, so students have little opportunity to improve.).

\textsuperscript{292}. See James Lindgren & Allison Nagelberg, Are Scholars Better Teachers?, 73 Chi.-Kent L. Rev. 823, 827–29 (1998) (explaining data showing that major scholars, despite teaching larger and more classes, are perceived as better teachers than less-cited scholars); see also Deborah Jones Merritt, Research and Teaching on Law Faculties: An Empirical Exploration, 73 Chi.-Kent L. Rev. 765, 807 (1998) (showing that faculty who taught at prestigious universities taught fewer credit hours than faculty at less prestigious institutions); Fred R. Shapiro, They Published, Not Perished, But Were They Good Teachers?, 73 Chi.-Kent L. Rev. 835, 839 (1998) (observing that “while, generally, praise of teaching is a nearly universal feature of tributes to law faculty, for the most highly cited scholars, it is often completely absent from their tributes, and this despite the fact that such scholars typically are accorded much longer tributes than is the norm”).


[A] number of young lawyers flee the ‘rat race’ after only a short stay. Often, very bright young lawyers seek to move from law practice to law teaching as quickly as possible, with little practical knowledge or professional experience. This creates a conundrum whereby many of these smart young people who escape to academia have nothing good to say about practice, though they are the ‘teachers’ of the next generation of the legal profession.

Id.; see also Ann Juergens, Using the MacCrate Report to Strengthen Live-Client Clinics, 1 Clinical L. Rev. 411, 412–13 (1994) (“[A]n excellent academic record at an ‘elite’ law school, law review, clerkship for an appellate judge, perhaps a few years at a big law firm, but relatively little first-hand knowledge of dealing with clients, transactions, the courtroom, real-life conflict and problem solving.”).
much\textsuperscript{294} that most graduates have mammoth, mortgage-like debts that limit their economically viable options after graduating.\textsuperscript{295} This is no way to produce competent legal professionals. A closer look at some of the pillars of law school education—its curriculum, its pedagogy, its assessment methods, and its faculty—reveal that what is past is not only prologue, but it is largely our present, a fact that is problematic given the death of the apprenticeship and the dramatic changes in the law and legal practice since the time of Langdell.

\section*{A. The Curriculum}

Contemporary law school curricula are dominated by legal doctrine, as was the case during the nineteenth century. Currently, law school is typically a three-year program (when pursued full-time)\textsuperscript{296} that offers roughly the same basic set of


first-year doctrinal courses—featuring traditional common law subjects—and legal research and writing instruction, followed by electives that offer doctrinal, skills, or professional/clinical instruction at the election of the student, culminating with a major writing requirement of some kind. Students generally are not required to take particular courses beyond the first year—professional responsibility or “PR” being a typical exception—nor is their course load typically organized around tracks or concentrations, though many schools do offer the option
of pursuing courses within a specified area of focus. Although experiential learning is becoming something that many law schools are requiring students to have at least one encounter with, few schools require extensive practical skills and clinical training. Under this traditional approach, the primary focus within the law school curriculum is on doctrinal courses and the

---

302. Many schools offer the opportunity to pursue a concentration, which may or may not lead to a certification to that effect. See, e.g., Programs of Study, HARVARD L. SCH. (Aug. 31, 2012), http://www.law.harvard.edu/academics/degrees/jd/pos/index.html (last visited Sept. 14, 2012) (on file with the Washington and Lee Law Review). Harvard explains:

To guide you in pursuing deepening knowledge and progression as you move through the three years of law school and to create a tool for better coordination and collaboration between faculty members, the faculty has developed ‘programs of study.’ Students do not sign up for any program; nor should any student feel compelled to adhere to one.

Id. However, the author is unaware of any school that requires all of its students to organize their courses around a substantive concentration.


304. Exceptions, of course, do exist and were mentioned previously. See supra note 283 and accompanying text (discussing experiential learning opportunities at Washington and Lee and Northeastern). Gonzaga University School of Law provides more practical skills training than is typical by supplementing its basic first-year legal research and writing courses with a “Litigation Skills & Professionalism Lab” (the life of a case) in the fall and a “Transactional Skills & Professionalism Lab” (the life of a commercial transaction) in the spring. Course Descriptions and Frequency, GONZAGA UNIV. SCH. OF LAW (Spring 2010), http://www.law.gonzaga.edu/files/course-descriptions.pdf.
transmission of substantive knowledge, with experiential or clinical courses aimed at transmitting practice skills being elective offerings mostly taught by a separate group of faculty members (clinical or adjunct professors) with different titles and status.

This focus on academic instruction is a legacy of the importation of legal education into the traditional university and the desire of early reformers such as Langdell and Harvard’s President Eliot, and later the ABA and the Association of American Law Schools, to raise the standards and status of legal education from its apprenticeship roots. One consequence of this doctrinal approach is that the study of law is conceptualized as the study of legal rules—a Langdellian innovation—rather than a broader study of legal practice.
involving the study of legal regulation as a social phenomenon and training in the full array of methods and techniques that legal practitioners must be able to employ. Another consequence of this academic, doctrinal dominance is that law faculties built to deliver such curricula tend not to consist of experienced practitioners but rather career academics focused on legal scholarship.

However, as the 2007 Carnegie Report and other reports and commentators have emphasized, substantive legal knowledge—although a critical element of professional training—is not the sole or principal component of learning to be a practicing lawyer, nor is it realistic to expect that law school can impart the sum total of substantive knowledge needed for practice. To the contrary, there are several related but distinct levels of training that are necessary to become a competent legal professional: the acquisition of foundational knowledge and analytical abilities, the development of certain practical skills, and the formation of the professional values and judgment that define legal practice. The traditional law school curriculum

one on contracts—the patient explication of legal doctrine, the rule of law as the law of rules.

311. See Rubin, supra note 101, at 640 ("[S]ocial science has taught us to regard law as a social practice . . . . While law . . . may be understood as definitive statements by authoritative sources, legal practice is the total set of behaviors that are prevalent among those trained professionals.").

312. See Jean R. Sternlight, Symbiotic Legal Theory and Legal Practice: Advocating a Common Sense Jurisprudence of Law and Practical Applications, 50 U. MIAMI L. REV. 707, 712 (1996) ("[L]aw schools are often highly reluctant to hire 'the practitioner,' and . . . particularly at elite law schools, professors often sneer at books and articles oriented toward practice, doctrine or (perhaps worse) teaching." (citation omitted)).

313. See CARNEGIE REPORT, supra note 22, at 25–29 (describing necessary components of legal education that are lacking in the current system).

314. See KARL N. LLEWELLYN, THE BRAMBLE BUSH: SOME LECTURES ON LAW AND ITS STUDY 93 (1930) ("No, the nature of our system of multiple jurisdictions, the accidental constellation of our statutes, the inductive concrete method of our case-materials—these make the learning of our law entire, as information, hopeless."(emphasis added)); see also Frank, Why Not a Clinical Lawyer-School?, supra note 204, at 726 ("Of course it is impossible in three years, or indeed in thirty-three years, to give or take courses in all the subjects into which the subject we call 'law' can be subdivided.").

315. See CARNEGIE REPORT, supra note 22, at 27 ("For the sake of their...")
does not reflect the relative importance of each of these levels of learning because it focuses the lion’s share of attention on doctrinal instruction.\textsuperscript{316} Fixing that imbalance does not mean that law school should be entirely handed over to “vocational” or skills training, the reflexive critique seemingly lodged at any effort to make legal education more relevant to legal practice.\textsuperscript{317} Rather, it means that the current relationship between doctrinal, practical, and professional instruction must become more integrated and balanced: Skills-based and practice-centered instruction and training should be a more substantial part of the law school experience, with substantive knowledge instruction serving as the foundation for and compliment to practical professional education. Further, to facilitate the transition from student to practitioner, doctrinal instruction must move more quickly into being taught in context from the operational perspective,\textsuperscript{318} rather than more abstractly through the prism of judicial opinions and the case method.\textsuperscript{319} Finally, students must

\textsuperscript{316} See Margaret Martin Barry et al., \textit{Clinical Education for This Millennium: The Third Wave}, 7 CLINICAL L. REV. 1, 32 (2000) (“The analysis of legal doctrine as presented in appellate decisions digested in casebooks . . . continues to frame most classroom discourse . . . ”).  

\textsuperscript{317} See BLACKSTONE, supra note 39, at *32 (“If practice be the whole he is taught, practice must also be the whole he will ever know: if he be uninstructed in the elements and first principles upon which the rule of practice is founded, the least variation from established precedents will totally distract and bewilder him.”); see also Fuller, supra note 315, at 191 (“[A]s soon as an attempt is made to employ the skills-and-techniques conception as the exclusive standard for organizing legal education, the whole educational process is disoriented and cheapened.”).  

\textsuperscript{318} Kristen Holmquist refers to this as “applied learning opportunities.” Kristen Holmquist, \textit{Challenging Carnegie}, 61 J. LEGAL EDUC. 353, 357 (2012).  

\textsuperscript{319} See STUCKEY ET AL., supra note 30, at 99 (“Law schools cannot prepare students for practice until they teach doctrine, theory, and practice as part of a unified, coordinated program of instruction.”).
have more opportunities to collaborate in team settings and to work on solving problems that blend legal and other issues in a single setting. Many law schools seem to be making moves in these directions, though how thorough and effective these changes will be at each school remains to be seen.

B. The “Signature Pedagogy” of Law School

“[I]t is obvious that man could hardly devise a more wasteful method of imparting information about subject matter than the case-class.”

—Karl Llewellyn, 1948–1949

Although the case-dialogue method has been criticized and modified in many ways over the years, it retains its basic hold as the fundamental framework for teaching law students legal doctrine and analysis to this day. Indeed, notwithstanding the

320. See Michael Hunter Schwartz, Sophie Sparrow & Gerald Hess, Teaching Law By Design: Engaging Students from the Syllabus to the Final Exam 7 (2009) (“Students engage in crucial mental activity when they negotiate meaning and seek to synthesize their personal understandings. The hundreds of studies demonstrating the superiority of cooperative learning groups compared to all other teaching methods support this assertion.”).

321. See, e.g., Romero, supra note 280 (announcing that Stanford Law School is “transforming its traditional law degree into a multi-dimensional J.D., which combines the study of other disciplines with team-oriented, problem-solving techniques together with expanded clinical training that enables students to represent clients and litigate cases while in law school”).


323. See, e.g., Carnegie Report, supra note 22, at 56–59 (describing the case-dialogue method as insufficiently connected with practical context and ethical values); Elizabeth Mertz, The Language of Law School: Learning to “Think Like a Lawyer” 190, 202 (2007) (explaining that the current methods used in law schools to get students “thinking like lawyers” have disparate effects depending on students’ race and gender); Redlich, supra note 95 (criticizing the case method for failing to give students a picture of the law in its entirety); Stuckey et al., supra note 30, at 41–47, 132–34 (noting the negative impacts of the case-dialogue method on the students’ education).

324. See Carnegie Report, supra note 22, at 50 (“[L]aw schools use case-dialogue teaching almost exclusively in the first phase of doctrinal instruction.”); see also Mertz, supra note 323, at 41 (describing the Socratic method of teaching as the “discourse for which law school is famous”). Edward Rubin has
myriad changes in the legal profession and in our understanding of how people learn, the contemporary law school remains remarkably Langdellian in its design as a three-year system in which doctrinal legal knowledge and legal analytical abilities are transmitted to students mostly via a traditional or modified case-dialogue approach, supplemented with optional or mandatory experiential learning components. It must be acknowledged, however, that most professors vary from a pure Socratic method in their doctrinal courses, and that the method has virtually no place in experiential courses. Thus, it would simply be written that the case method has reached venerated status, making any critiques a threat to an esteemed tradition:

Continuing on for another seventy years or so, [the Langdellian model] has ceased to be viewed as a particular approach to legal education—as last generation’s innovation—and has become a venerable institution that gains gravity and prestige from its antiquity. As such, this approach has the remarkable capacity to make suggested changes seem jejune and to reduce reform initiatives to quixotic ventures that can be dismissed with knowing guffaws from its wiser, more experienced supporters.

Rubin, supra note 101, at 613.

325. See infra notes 350–53 and accompanying text (explaining and differentiating learning techniques).

326. See Llewellyn, The Bramble Bush, supra note 314, at 145 (“Our teaching technique we have . . . carried for the most part unaltered into wholly changed conditions . . . .”); see also Todd D. Rakoff & Martha Minow, A Case for Another Case Method, 60 VAND. L. REV. 597, 597 (2007) (“The plain fact is that American legal education, and especially its formative first year, remains remarkably similar to the curriculum invented at the Harvard Law School by Christopher Columbus Langdell over a century and a quarter ago.” (citation omitted)).

327. See Rubin, supra note 101, at 610 (“[T]he basic educational approach that law schools use remains essentially unchanged from the one that C.C. Langdell introduced at Harvard in the years following the Civil War.”).

328. See Carnegie Report, supra note 22, at 34–43 (describing programs at NYU and CUNY designed to provide students with practical, hands-on legal experience during law school).

329. See Mertz, supra note 323, at 142 (indicating that “pure” Socratic teaching was encountered in only one class within the study); see also id. at 142–69 (describing variations of the Socratic teaching style as “pure” or “traditional,” “modified,” a mixed method of “dialogue, lecture, and conversation,” and a “dialectic” lecture method).

330. See Carnegie Report, supra note 22, at 59 (“[C]ase-dialogue teaching is seldom explicitly connected with clinical teaching . . . .”).
inaccurate to characterize all of law school education as merely a sequence of ineffectual classically Socratic experiences. That said, it is true that traditional doctrinal courses—which do in fact dominate the curricula at virtually all law schools—tend to involve the study of doctrine through the lens of cases and casebooks\(^{331}\)—even in statutory courses\(^{332}\)—and that professors typically use a mix of lecture and Socratic questioning as the principal means of covering the material in their courses throughout law school.\(^{333}\)

While there may be strengths to the case-dialogue method,\(^{334}\) there are several shortcomings worth noting here. It is often said that the virtue of the case method is the training it can impart in the skills of legal reasoning and analysis—“thinking like a lawyer”—that are critical components of professional

---

331. See id. at 55 (“The legal texts that form the basis for the case-dialogue method are found in a unique invention of legal pedagogy—the case book.”). The material used in a typical doctrinal law school course is some variant of a casebook—a text collecting edited cases organized by topic and supplemented with notes and questions—rather than a legal treatise or traditional textbook of the kind found in other disciplines such as biology, history, or accounting. The inclusion of such “materials” (notes, questions, article excerpts, etc.) in modern casebooks is what distinguishes them from the very first casebook, which was developed by Langdell himself and solely consisted of cases organized by the legal principles to which they related. See C. C. Langdell, Selection of Cases on the Law of Contracts (Neil H. Alford, Jr. et al. eds., The Legal Classics Library 1983) (1871).


333. See Carnegie Report, supra note 22, at 3 (“Law schools use Socratic case-dialogue instruction in the first phase of their students’ legal education. During the second two years, most schools continue to teach, by the same method, a number of elective courses in legal doctrine.”); see also Mertz, supra note 323, at 144–69 (finding, among the classrooms contained in the study, that most professors used variations of the Socratic method that blended questioning about cases with discussion and lecture).

334. Rakoff & Minow, supra note 326, at 598

The Langdellian case method afforded a way to communicate information; to cultivate a style of reasoning and questioning that was intellectually respectable, yet also well-suited to the paradigmatic law practice of adjudication; and to engage the attention and interests of large numbers of students at relatively little expense for instruction and materials.
These are skills that all law students must acquire if they are to become fluent in the language of the law, jurisprudence, and the art of using cases to derive and develop legal principles. However, the ability of the case-dialogue method to transmit analytical skills effectively has never been demonstrated. Indeed, Elizabeth Mertz, in her seminal study of the case method, *The Language of Law School*, describes studies of teaching methods that fail to show any connection between the method used and the ability of students to engage in effective legal analysis. Additionally, the type of thinking promoted by the method is limited to certain kinds of legal analysis, neglecting some of the basic problem-solving skills that today’s practitioners need to develop solutions to their clients’ problems.

---

335. See, e.g., CARNEGIE REPORT, supra note 22, at 53 (describing the “deep structure” of the case-dialogue pedagogy as “the teaching of legal reasoning”); id. at 54 (indicating that the “dispositions and attitudes” modeled by the case-dialogue pedagogy are “habits of legal thinking” such as distancing from extraneous detail to focus on points of legal argument); CRAMTON REPORT, supra note 22, at 13 (“The traditional ‘socratic method’ of legal instruction continues to be used in first year law classes as an extremely effective technique for developing analytical skills . . . .”); LLEWELLYN, THE BRAMBLE BUSH, supra note 314, at 116 (“The first year . . . aims, in the old phrase, to get you to ‘thinking like a lawyer.’”); but see Rubin, supra note 101, at 610–11

What one sometimes hears is that the current law school curriculum teaches students to ‘think like lawyers.’ Any systematic demonstration that such an outdated approach to legal education develops skills that are central to the very different world of modern legal practice would be interesting to see, but no such demonstration has been offered. (citation omitted).

336. See, e.g., Frank, Why Not a Clinical Lawyer-School?, supra note 204, at 910 (“[N]o sane person will deny that a knowledge of those rules and principles, of how to ‘distinguish’ cases, and of how to make an argument as to the true ratio decidenendi of an opinion, is part of the indispensable equipment of the future lawyer.”).

337. See MERTZ, supra note 323, at 28 (“[C]ontrolled experiments in which first-year classes were divided into separate groups, some taught Socratically and others not, resulted in generally similar performances.” (citing Edward Kimball & Larry Farmer, *Comparative Results of Teaching Evidence Three Ways*, 30 J. LEGAL EDUC. 196, 196–212 (1979); Willard Lorenson, *Concentrating on a Single Jurisdiction to Teach Criminal Law—An Experiment*, 20 J. LEGAL EDUC. 361, 361–65 (1968))).

338. See Kagan, supra note 283, at 477 (“[W]hile the case method does a great job teaching students a certain type of legal reasoning, it fails to equip them fully to serve as active problem solvers, able to engage a range of resources
Another weakness of the case-dialogue method, as currently employed in law schools, is that whatever benefits might accrue from the method are difficult to achieve among large groups of students. When the method is used in the context of large classes—which is typical—it tends to focus attention on a discussion between a few students and the professor. Although the students involved may benefit to some extent, the method is less effective in instilling legal analytical skills vicariously to observers not involved in the discussion, creating diminishing returns as the class grows in size. Relatedly, large classes conducted under the Socratic method involve sizeable audiences of peer onlookers, potentially contributing to the stress and anxiety of students expected to respond to the professor's questioning and creating an intimidating environment that

339. See MERTZ, supra note 323, at 7 (“Although many law schools are now experimenting with smaller first-year classes, it is still common to find the bulk of a first-year student’s time spent in larger classes of seventy to one hundred students.”); id. at 202 (identifying patterns in her study linking “increased class participation and classroom presence with traditional insiders . . . ; that is, white male students tend to predominate.”); id. at 177 (“Use of recitation (the closest analogue to Socratic dialogue), with its intensely public potential for evaluation of responses (both by teachers and peers), tends to encourage the formation of entrenched, segregated groups.” (citing STEVEN BOSSERT, TASKS AND SOCIAL RELATIONS IN CLASSROOMS: A STUDY OF INSTRUCTIONAL ORGANIZATION AND ITS CONSEQUENCES (1979))).


341. See STUCKEY ET AL., supra note 30, at 222 (“[I]nevitably, many students have not participated in the dialogue; some, overwhelmed by the relief that they were not the one called on, have not even listened attentively.” (citation omitted)); Schwarz, supra note 340, at 351 (“[L]aw teaching requires students to learn vicariously.”).

342. See STUCKEY ET AL., supra note 30, at 134 (“The potential value of the Socratic dialogue and case method is diminished, however, because we use it in large classroom settings . . . .”).

343. See Jennifer Jolly-Ryan, The Last Taboo: Breaking Law Students with Mental Illnesses and Disabilities Out of the Stigma Straitjacket, 79 UMKC L. REV. 123, 144 (2010) (“[D]ue to the nature of a legal education, such as stress, competition, and longstanding traditions including the Socratic Method and grading policies, depression and anxiety may develop.”); see also Lawrence Silver, Comment, Anxiety and the First Semester of Law School, 1968 Wis. L.
may be more discouraging to women\textsuperscript{344} and students of color.\textsuperscript{345} These dual challenges surrounding the use of the method in large groups indicate that law schools take what theoretically may be a sound pedagogical approach and dilute it, suggesting that moving the Socratic case-dialogue out of the large classroom into a much smaller setting with only a handful of students might be a structural alteration that would allow the approach to bear its fullest fruit.

Next, the case-dialogue method is an inefficient means of transmitting substantive information and is limited in its ability to impart the full range of competencies that students need to become successful legal professionals.\textsuperscript{346} There have been many advances in learning theory and pedagogy since Langdell’s time with which the case-dialogue method is out of step.\textsuperscript{347} Yet the

\begin{itemize}
\item \textsuperscript{344} See \textsc{Mertz, supra} note 323, at 190 ("Our data tend to confirm the findings of previous studies . . . . that male law students generally participate at greater rates than females."); see also \textsc{Lani Guinier, Michelle Fine & Jane Balin, Becoming Gentlemen: Women, Law School, and Institutional Change} 49 (1997) ("The pedagogical structure of the first year—large classes, often constrained by limits on student participation, fierce competition, a mandatory grading curve, and few women faculty—produces alienation and a gender-stratified hierarchy.").
\item \textsuperscript{345} See Carol J. Buckner, \textit{Realizing Grutter v. Bollinger’s “Compelling Educational Benefits of Diversity”—Transforming Aspirational Rhetoric into Experience}, 72 \textsc{UMKC L. Rev.} 877, 911 (2004) ("The academic achievement of African American students improves when teachers use cooperative rather than competitive learning strategies . . . ." (citation omitted)); see \textit{also id.} at 905
\item Hispanic students do not like to be singled out; they function more effectively working in groups to achieve a common goal and are receptive and susceptible to thoughts and attitudes expressed by others. Encouragement, group work and establishing a sense of ‘belonging’ all help to create a positive learning experience for Hispanic students. These learning preferences describe the direct antithesis of the typical law school environment, and explain the dysfunction and disparate achievement of Hispanic students. (citations omitted).
\item \textsuperscript{346} See \textsc{Carnegie Report, supra} note 22, at 58–59 ("[L]earning the law is an ensemble experience . . . . [T]he apprenticeships of cognition, performance, and identity are not freestanding . . . . Because case-dialogue teaching is seldom explicitly connected with clinical teaching, few law schools achieve the full impact that an integrated ensemble could provide.").
\item \textsuperscript{347} See Schwarz, \textit{supra} note 340, at 383
\end{itemize}
Langdellian case-dialogue method—a relic of a bygone era—persists. Insights from learning theory reveal that teaching focused mainly on purely abstract concepts divorced from their context—something that fairly characterizes the case method—is less effective than teaching that recognizes that we experience information in many different ways and at different levels of abstraction. Edgar Dale visualized these various levels of encountering information with his Cone of Experience, which depicted “a range of experience from firsthand action to observation (iconic experiences) on to symbolic communication.” For example, one can understand what a knot is directly by tying the knot (referred to by Dale as an “enactive” experience), visually by simply seeing a picture of a knot (an “iconic” experience), or abstractly by simply hearing or seeing the word “knot” or a verbal description of the phenomenon (a “symbolic” experience). Learning often moves from direct experience or iconic experience toward abstractions as words or symbols whose meaning we come to understand, although this

All three learning theories discussed—behaviorism, cognitivism, and constructivism—are relevant to law school instruction, yet, for the most part, the legal academy has ignored these theories. Together, the theories suggest that instruction should cause students both to build their skills from base levels to the highest levels, and to move from simply knowledge of information to the creative problem-solving contemplated by the constructivist model.

348. See Rubin, supra note 101, at 611 (“Langdell’s design for legal education, although innovative in its own time and on its own terms, is more closely connected to modes of thought that prevailed in the Renaissance, the Middle Ages, and ancient Greece and Rome than to our current ways of thinking.”).

349. See Mertz, supra note 323, at 26 (“[D]espite a number of arguably successful attacks on the substantive underpinnings of Langdell’s approach, the method itself appears to have outlasted its theoretical rationale.”).

350. Edgar Dale, Audiovisual Methods in Teaching 110 (3d ed. 1969). The levels of the Cone, from bottom to the top, are direct purposeful experiences, contrived experiences, dramatized experiences, demonstrations, study trips, exhibits, educational television, motion pictures, recordings, radio and still pictures, visual symbols, and verbal symbols. See id. at 107 (illustrating the eleven levels of Dale’s “Cone of Experience”).

351. Id. at 108. Dale borrows these terms from Jerome S. Bruner, Toward A Theory of Instruction, 10–11 (Harvard Univ. 1966).

352. Id. at 108–09.
is not always the case. However, when learning consists solely of abstractions untethered from lower-order experience, problems can arise:

Difficulties arise when abstractions have inadequate foundations. If a learner has had too little enactive or iconic experience in acquiring a particular summarizing idea, the word or formula will probably have no real meaning for him. Because a verbal symbol does not resemble anything the student can do or see, he may have difficulty in relating it to his own experience. If a symbol is to stand for something, it must stand on something—a firm foundation of relevant experience.

This is not to say that enactive or iconic experiences must always precede abstract learning; rather, abstract learning can be enhanced by direct or visual experiences that can concretize and deepen the understanding of the abstract concept at issue.

The application of Dale’s insights to legal education is clear: Law school learning exclusively rooted in symbolic, abstract experience is less likely to be effective in giving students the depth of understanding requisite for moving towards proficient legal practice. Further, to the extent that legal learning is exclusively at the abstract level, it becomes difficult for students to synthesize learning from different areas or to operationalize concepts for practical application and the resolution of real-world legal problems. To be clear, this is not an argument that the

---

353. See id. at 128 ("Does the Cone device mean that all teaching and learning must move systematically from base to pinnacle? Emphatically no . . . . We continually shuttle back and forth among various kinds of experiences.").

354. Id. at 109.

355. See id. at 128–29

In our teaching, then, we do not always begin with direct experience at the base of the Cone. Rather, we begin with the kind of experience that is most appropriate to the needs and abilities of a particular learner in a particular learning situation. Then, of course, we vary this experience with many other types of learning activities.

356. See id. at 132 ("Even the most advanced student, therefore, can deepen his understanding of concepts . . . by participating in experiences all along [the] Cone.").

357. See id. at 134

A teacher may move students so swiftly to the symbolic level of thought, and with so little preparation, that their concepts will lack
best learning occurs at the lower, more direct experience levels of Dale's Cone, nor is it the often misattributed notion that we remember more of the things that we learn from direct experience. Rather, the claim is that different types of learning experiences are possible and that legal teaching needs to make an intelligent use of a mix of these experiences to give students the level of understanding needed for effective learning and translation into practical application.

These deficits of the case-dialogue method become harder to tolerate once students have acquired the legal analytical skills that the method is designed to impart. Indeed, there may be deep roots in direct experience. These rootless experiences will not have the generative power to produce additional concepts and will not enable the learner to deal with the new situations that he faces.

358. Many have misappropriated Dale's Cone of Experience to present a revised pyramidal image that depicts the levels of experience with indications of the percent of information retained when learned through each approach. See, e.g., Learning Objectives, MINN. STATE UNIV. MOORHEAD, http://web.mnstate.edu/InstrTech/SCModules/LearningObjectives/index.html (follow “Dale’s Cone of Experience” hyperlink) (last visited Aug. 24, 2012) (on file with the Washington and Lee Law Review). Dale himself never attributed such percentages of retention to his Cone nor did he support the implication that experiential learning was necessarily superior to more abstract learning. See DALE, supra note 350, at 128–30; see also Michael Molenda, Cone of Experience, in EDUCATION AND TECHNOLOGY: AN ENCYCLOPEDIA 161, 164 (Ann Kovalchick & Kara Dawson, eds., 2003) (discussing the common misappropriation of the theory, noting that “[a]t some point[,] someone conflated Dale’s Cone of Experience with a spurious chart that purports to show what percentage of information people remember under different learning conditions”). This misrepresentation of Dale’s Cone was unfortunately featured in a recent article on legal education reform. See Sonsteng et al., supra note 286, at 309 (attributing to Dale the idea that “the least effective methods of instruction include reading text and listening to lectures,” notions that Dale himself expressly disavowed in the very work cited by Sonsteng).

359. See STUCKEY ET AL., supra note 30, at 132 (“Law teachers need to be multi-modal in our teaching and reduce our reliance on the Socratic dialogue and case method. There are many tools for reaching students than one finds in the typical law school classroom.”).

360. Jerome Frank estimated that students could be taught the dialectical method of legal analysis imparted by the case-dialogue method within six months, after which time other more effective methods of instruction should be used. See Frank, What Constitutes a Good Legal Education?, supra note 210, at 726. Intelligent men can learn that dialectical technique in about six months . . . . Teach them the dialectic devices as applied to one or two
declining benefits associated with a continued employment of the case-dialogue method throughout the second and third years of law school. Thus, once this point has passed and basic foundational legal principles have been explored using the method in the first year, the pedagogy should shift toward alternative methods that help students acquire knowledge and skills in the manner and contexts that will be required of them as legal practitioners. In other words, a thorough understanding of legal principles and the ability to “think like a lawyer” need to become the foundation for the next step in professional development—developing the ability to handle complex problems of clients in a skilled and professional manner.

Finally, the case-dialogue method presents the law through the lens of litigation, most of which is appellate fields and they will have no trouble applying them to other fields. But in the law schools, much of the three years is squandered in applying that technique over and over again to a variety of subject matters . . . 

See also Carnegie Report, supra note 22, at 77 (speaking of the “diminishing returns” problem in legal education characterized by the “drop-off in interest and effort in classroom learning as students move through law school”).

361. See, e.g., Stuckey et al., supra note 30, at 140 (“Unfortunately, many law teachers continue to rely exclusively on the Socratic dialogue and case method, not just in the first year, but also in the second and third year courses long after students become competent in case analysis and ‘thinking like a lawyer.’ This contributes to student boredom and loss of interest in learning.”). This is not a new observation; Harvard students in 1935 complained that after the first year, the benefits of the case method were dramatically reduced. See Stevens, supra note 123, at 161 (“After the first year the case method allegedly lost its value; the students thought it should be dropped or modified in the second and third years.”); see also id. at 246 (“In the early 1970s, there was extensive evidence that outside those working on law review there was a dramatic falloff in energy levels and work at the end of the first year, if not in the first semester.” (citation omitted)).

362. See Carnegie Report, supra note 22, at 82 (“Practice requires not the distanced stance of the observer and critic but an engagement with situations.”).

363. This term is not used here to suggest that the universe of critiques of the case-dialogue method have been exhausted. Quite to the contrary, extensive critiques of the method abound. Elizabeth Mertz’s work helpfully summarizes the most pertinent of these and offers an extensive argument in favor of her own critique. See Mertz, supra note 323. One of her most incisive observations is that “thinking like a lawyer” is not a superior mode of analysis but rather is a kind of analysis that preferences certain bases for authority, modes of understanding, sources of knowledge, and types of arguments over others for
litigation, and does so in a highly formalized and acontextual manner that skews students’ perspective away from the realities and complexities of raw facts, clients, and professional responsibility.\textsuperscript{364} The “thinking like a lawyer” modeled by the case-dialogue method often strips disputes from their context and emphasizes formal and procedural issues over other moral or personal factors that might bear on reaching a more complete appraisal of the justice of an outcome.\textsuperscript{365} Further, by being rooted in court decisions, the law is learned as the product of conflict, as a battle among adversaries that yields legal pronouncements and interpretations. But law today is not developed simply or even largely through litigation but through legislation, regulation, and negotiated agreements. These aspects of law are underappreciated in the existing curriculum, as even the bodies of law that emanate from these latter sources are studied through the eyes of court decisions interpreting them.

\begin{quote}
\textsuperscript{364} See \textit{Carnegie Report}, \textit{supra} note 22, at 53 ("[S]tudents tend to think of legally relevant facts as they are presented in the appellate opinions that they typically read for class discussion . . . . As a first encounter with legal facts, this can give the misleading impression that facts are typically easy to ‘discover,’ rather than resulting from complex processes of interpretation that are shaped by pressures of litigation."); see also \textit{Mertz, supra} note 323, at 95 ("[P]rofessors are conveying a linguistic ideology centered on the crucial structuring role of layers of authority, discernable in the text. Emotion, morality, and social context are semiotically peripheralized in this process.").

\textsuperscript{365} See \textit{Carnegie Report}, \textit{supra} note 22, at 52 (summarizing Mertz’s description of an ethos “that emphasizes the formal, procedural aspects of legal reasoning as the central focus, making other aspects of cases peripheral or ancillary”).
\end{quote}
In the mid-twentieth century, a faculty committee at Harvard Law School described some of the benefits of the case method when it wrote:

The problems most naturally raised by reading a series of appellate cases are: Were these cases “rightly” decided? Are they consistent with one another? Can a pattern of decision be discerned that will reconcile them, even though their language is in conflict? On the basis of these decisions, how would this hypothetical case be decided? All of these inquiries are eminently worthwhile, and afford a useful training for the lawyer.366

This view is sound; the study of cases can be a worthy vehicle for learning about the law and legal analysis. However, the case method as practiced focuses on appellate opinions of judges, which do not reflect the sum total of what factored into the how particular cases were litigated or decided,367 including considerations of basic justice and fairness.368 Omitted is any consideration of the underlying record, including documents, evidence, pleadings, trial transcripts, trial court rulings, and the like, or the raw client narratives and other facts that faced the practitioner at the pre-litigation, problem-solving phase of the representation,369 in favor of a retrospective view that stymies the

---

367. See id. at 35–36
   In general American legal education with its emphasis on the appellate phase of litigation trains the lawyer in testing the validity of a position already taken, not in the problem of deciding what position to take or what course to follow. . . . Because law school instruction is largely based on appellate decisions, its focus is inevitably upon this last phase of a controversy, and the student receives little direct training in the choices that have to be made before this final phase is reached.
368. See MacCRA TE REPORT, supra note 22, at 236
   Too often, the Socratic method of teaching emphasizes qualities that have little to do with justice, fairness, and morality in daily practice. Students too easily gain the impression that wit, sharp responses, and dazzling performance are more important than the personal moral values that lawyers must possess and that the profession must espouse.
369. See FULLE R ET AL., supra note 26, at 41 (“The rapid exchange of intellectual repartee that has characterized some of the best case-method instruction in the past may be excellent training for the appellate advocate, but
development of the "legal imagination" needed to develop solutions to legal problems prospectively. A true case method, as Jerome Frank recognized long ago, would entail a study of the entire "case" rather than the edited and refined representation of a dispute one finds in appellate opinions. Such an approach would permit students to assess how facts, legal doctrine, and other factors such as arguments raised, questions asked, and strategic decisions made all combined to yield a given result in one case versus another, as well as permitting them to think about how they would use the facts and the law to shape an alternate approach to the matter that might have led it down a different path entirely. Additional alternative approaches to teaching the law have been richly covered by other critics including those detailed in the 2007 Best Practices Report as well as in work such as Teaching Law by Design by Professors Schwartz, Sparrow, and Hess. However, these alternatives have not permeated the law school culture, where the legacy of the Langdellian case-dialogue method retains its sway.

it hardly furnishes the appropriate atmosphere for a discussion of the soundest solution for a practical problem of legal planning.

370. See Rakoff & Minow, supra note 326, at 602 (describing "legal imagination" as "the ability to generate the multiple characterizations, multiple versions, multiple pathways, and multiple solutions, to which [students] could apply their very well honed analytic skills").

371. See id. at 600

By taking a retrospective view of facts already found and procedures already used by a court, the appellate decision does little to orient students to the reality of unfolding problems with facts still to be enacted, client conduct still to take place, and procedural settings still to be chosen and framed.

372. See Frank, Why Not a Clinical Lawyer-School?, supra note 204, at 916.

373. See Stuckey et al., supra note 30, at 105–234 (suggesting the use of multiple methods of instruction, including context-based instruction, to teach law school courses).

374. Schwartz et al., supra note 320, at 107–34 (discussing a mixture of modeling, coaching, lecture, questioning, story-telling, visual-aids, and simulations to teach law school classes).
C. Assessment in Law School

“Our machinery for checking our results [of our teaching] with you [the students] would set an intelligent ass to braying.”

—Karl Llewellyn, 1930

Karl Llewellyn, who in the above quote in no uncertain terms said that law schools put the “ass” in “assessment,” recognized over sixty years ago that the traditional method of assessing student performance in law school was an ineffective means of measuring student learning. More specifically, there was and still is a gap between the professed learning objectives of many law school classes—teaching students to think like lawyers and to master certain legal doctrines—and the dominant method of measuring students’ attainment of that learning—the final essay exam, which tests more so what a student knows rather than what a student can do. Recall that the Carnegie Report emphasized the multiple apprenticeships of professional legal education—the cognitive, the practical, and the professional/ethical. Becoming a competent practitioner requires training in all three areas, though—as has been discussed—law school is disproportionately oriented towards the cognitive. Thus, the single final essay exam is typically drafted in a manner that requires students to display their mastery of legal concepts and doctrines learned during the semester through the analysis of various hypothetical problems drafted by the

---

375. Here, I am referring to the individualized assessment of student learning. Supplemental to such assessment is the law school’s assessment of its own institutional effectiveness, which can be measured in several ways. See 2011 Proposed Draft, supra note 195, Interpretation 305-2 (listing reviewing individual student assessment, evaluation of student portfolios, student surveys, student performance in capstone courses or other skills courses, bar exam passage rates, and bench and bar surveys of student performance as acceptable methods of measuring institutional effectiveness).

376. LLEWELLYN, THE BRAMBLE BUSH, supra note 314, at 139.

377. See David P. Bryden, What Do Law Students Learn? A Pilot Study, 34 J. LEGAL EDUC. 479, 480 (1984) (“There is, then, a serious dissonance between our higher aspirations as teachers and our examination and grading practices. We aspire to teach mental habits that transcend substantive law but we do not try very hard to find out how well we are succeeding.”).

378. See supra notes 257–61 and accompanying text.
professor. These exams will either put the student in the role of a judge who must resolve a particular legal question, as an advocate who must argue in support of a particular position or decide on a course of action, or sometimes in the role of a policymaker who must decide how the law should be designed to deal with a given situation. Such questions test the students’ understanding of the law and their ability to engage in proper and rigorous legal analysis.

That said, the limitations of this type of assessment are twofold. First, it is purely summative, in that it comes at the end of a course and attempts to measure learning after the course has been completed. In most courses, formative assessments—which measure student learning along the way—are underutilized or neglected entirely.379 But formative assessment is an important component of the learning process, as students need to have the opportunity to measure their understanding—or lack thereof—at a point in time when they still have the opportunity to make corrections and improvements. It is better to learn early on that one’s understanding of a concept or doctrine is confused so that the lessons of the remainder of the semester can build on a solid foundation of understanding rather than simply cumulating atop confusion and uncertainty. Formative assessment is critical to coaching and guiding students through their exploration of a topic, taking care to notice when a lesson has not been fully learned and prodding them in the right direction as they proceed through the course.380 Thus, law school courses should have multiple assessment exercises along the way—graded or ungraded—that permit the instructor to determine students’ learning levels and give feedback that will permit the student to adapt and improve.

Second, while typical essay exams do, to some extent, engage the analytical abilities needed of a judge or an advocate arguing a

379. Recall that unlike most law school courses today, legal instruction at the Litchfield Law School in the early nineteenth century included regular examinations every Saturday. \textit{Supra} note 64 and accompanying text.

380. \textit{See} \textit{Carnegie Report}, \textit{supra} note 22, at 164–67 (discussing criticisms of summative final exams and how students, “in the absence of feedback during the semester, [have] no basis on which to gauge whether they [are] mastering the material”).
legal point, such skills are not the sole or principal skills required of most competent practitioners. In the litigation context, practitioners often begin with an unfiltered narrative from a client, which must be distilled into a coherent set of facts that can then be translated into potential legal claims. As an attorney proceeds with or responds to a claim, legal doctrine is not the only determinant of how the case should be litigated or of how issues that arise should be resolved. There are ethical and strategic considerations as well as intangible factors such as the profile of the parties and the judge, the nature of the claims, what happens to be at stake in the case, the relevance of issues in this case to other cases, or other larger policy concerns. All of these factors, and more, must be taken into account as an attorney decides what actions to take and what arguments to make. Engaging in this process certainly requires a good grasp of the relevant and applicable legal doctrine. But sensitivity to the full range of pertinent factors that bear on legal strategy and judicial decision-making takes practical experience and the development of sound professional judgment. Traditional doctrinal courses and their associated final exams tend to abstract all of these things out of legal problems, isolating doctrinal (and perhaps policy) analysis as the key to how any given issue is resolved.\textsuperscript{381} That is an unrealistic picture of most legal practice that hardly suffices to prepare a student for actual practice.

Compounding the deficiencies of summative assessments in law school is the accompanying system of grading that characterizes most law school programs. A student’s grade in a course is typically determined by his or her performance on a

\textsuperscript{381} See Mertz, supra note 323, at 10

At the end of your first semester comes exam time. Now you have a chance to demonstrate your newly acquired legal vision of the world. In exam after exam, you are asked to respond to hypotheticals, stories made up by your professors. These stories are often replete with pathos and drama. Your job is to ignore as much of the emotional content as you can while hunting for the details that are relevant to the legal tests and frames, steadfastly averting your gaze from the human perfidy, misery, justice, or injustice found in the story. Once you’ve done that, if you’re very careful, you can throw in a little discussion of fairness, disguised as a “policy” argument, and sometimes get some extra points.
single final exam (perhaps with some credit or demerit for class participation), and that grade is constrained by grade normalization policies: Law schools enforce a grade distribution around a predetermined course mean (the curve). The forced distribution converts grading from an evaluation of the students' achievement to a system of ranking among course peers. The set mean—typically a B+ or 3.3 at most elite law schools—determines the label used to signal an average performance, divorcing traditional A through F letter grades from their essentialist identities (that is, A is excellent, B is good, C is average, D is below average, and F is a failure) and creating an alternative, discordant world in which high letter grades are used to give a faux sense of achievement that would be absent were the traditional labels for excellent, average, and below average used. This is quintessential grade inflation. Indeed, the forced mean all but guarantees that the grades for all students will hover around the upper echelon; actual failures will be rare or nonexistent, even if a student’s performance in truth reflects a failure to demonstrate proficiency with reference to the desired learning objectives for the course.

What are the problems with such an approach to grading? The principal problem is that what starts off as a poor assessment of the knowledge, skills, and values that an attorney must have to perform competently in any given situation becomes not even that; in reality the assessment is one of relative merit.


383. See Catherine Rampell, In Law Schools, Grades Go Up, Just Like That, N.Y. Times, June 22, 2010, at A1 (describing how several law schools have retroactively increased the mean G.P.A. in an effort to improve the students’ chances with employers and the schools’ reputation).

384. When there is a forced B+ mean, giving failing grades or D grades create such a drag on the course mean that more top grades have to be given to more people, weakening the distinction between top performers and above-average or mediocre performers.

385. Are there good reasons to measure relative merit from an institutional perspective? If there is a system of rewards the distribution of which depends on
not of learning or achievement.\textsuperscript{386} That is, in a \textit{B+} curve grading
system, an \textit{A-} does not represent any particular level of
competence or understanding but rather is a mere indication that
the student outperformed her average peers but was not among
the very top performers in the class. Indeed, such a student may
have been deficient in many respects in her performance on the
exam, but so long as most students were even more deficient, she
is able to earn the label given to those who moderately
outperform, an \textit{A-}. This grade tells an observer nothing about the
student’s mastery or competence in a given topical area, a fact
that is dissonant with our essentialist notions of the level of
quality that an \textit{A-} label represents. The result is that legal
educators—and prospective employers—lack any true measure of
the learning or capabilities of their students.

Proper assessment is about evaluating a student’s
attainment of specified learning objectives.\textsuperscript{387} It involves the
setting of clear goals regarding what students are supposed to
learn and know how to do after completing a course, followed by
the administration of an instrument that measures their
performance against those stated objectives. For example, in a
civil procedure course, a learning objective might be that a
student is able to determine whether a federal court would have
subject matter jurisdiction over various claims that are asserted
in an action. The assessment tool could describe the nature of
certain claims asserted by a set of fictitious parties and ask for an
analysis of each claim, or the exam might ask a student to take a
certain factual situation and make a determination of where to
assert the claims and how to respond to any ensuing

\textsuperscript{386.} See \textsc{Stuckey et al.}, \textit{supra} note 30, at 237 (“The scaled grading system
allows law schools to sort students for legal employers, but it impedes learning,
community building, and moral development.”).

\textsuperscript{387.} See \textit{id.} at 235 (“The main purpose of assessments in educational
institutions is to discover if students have achieved the learning outcomes of the
course studied.” (citation omitted)).
jurisdictional challenges. If the student were able to engage in
that analysis or make those determinations correctly, the student
would be entitled to be judged proficient in that area, having
achieved the learning objective. A standard would have to be set
for what level of success the student must achieve to warrant a
conclusion that he or she was proficient in the area—perhaps
based on a certain percentage of correct responses or, if the
assessment instrument is more oriented toward performance,
there might need to be a more qualitative standard for
proficiency. With each topic or subject matter, professors would
have to make ex ante determinations of what a minimum
acceptable level of performance would be in any given situation;
after such line-drawing, the instructor’s task would be to
determine whether that level was achieved or not.

Ultimately, then, what we know about proper assessment—
that it must be focused on measuring performance against clear
learning objectives—points towards the development of a series of
such objectives within each course, the development of formative
and summative assessment measures that permit students to
demonstrate their attainment of those goals, and the use of
something approaching a binary or three-level\textsuperscript{388} pass-fail system
that reflects achievement rather than relative merit,\textsuperscript{389} where

\begin{itemize}
\item \textsuperscript{388} The Appalachian School of Law uses a pass–fail system that consists of
the designations of “Proficient,” “Not Proficient,” and “Fail.” Clinton W. Shinn,
\textit{Lessening Stress of the 1L Year: Implementing an Alternative to Traditional
Grading}, 41 U. TOL. L. REV. 355, 368 (2010). The problem with that approach is
that the “Not Proficient” category is treated as a passing grade, even though the
student is only marginally competent. Id. Further, this category is hardly a
grade that will put the student in good stead with prospective employers. I
would thus label not-proficient performers as failures; if distinguishing among
passers is desired, “Proficient” and “Highly Proficient” categories might better
serve the students.
\item \textsuperscript{389} There are risks to moving to a pass–fail system—particularly for non-
elite law schools—which include the muddled signals employers receive about
students’ ability without grades. One possibility if a school is committed to using
letter grades as a proxy for the pass-fail categories would be to simplify the
grading system by having only A, B+, and B- grades to reflect exemplary,
proficient, and marginal performances, reserving a C for failures. But it is also
critical that there are ways beyond grades that students are able to demonstrate
their value to prospective employers, such as with portfolios that include work
done by the student and more meaningful faculty evaluations and
recommendations. Of course, evaluating such qualitative materials will take


failure is a real possibility for substandard performances. Rather than a single final essay exam, doctrinal courses should be characterized by a series of opportunities for formative assessment—graded or ungraded—consummated by a summative evaluation designed to mirror more closely what will be expected of a student once he enters practice and confronts such situations. That means that not only is analytical ability and facility with doctrine assessed, but also students’ ability to gather, interpret, and present facts or to assess a situation to determine what action should be taken to achieve a particular goal for one’s client. Quizzes, problems, and other meaningful

more effort and commitment on the part of employers who may be accustomed to relying on grades as simple proxies for merit.

390. See Shinn, supra note 388, at 369 (explaining that at least 15% of students in 1L classes at Appalachian School of Law must receive grades of “Not Proficient” or “Fail,” but that professors are not required to give any grades of “Fail”). Therefore, for this system to be meaningful, a “Fail” or “Not Proficient” grade would have to be real, meaning that professors were willing and able to assign that grade to students whose performance fell short of a predetermined level of sufficiency. Although other disciplines such as medicine exhibit less hesitation in awarding failing grades to students, law schools are institutionally averse to failures, which is likely for at least three reasons.

First, it may be that an F grade is seen as deserved only for a complete lack of knowledge or competence—scoring near zero percent correct if you will—rather than as being appropriate for poor performances that include instances of lucidity—let us say a 60% or 50% performance.

Second, because of the forced mean, if an F grade is given to a student, many very high grades must be given to offset the downward pull such a grade has on the class average. For example, in a class of ten students, it would be mathematically impossible to give more than one student an F grade if the required mean were B+ or a 3.33; even if eight students were given A grades, giving the remaining two students F grades would result in a mean of 3.2 [(8 x 4.0 + 2 x 0)/10 = 3.2].

Finally, failures mean that courses must be repeated, a consequence that law schools are not designed to accommodate in large numbers. Further, were students to regularly be subjected to failing grades across courses, they would have to leave the law school. Such attrition, were it to be of a significant number of students, would not be consistent with the business model of most law schools, which rely on maintaining a certain level of enrollment to support their budgets.

391. See 2011 Proposed Draft, supra note 195, Standard 304 (showing that a revision to the American Bar Association Standards recently under consideration would require law schools to “apply a variety of formative and summative assessment methods across the curriculum to provide meaningful feedback to students”).
assignments along the way will help the professor gauge students’ understanding and progress throughout the term.392

Although achievement-based assessment and grading is preferred, enforced means are an important tool for equalizing grades across courses taught by different professors who may have varying views of what level of performance deserves a given grade. A similar problem could arise under a pass-fail system, where different professors reach different judgments about what level of performance is necessary to demonstrate proficiency in a topical area. However, enforcing a mean across all courses is not the solution to this problem because of the impact such means have on moving assessment away from being an evaluation of achievement against learning goals towards a system that signals relative merit. A more appropriate solution would be to have learning objectives and determinations of proficiency levels centrally or collectively determined.393 That is, a law school would—either through a group of professors394 who teach the same subjects or collaboratively among relevant professors and administrators—be responsible for setting and/or scrutinizing the learning objectives identified by professors and for setting or approving of the standards for proficiency and distinctiveness applied to all assessment instruments, rather than leaving such determinations solely to the discretion of the individual professor. The professor would certainly be the originator of these standards and should have a say in their ultimate form. However, there must be collective agreement across the faculty regarding these matters to ensure that a performance that earns a “pass” in one course would not earn a “fail” in another course. As setting such standards for achievement is an inherently subjective process,

392. For a fuller discussion of more effective assessment methods in legal education, see Schwartz et al., supra note 320, at 135–63; Stuckey et al., supra note 30, at 235–63.

393. See 2011 Proposed Draft, supra note 195, Standard 302 (showing that a revision to the American Bar Association Standards recently under consideration would require law schools to “apply a variety of formative and summative assessment methods across the curriculum to provide meaningful feedback to students”).

394. If this would not be the province of any existing law faculty committee, perhaps a new “assessment” or “institutional standards” committee could be established.
doing so collectively across different types of courses will be challenging. But such an approach is superior to the blunt instrument of an enforced curve because it makes assessment and grading a measure of substantive versus relative achievement, something that is much more meaningful and the true mark of quality assessment in education.

Finally, it must be acknowledged that forced means also are a bulwark against substandard professor instruction, which might result in many students being deficient in their knowledge in part due to the failures of the professor. If a modified pass–fail or simplified letter-grade system is pursued, professors would have to use more formative assessments with feedback to make sure that their students were learning and progressing along the way. They would also need to make sure that the final assessment is properly calibrated to measure what was successfully taught and learned during the course. That said, there would be no guarantee that student performance would collectively rise to the proficient level. Under such circumstances, the professor’s teaching abilities would have to be evaluated but so too would the content and pace of the educational program being delivered to the students to make sure that it is commensurate with their aptitude to learn.

D. Law School Faculty

“How can we expect law students to become competent practitioners if the core of full-time law faculties, notwithstanding its scholarly prowess, does not itself possess even the basic skills required to practice the type of law about which it teaches and writes?”

—Brent E. Newton, 2010

395. Within a particular subject, one could imagine the possibility of having school-wide, standardized exams for the basic required courses that are graded by a common rubric. This would make grade normalization more rational across different sections. However, such an approach would require each professor to cover the same material in the same courses, which would impinge on individual professors’ freedom to select areas of emphasis.

396. Brent E. Newton, Preaching What They Don’t Practice: Why Law Faculties’ Preoccupation with Impractical Scholarship and Devaluation of
As we have seen, Langdell introduced the idea that full-time academics with little or no experience were to be preferred when staffing a law school faculty. More or less true to this vision, contemporary law school faculties are dominated by tenured and tenure-track professors who are less experienced practitioners than they are highly credentialed legal scholars. Traditional doctrinal law faculty currently maintain an obligation to contribute in the areas of teaching, scholarship, and service, carrying a typical teaching load of three to four courses per year and being expected to produce scholarly publications on a regular basis. In return, this category of professors is highly compensated, in an effort to attract the most highly compensated.

---

397. *Supra* notes 103–08 and accompanying text.

398. *See infra* note 407 and accompanying text (providing data on the average practice experience of law professors).


401. *See, e.g.*, *University of Richmond Faculty Handbook*, at Appendix: The T.C. Williams School of Law Personnel Policies and Procedures, Standards and Procedures for Reappointment, Promotion, and Tenure, UNIV. OF RICHMOND, http://facultyhandbook.richmond.edu/Ch_VI/law.html#tenure (last visited Sept. 3, 2012) (“As part of the application for promotion from Associate Professor to Professor and for the award of tenure, the applicant must submit at least four published scholarly works of high quality.”) (on file with the Washington and Lee Law Review).

402. *The Chronicle of Higher Education* reports that for the 2010–2011 school year law professor salaries were the highest of all disciplines, varying positively 59.5% above the average full-time professor of English language and
credentialed and most capable scholars to a school. Why is hiring expensive, inexperienced high-quality legal scholars important to the modern American law school? As the Carnegie Report explains:

Within academic circles, legitimacy and respectability accrued to whatever could be assimilated to the model of formal, science-like discourse . . . . Since the coin of that realm is productivity in scholarship and research, it is not surprising that law schools have increasingly emphasized this dimension of their faculties’ work and identity.403

Certainly, hiring in this manner is instrumental to achieving a mission oriented toward producing legal scholarship, which

---

many, if not most, law schools embrace. More cynically but no less verily, the *U.S. News & World Report* ranking system, with its heavy emphasis on peer reputation, makes it critical that law schools attract and retain a faculty regarded as productive of high quality scholarship if they wish to maintain or enhance their position in these standings.

Are law faculties—as currently constituted—up to the task of delivering the balanced, integrated curriculum suggested by the 2007 *Carnegie Report*? Unfortunately, there are two main problems with relying on traditional law professors of this mold to deliver a revised curriculum. First, as just mentioned, traditional doctrinal professors are not typically hired for their practice experience, of which they tend to have little or

---


407. See Redding, *supra* note 293, at 601 (showing that for those law professors hired between 1996 and 2000, of those with any practice experience (86.6% of the hires), the average number of years of experience was 3.7). A 1991 study found that all law professors at that time had an average of 4.3 years of practice experience, with the experience of hires at law schools ranked in the top 25 being only 1.4 years. See Robert J. Borthwick & Jordan R. Schau, Note, *Gatekeepers of the Profession: An Empirical Profile of the Nation’s Law*
Rather, law professors are hired mostly based on their academic credentials and their promise as legal scholars. A corps of instructors thusly qualified is not ideally suited for delivering a curriculum of expanded practical experiential learning, and arguably is not optimal for practice-oriented, doctrinal instruction, meaning that adjuncts or differently

---

None of the text is repeated.
viewed clinical faculty have to pick up the slack. Second, traditional law faculty members are expensive from the perspective of the law school, as their salaries account for a large share of a law school’s budget and tend to be impervious to dramatic reductions. Transforming law school to a system that involves much greater skills training and a decrease in courses taught through the large-class Socratic case-dialogue method would require much smaller faculty-student ratios, a proposition that would be quite expensive if the hiring of additional traditional faculty were the means undertaken to achieve that goal.

How, then, can law school faculties hope to offer an improved curriculum? Some schools might favor relying on the heavy use of adjunct professors for practice-oriented courses, given the meager wages generally paid to such instructors. This is a sub-optimal


solution, however, because adjuncts are typically unable to give the time and attention necessary to provide high quality experiential instruction to students compared with properly qualified full-time instructors dedicated to such courses. Further, the ABA Standards strongly discourage the use of adjuncts in the first-year curriculum and provide that “full-time faculty shall teach the major portion of the law school’s curriculum,” while also requiring that full-time law school employees be the supervisors of all clinics. That said, adjuncts can and must have some role in delivering experiential learning given the reality of limited resources; the perfect scenario (using all full-time faculty) should not be made the enemy of the good scenario (using adjuncts), and the ABA standards should be revised to permit a greater role for part-time faculty or adjunct instructors. Another possibility would be to expand the ranks of clinical faculty in proportion to the rest of the faculty. Their salaries typically are lower than traditional doctrinal faculty, but their reach in terms of numbers of students is more limited given the small size that clinics must be to be effective.

Ultimately, schools interested in moving their curriculum in a more practice-oriented direction will have to give serious thought to revising their hiring patterns to identify experienced

415. This differential flows largely from the fact that adjuncts will have other full-time employment that will not necessarily accommodate the need to dedicate large amounts of time to supervising law students and they are not physically located at the law school where they would be more readily accessible to their students. See id. Adjuncts may also be less experienced in teaching than full-time faculty and, thus, may do a poorer job, although this is hardly a universally true. Full-time faculty members are not necessarily outstanding classroom teachers guided by the latest research on effective pedagogy nor may they be qualified to guide students through simulations of practice scenarios they have not themselves experienced.


417. Id. Standard 304, Interpretation 304-3(e).

practitioners who have the potential to be great classroom teachers.\textsuperscript{419} Although scholarly potential could remain a factor, exalting that above the ability to deliver practice-relevant training and experiences to students would not be sensible if this type of curricular reform were the goal. To be clear, this is not a goal all law schools may want to embrace; law schools should be free to pursue a variety of missions, including serving as a preeminent center of scholarly research on the law. But for those schools wishing to emphasize a mission of preparing students for practice, hiring differently—meaning a shift toward more experienced lawyers—will have to be part of the equation.\textsuperscript{420}

V. The Next Century in Legal Education

What have we learned? The four pillars of law school education—its curriculum, its pedagogy, its mode of assessment, and its faculty—all have roots in the Langdellian reforms of the late nineteenth century. The justification for the design of the Langdellian law school—that law is a science best learned from studying “original sources” at the feet of masters of learning rather than masters of practice—has been called into question ever since that time, but the basic model has endured. Its resilience seems to be linked to a variety of factors: The consonance of the Langdellian model with faculty backgrounds and aspirations makes it fairly self-perpetuating; the economics of the approach have been heretofore unquestionably superior to more effective alternatives; students and employers have historically been unresponsive to law school curricular reform as

\textsuperscript{419} A 1980 report of the ABA Special Committee for a Study of Legal Education recommended “the law school recruitment process for full-time faculty increasingly look to the practicing segment of the profession for its potential faculty members.” \textit{Am. Bar Ass'n Special Comm. for a Study of Legal Educ. supranote 165, at 105.}

\textsuperscript{420} Jerome Frank made a similar proposal to expand practice-experienced faculty many years ago. \textit{See Frank, Why Not a Clinical Lawyer-School? supranote 204, at 914 (“A considerable proportion of law teachers in any law school should be men with not less than five to ten years of varied experience in the actual practice of law.”).}
they continue to prioritize school prestige—not the quality of training—in making enrollment and hiring decisions; and the deficiencies of the approach were less consequential in a world in which the bar understood and fulfilled its duty to complete the training of lawyers during their first couple years of practice.

Unfortunately, the fraying of the foundation for the justification and perpetuation of the Langdellian model is not likely to usher in fundamental change with ease. Law faculty benefit from the current structure of the course delivery system and may be loath to take on work that will compromise time for other pursuits or impose burdens without increasing compensation. Further, the profile of current law faculty—

421. See NALP, AFTER THE JD, supra note 5, at 79 (reporting survey results that “indicated that two credentials are crucial to finding the first job after law school: the reputation of the school and law school grades”); see also Sloan, supra note 9 (“The most prestigious law schools still dominate when it comes to placing graduates.”); Leichter, supra note 17, (“Why aren’t California’s more price-sensitive firms hiring grads directly from the People’s College of Law in Los Angeles rather than from UCLA . . . ? Firms could do this but instead, they prefer ABA grads from highly regarded law schools.”).

422. See, e.g., MacCrate Report, supra note 22, at 6–7 (“[F]ew employers appear interested in whether students have enrolled in [skills] courses or how they perform in them.”); id. at 7 n.2 (“The American Bar Foundation survey of hiring partners found . . . that this selection of particular courses has little or no impact on hiring decisions.”). This stubborn fact, that legal employers complain about the quality of legal education but do not alter their hiring patterns in response to law school reforms, is a major contributor to the complacency among many legal academics, particularly at elite schools. This commitment to credentials—law school ranking and class ranking—as the determiners of hiring decisions means that doing a better job of training lawyers for practice receives little reward, at least in the short term. As a result, there is less incentive to pursue such reforms, at least among top-tier schools, particularly to the extent they raise the cost of legal education or place cherished school rankings at risk. That said, the competition for students among such schools is fierce, particularly in an environment of declining Law School Admission Test (LSAT) takers. See David Segal, For 2nd Year, a Sharp Drop in Law School Entrance Tests, N.Y. Times, Mar. 19, 2012, at B1 (“In all, the number of test takers has fallen by nearly 25 percent in the last two years.”). To the extent that curricular and other reforms can attract more and better students to one’s school, an increasing number of elite schools may begin to pursue reforms over time.

423. As Upton Sinclair once observed, “It is difficult to get a man to understand something, when his salary depends upon his not understanding it!” UPTON SINCLAIR, I, CANDIDATE FOR GOVERNOR: AND HOW I GOT LICKED 109 (Univ. of Cal. Press ed. 1994); see also David Segal, What They Don’t Teach Law Students: Lawyering, N.Y. Times, Nov. 19, 2011, at A1 (“Professor Rubin failed
having been educated under the Langdellian system and having had little to no practice experience—renders them less sympathetic to the urge toward practice-relevance and less competent to devise and deliver a program with such an orientation.424 This point was apparent to Jerome Frank, who long ago lamented that inexperienced teachers who learned only the law in books—the so-called “book lawyer” or “library-law teacher”—controlled law schools and, thus, could thwart the reform process:

Unfortunately, attempted reform of legal pedagogy is frequently in the hands of the “library-law” teacher. With the best will in the world, such a teacher often finds it almost impossible to warp over the old so-called case-system so as to adapt it to the needs of the future practicing lawyer. For, as above noted, that system is centered in books. So long as teachers who know nothing except what they learned from books under the old case-system are in control of a law school, the actualities of the lawyer’s life are likely in that school to be considered peripheral and as of secondary importance.425

Perhaps focusing on the “law in books” was appropriate in a time when students would go on to learn the “law in action” during the first years of practice; academic legal education was originally meant to precede and supplement law office training, not supplant it.426 Because that tandem relationship between the two spheres has shifted—from formal, to informal, to optional, to nonexistent—law schools must reform the Langdellian model to fill the void.


425. Frank, Why Not a Clinical Lawyer-School?, supra note 204, at 915.

426. See supra notes 57–60 and accompanying text.
What should that reform look like? I have sketched out some thoughts above—that schools should give some consideration to expanding practical skills training,\textsuperscript{427} diversifying pedagogical methods, developing more meaningful assessment techniques, and considering the benefit that more experienced practitioners could bring to a law school faculty—but these and other ideas require much more thorough treatment than I have given them here and will have to await future work.\textsuperscript{428} Further, beyond law school, there are other improvements that need to be made in the areas of pre-law education,\textsuperscript{429} bar admissions standards,\textsuperscript{430} and continuing legal education while in practice.


\textsuperscript{428} That said, the work of the Best Practices Report, the Carnegie Report, and those focusing on pedagogy such as Professors Schwartz, Sparrow, and Hess in \textit{Teaching Law By Design}, provide a solid vision of the direction legal education needs to take going forward. See generally SCHWARTZ ET AL., supra note 320; CARNEGIE REPORT, supra note 22; STUCKEY ET AL., supra note 30. Washington University law professor Brian Tamanaha has also offered some suggestions for the future, including giving schools the flexibility to pursue differing objectives, such as a research-orientation or a practice-orientation, which will result in their faculties having different backgrounds. See BRIAN Z. TAMANAH, FAILING LAW SCHOOLS 173–80 (2012).

\textsuperscript{429} See HARNO, supra note 23, at 127 (“Many of the problems of legal education owe their being to deficiencies in the pre-legal period.”); see also MACCRATE REPORT, supra note 22, at 230 (“Most prelaw counseling takes place only after individuals have already decided to become lawyers . . . . The need for advice at an earlier time in the decision-making process is apparent.”).

\textsuperscript{430} See, e.g., HARNO, supra note 23, at 155 (“[T]he bar examiners [should] introduce examinations, in addition to those they now give, to test applicants for admission on their proficiency in the simpler skills of the profession.”); see also REPORT OF THE TASK FORCE ON THE FUTURE OF THE LEGAL PROFESSION, supra note 21, at 68–69 (recommending that “the New York State Board of Bar Examiners begin assessing professional skills”). California has long had a practice component to its bar exam. See Kristin Booth Glen, \textit{Thinking Out of the Bar Exam Box: A Proposal to “Maccrate” Entry to the Profession}, 23 PACE L. REV. 343, 410–11 (2003) (detailing the parts of the 1980 California Bar Experiment that were adopted as part of the California bar exam). More recently, a Multistate Performance Test (MPT) has been developed as part of the Uniform Bar Exam (UBE), in an effort to universalize performance evaluation in the bar admissions process. See \textit{Uniform Bar Examination (UBE)}, NAT'L. CONF. OF B. EXAMINERS, http://www.ncbex.org/multistate-tests/ube/ (last visited Sept. 9,
But our concern here has been the Langdellian model of law school education. If that model is fundamentally broken—a question that will likely remain the subject of great debate—a then only fundamental change will do, rather than the incremental change we have seen over the past 130 years. Fundamental change means rethinking our categorization of doctrinal subjects and commitment to them as the dominant component of training for legal practice. It means acknowledging what other disciplines seem to know about how people learn, and giving in to the need to bring that knowledge into our own classrooms. It means going beyond traditional classroom dynamics and physical libraries to approaches that leverage technology to deliver content while focusing face time on meaningful discussions and problem solving. It means making the effort to evaluate students against learning objectives in a way that measures and supports their growth and development. It means opening our minds to the notion that using experienced lawyers to educate novice lawyers-in-training is not some radical proposition, but an approach that bears a greater promise of

431. Although there may be clear deficits with contemporary approaches to legal education, the question is whether these fundamental problems adversely affect law school graduates in an enduring manner or do the deficits simply wear off over time as graduates gain experience through legal practice. If the latter is the case, one could argue that this is a short-term problem for each individual that does not warrant an overhaul of how we train lawyers.

432. For example, should we continue to teach contracts, property, torts, criminal law, and civil procedure in the first year or should we be teaching American legal system, introduction to the common law, transactional law, business law concepts, American public law, transnational law, constitutional rights, and civil litigation?

433. For example, having students learn by doing, in context, more so than learning by reading and hearing alone.

434. See generally THOMSON, supra note 7 (discussing how technology can be leveraged to improve legal education); see also Robert Talbert, Thoughts on the Culture of an Inverted Classroom, CASTING OUT NINES (May 25, 2011, 8:23 AM), http://chronicle.com/blognetwork/castingoutnines/2011/05/25/thoughts-on-the-culture-of-an-inverted-classroom/ (last visited Sept. 15, 2012) (discussing “inverted classrooms,” in which lectures are offered online and classroom time is used for discussion, working on problems, and helping individual students) (on file with the Washington and Lee Law Review).
inculcating students with the tools they need for practice. It means freeing law schools to focus on their respective missions and areas of strength, rather than playing to a unitary, Harvard-based model of legal education.\textsuperscript{435} And it means that legal employers—who complain incessantly about the quality of legal education—will have to start putting their jobs where their mouths are, hire based more on the quality of training received than on one’s class rank and school prestige,\textsuperscript{436} and reclaim some responsibility for the continuing education of their new hires.\textsuperscript{437}

\textsuperscript{435} This was one of the key observations of Alfred Reed in his 1921 Carnegie Foundation report on law school. See Reed Report, supra note 1, at 417–18 (“Attempts by each type of law school to carry the entire burden of legal education produce such unsuccessful results as to bring the entire body of practitioners into disrepute.”). Of course, to realize this end, some reform would have to take place in the \textit{U.S. News & World Report} ranking system, which ranks all law schools along a single scale. Perhaps moving toward the approach \textit{U.S. News} takes with undergraduate rankings—categorizing them as national universities, national liberal arts colleges, regional universities, and regional colleges—would be something that could be tried for law schools. Unfortunately, no law school is likely to embrace the label “regional law school” or “local teaching law school.”

\textsuperscript{436} A promising area for future research would be to examine the hiring patterns of employers across and within law schools over time to see if there has been any migration towards job applicants with more extensive practical training; if not, the question of what they truly value in a potential hire—versus what they say they value—arises.

\textsuperscript{437} Something else employers could consider would be to hire new associates at dramatically reduced pay and offer them extensive practical training experiences, akin to the apprenticeship model used in other countries such as Canada (“articling”) and England (“pupilages” for barristers). See supra notes 195–96 and accompanying text. Some American law firms have experimented with this approach, although the model has not become widespread. See Elie Mystel, \textit{Howrey First Years to $100K}, \textit{Above the Law} (June 22, 2009, 3:26 PM), http://abovethelaw.com/2009/06/howrey-first-years-to-100k/ (last visited Sept. 3, 2012) (“[Howrey] is moving to more of an apprenticeship model. New Howrey associates will receive an emphasis on training and take a significant reduction in salary.”) (on file with the Washington and Lee Law Review); see also Elie Mystel, \textit{Salary Cut Watch: Drinker Biddle Cuts Salaries AND Rates}, \textit{Above the Law} (May 11, 2009, 4:22 PM), http://abovethelaw.com/2009/05/salary-cut-watch-drinker-biddle-cuts-salaries-and-rates/ (last visited Sept. 3, 2012) (“Rather than immediately assign the incoming lawyers to client matters, [Drinker Biddle] will enroll its hires in a new training program that will provide courses on taking depositions, writing briefs, and meeting client needs.”) (on file with the Washington and Lee Law Review). Howrey was dissolved on March 15, 2011. See Brian Baxter, \textit{Howrey to Dissolve Effective March 15}, \textit{AmLaw Daily} (Mar. 9, 2011, 5:45 PM),
Many readers will want a bit more specific advice regarding legal education reform than I have given here. I will thus offer some reforms that a law school wanting to do a better job of preparing its students to become practitioners could undertake:

- Modernize the first year to include an introductory overview of the legal system and the legal profession, as well as subjects more pertinent to contemporary legal practice such as transnational law and administrative law;
- Impose a live-client experience requirement, having all students participate in either a clinical course or an externship;
- Extend legal research and writing education into the second year, featuring more extensive simulation training focused on certain areas such as litigation and transactional skills;
- Redesign the content of traditional courses away from an emphasis on cases toward more source material and practice documents, while redesigning the delivery of courses around more group work and problem-solving exercises in the lawyer role during class meetings;
- Hire full-time, part-time, and adjunct faculty who can bring more extensive and contemporary practice experience to bear on the design and delivery of the curriculum;
- Develop capstone courses that enable third-year law students to synthesize their learning across courses and apply it in practice settings.

These are not steps that all law schools must take. Rather, these are simply some possibilities that some schools could consider; there are surely other ways to improve the ability of legal education to prepare students, and schools should be free to

pursue them. Further, while the above reforms may be worthwhile improvements, work remains to be done that can demonstrate their efficacy in better preparing students for practice, at least sufficiently to justify the cost and potential disruption that might accompany some of these efforts.

Unfortunately, several external constraints facing law schools make fundamental reforms difficult to embrace: ABA standards limit the ability to use active practitioners as part-time faculty, require the commitment of extensive resources to physical libraries, and limit the amount of distance education that students can apply toward their degrees; the U.S. News rankings lump all schools into a unitary system that rewards things like expenditures per student, faculty scholarship and prestige, and LSAT scores rather than qualities that relate more directly to a school’s ability to prepare its students for practice; and bar exams continue to focus almost exclusively on substantive knowledge rather than practice competency, a focus that law schools must mirror to some extent if their graduates are to be able to pass the bar. Hopefully over time these and other external factors will evolve in ways that facilitate the more effective legal education that reformers have been urging for the past century.

VI. Conclusion

Traditional legal education remains bound up with many of the fundamental attributes designed by Langdell at Harvard Law School more than a century ago. It is a decidedly academic, or cognitive, model of legal education—centered on legal doctrine and case law—with varying degrees of elective opportunities to attain practical and professional competence. To be truly effective, however, professional legal education must give more attention to transmitting the skills and values that are essential compliments to doctrinal instruction. Mastering the cognitive, practical, and ethical dimensions of legal practice are what professional legal education must be about; focusing largely on the law in books cannot do the job. Perhaps the late (and great) Derrick Bell said it best when he wrote the following:
By... giving priority to "learning by doing" simulations, students mimic the kind of process that an attorney, researching an unfamiliar area of law, might utilize to investigate prior decisions. In practice, lawyers are called to research and to write; to comprehend legal arguments; to guess at the probable effect of and interaction between judicial, statutory, legal and policy arguments in court; to argue, persuade and debate; to work cooperatively with colleagues; and for some, to judge those arguments and decide cases and issues of law. This is as true in the practice of constitutional law as in any other. Once their research skills are in place, most students are aware that they have the capacity to learn, relatively quickly, whatever they need or want to know regarding any legal question.438

This full range of abilities gained through experiential learning is what law schools should strive to deliver if their goal is to produce competent attorneys. Students need to learn how to "work like a lawyer," not just how to "think like a lawyer."439

This has been understood by many since the time of Langdell, as evidenced by the continual criticism emanating from the ABA, the Carnegie Foundation, and legal commentators since the late nineteenth century. What makes change possible now is that the unprecedented confluence of disintermediation440 in the legal profession, the stagnation of incomes in the legal job market, a bubble in law school tuition and attendant student borrowing, and the prospect of a decline in law school applications and enrollments that will require all but perhaps the most elite and secure law schools to innovate or die. I have no doubt that many law professors will react to these admonitions much as most law professors have reacted to previous efforts to


439. Zandt, supra note 31, at 1133–34

The excellent legal analysis and advocacy skills that are the hallmark of law school programs must remain an essential element of legal education, but today’s law students also need a much more sophisticated understanding of what it means to work as well as think like lawyers in their multi-job careers.

440. See SUSSKIND, supra note 12, at 6 (“Lawyers, like the rest of humanity, face the threat of disintermediation (broadly, being cut out of some supply chain) by advanced systems . . . .”).
improve legal education—with denial or sighs of impossibility or indifference—given the many obstacles to reform. It may require bold leadership from deans to make the case for a new vision of legal education and an insistence on the adoption of certain measures, perhaps as a condition of their taking on or continuing to serve in the dean role. Certainly, there may be faculties that take the lead in responding to the need for significant change. However we get there, it is clear that we need to get beyond the Langdellian model toward a truly twenty-first century program of professional legal education that prepares graduates for practice; the time is ripe for getting there if we can all collectively muster the will to take the first steps.

441. See Schwartz, supra note 340, at 360 (“The legal academy’s policies regarding law school hiring, promotion, and tenure practices, law school textbooks, law school accreditation practice, and law school economics have created an environment in which change is very unlikely to occur.”).

NOTES
The Power of a Suggestion: The Use of Forum Selection Clauses by Delaware Corporations

Thomas T. McClendon*

Table of Contents

I. Introduction .......................................................... 2068
II. Problem of Cases Increasingly Decided Outside the State of Incorporation ........................................... 2070
   A. Evidence of Plaintiffs Fleeing Delaware Jurisdiction .............................................................. 2070
   B. Why Are Plaintiffs Fleeing Delaware? ................................................................................. 2074
   C. Is the “Flight from Delaware” Harmful? ........................................................................... 2078
   D. Internal Affairs Doctrine ....................................................................................................... 2081
   E. Forum Non Conveniens ....................................................................................................... 2086
III. Use of Forum Selection Clauses in Governing Documents .......................................................................................................................... 2090
   A. Policy Supporting Forum Selection Clauses ........................................................................ 2090
   B. Legal History of Forum Selection Clauses .......................................................................... 2094
      1. Federal Courts ...................................................................................................................... 2094
      2. Delaware Courts ............................................................................................................... 2098
   C. Enforcement of Forum Selection Clauses .......................................................................... 2100
   D. Previous Use of Forum Selection Clauses by Corporations ............................................. 2105
   E. Content of Forum Selection Clauses .................................................................................. 2110
      1. Claims Covered by a Forum Selection Clause .................................................................. 2110
      2. Mandatory or Elective Provision ..................................................................................... 2113

* Candidate for J.D., Washington and Lee University School of Law, May 2013. The author wishes to thank Professor Christopher M. Bruner for his guidance, encouragement, and helpful comments and criticism to strengthen this Note. The author especially wants to thank his wife Kirsten for her encouragement and understanding through the hours, nights, and weekends that this Note required.
Delaware Vice Chancellor J. Travis Laster suggested in his decision *In re Revlon Inc. Shareholders Litigation*\(^1\) that “if boards of directors and stockholders believe that a particular forum would provide an efficient and value-promoting locus for dispute resolution, then corporations are free to respond with charter provisions selecting an exclusive forum for intra-entity disputes.”\(^2\) Since *Revlon* was decided on March 16, 2010, over 195 Delaware corporations, including 27 in the S&P 500, have followed the Vice Chancellor’s suggestion and adopted a forum selection clause in their governing documents.\(^3\)

Vice Chancellor Laster’s suggestion is a response to the perceived two-pronged problem of (1) duplicative litigation over corporate transactions in Delaware and other forums and (2) other state courts applying Delaware law.\(^4\) From the viewpoint of a Delaware corporation, both prongs are problematic. First, duplicative litigation is more expensive than litigation in a single state and may even result in a split decision.

---

\(^1\) See *In re Revlon Inc. S’holders Litig.*, 990 A.2d 940, 961 (Del. Ch. 2010) (holding that plaintiff’s lead counsel’s efforts to litigate the case had been inadequate and replacing the firm with a new lead counsel).

\(^2\) *Id.* at 960.


\(^4\) While the inverse—Delaware applying another state’s law—is also problematic because Delaware’s corporate law is universally accepted as the most highly developed corporate case law, the application of Delaware’s law is a larger problem than the inverse.
with one state’s court system siding with the corporation and the other state’s court system siding with the plaintiff. Second, other state courts applying Delaware law is—as one litigator has put it—“like taking Gallatorie’s secret recipes and giving them to a Jack-In-The-Box short-order cook. It doesn’t always work so well.” While this statement shows the very high regard in which Delaware lawyers hold Delaware judges, this statement also illustrates the advantage of a system in which only the courts of the state whose law is to be interpreted rule on that law.

This Note will argue that corporations and society in general are detrimentally affected by plaintiffs’ bar filing in multiple jurisdictions against a single action or transaction; that Delaware corporations should enact a forum selection clause to protect themselves from these useless expenses; and that courts should enforce these clauses.

Part II will open by examining the evidence that plaintiffs’ attorneys are choosing to file their cases against Delaware corporations outside of Delaware and will look at several explanations of their motivation in doing so. It will then argue that this trend is detrimental to corporate defendants and society in general. It will then examine the strength and utility of the internal affairs doctrine and forum non conveniens in preventing duplicative litigation.

Part III will propose the use of choice of forum clauses as the best potential solution to the aforementioned problem. This Note will examine the policy underlying choice of forum clauses. It will then examine the legal history of forum selection clauses in federal and Delaware courts. It will then discuss the likelihood of enforcement in both federal and state courts and the arguments


7. Of course, this statement ignores the response that the American judicial system allows the plaintiff to choose where to sue because failure to do so might prevent the plaintiff from having his day in court. This counterargument will be considered later in the Note. See infra Part III.A (discussing the policy behind forum selection clauses).
that should be made in a clause’s favor in different courts. It will look at the previous use of these clauses by corporations, and will discuss the different choices that a corporation must make in drafting a proposed forum selection clause.

II. Problem of Cases Increasingly Decided Outside the State of Incorporation

A. Evidence of Plaintiffs Fleeing Delaware Jurisdiction

Delaware commentators have consistently found that cases involving Delaware corporations are increasingly being decided outside of Delaware courts. A study of venue choice was conducted by John Armour, Bernard Black, and Brian Cheffins, who examined corporate law claims brought against directors of Delaware public corporations that resulted in written decisions from 1995 to 2009 and tracked the percentage of these decisions issued by Delaware courts, other state courts, and federal courts. The study found that the number of written Delaware court opinions remained steady throughout the period, but the percentage of all written opinions in the field that were issued by Delaware courts declined slowly from a high of 80% in 1995 to 65% in 2002. During 2005–2009, however, Delaware’s share had dropped precipitously to an average of 31% of cases involving claims against directors of Delaware public corporations.

---


9. Armour et al., Balancing Act, supra note 8, at 1353. The study included written decisions only because the authors wanted to focus on the significant decisions that would make a difference in the richness and value of Delaware law as precedent. See id. at 1353–54 (stating that written decisions were a “crude” proxy for determining whether Delaware’s rich body of precedent was under threat).

10. Id. at 1354. The study shows a surprising growth of federal court opinions. The authors posit that this growth may have occurred due to the
Mergers and acquisitions (M&A) is one area in which this trend is most pronounced. A study of merger transactions by Brian Quinn revealed the same trend away from Delaware found by Armour, Black, and Cheffins.\(^\text{11}\) Quinn examined 119 merger transactions of more than $100 million that involved solvent Delaware public corporations.\(^\text{12}\) Of 119 transactions, ninety-seven (82%) had subsequent merger-related litigation.\(^\text{13}\) Of those transactions with subsequent litigation, eighty-two (85%) involved multiple lawsuits, with an average of 5.3 per transaction.\(^\text{14}\) In addition, of those transactions that involved subsequent litigation, fifty were litigated in multiple jurisdictions.\(^\text{15}\) Forty percent of the cases were litigated exclusively outside of Delaware.\(^\text{16}\)

Jennifer Johnson conducted a similar study in 2010 by tracking in which forum M&A plaintiffs brought their claims. Johnson found that, of 193 M&A defendants incorporated in Delaware, roughly 40% of cases were filed in Delaware, another 40% were filed in other state courts, and the remaining 20% were filed in federal courts.\(^\text{17}\)

In their study on venue choice, Armour, Black, and Cheffins also examined the twenty-five largest merger transactions each number of federal securities claims and the availability of supplemental jurisdiction that allows plaintiffs to attach these claims to claims over which the federal courts have jurisdiction.

\(^{11}\) Compare id. at 1358 (noting the trend away from Delaware, although expressing uncertainty when it began), with Quinn, supra note 8, at 148 tbl.2 (finding the same trend away from Delaware in public company mergers litigation).

\(^{12}\) Quinn, supra note 8, at 147. The study excluded buybacks, exchange offers, and partial acquisitions. Id. at 147 n.35.

\(^{13}\) Id. at 148 tbl.1.

\(^{14}\) Id. at 148 tbl.2; see also Johnson, supra note 8, at 372 fig.7 (comparing state class action filings in the M&A context with federal class actions in the same context).

\(^{15}\) Quinn, supra note 8, at 148 tbl.2.

\(^{16}\) Id.

\(^{17}\) See Johnson, supra note 8, at 374 fig.8 (finding that, of 193 M&A filings against defendants incorporated in Delaware, 103 were filed in Delaware, 115 were filed in other states, and 47 were filed in federal court). Johnson's study also shows a marked increase in M&A litigation in federal court during 2009 and 2010. See id. at 372 fig.7 (comparing the number of claims brought in state courts against the number of claims brought in federal court).
year during 1994–2009. Of the 400 transactions studied, 256 involved companies incorporated in Delaware, and shareholders filed suit in 121 of 256 transactions (47%). Between 1994 and 2001, litigation contesting these mergers was filed in Delaware an average of 69% of the time. After 2001, the average dropped to 31%.

The study also notes that, prior to 2001, Delaware routinely hosted at least part of the litigation in any given case and was often the exclusive forum. Indeed, in 2001, Delaware provided the exclusive forum for all twenty-five mergers that the study examined during that year. After 2001, however, Delaware became one of many forums, and, in roughly half of the cases studied, plaintiffs did not use Delaware’s courts at all. Although the trend of completely excluding Delaware has shifted back significantly—only 20% of M&A litigation over the twenty-five largest transactions during 2009 was exclusively outside of Delaware—Delaware no longer regularly serves as the exclusive forum. In 2009, none of the twenty-five M&A transaction disputes were litigated exclusively in Delaware, even when many of the transactions included Delaware-incorporated companies.

Leveraged buyouts (LBOs) are another context in which the “flight from Delaware” has been found. Armour, Black, and Cheffins again provided the initial study. They studied 477 LBOs that occurred from 1995 to 2009, of which 300 involved

---

18. Armour et al., Balancing Act, supra note 8, at 1356. Armour, Black, and Cheffins studied the twenty-five largest mergers of publicly held, U.S.-based companies as measured by transaction value. Id.


20. Armour et al., Balancing Act, supra note 8, at 1356.

21. Id.

22. Id. at 1357.

23. Id.

24. Id. at 1358 fig.5.

25. Id. at 1358.

26. Id. at 1358 fig.5.

27. Id.

28. Armour, Black, and Cheffins only considered those public companies that had filings on EDGAR. Id. at 1359.
buyouts of Delaware corporations. Of these 300 LBO transactions, 141 (47%) resulted in litigation. Although the authors note that the frequency of LBOs rose or fell with market cycles, 73% of all LBO-related litigation filed before 2001 was filed in Delaware. From 2002 to 2009, only 45% of LBO-related litigation involving a Delaware corporation was filed in Delaware. For LBO-related litigation, however, the alternative venues were other state courts, not federal courts.

Options backdating litigation formed the third area of study for Armour, Black and Cheffins. They identified 127 Delaware corporations that faced 234 option backdating lawsuits asserting breach of fiduciary duty claims. Of these lawsuits, only 26 (11%) were brought in Delaware, while 115 (49%) lawsuits were brought in federal courts, and another 93 (40%) actions were brought in state courts other than Delaware. Compared to mergers and LBOs, the amount of options litigation in federal court is extremely high, but this is explained by the fact that options backdating cases generally meet federal diversity requirements. The point is, however, that breach of fiduciary

29. Id.
30. Id.
31. Id. at 1360.
32. Id.
33. Id. at 1360 & fig.7.
34. Option backdating occurs when the corporation gives an option to purchase stock to its directors or officers but falsifies records to show that the option was granted earlier when the stock price was lower. By backdating the grant of the option, the corporation did not have to report as compensation the difference between the exercise price (the market value on the date that the corporation supposedly gave the option) and the share's market value on the date when the executive exercises the option. This resulted in improper accounting and tax treatment, which in turn led to false compensation reports. See Christy L. Abbott, Comment, The Shareholder Derivative Suit as a Response to Stock Option Backdating, 53 St. Louis U. L.J. 593, 597 (2009) (defining option backdating); Armour et al., Losing Cases, supra note 8, at 14. Armour, Black and Cheffins state that they chose options backdating because it was derivative litigation, not direct litigation like mergers and LBOs. See Armour et al., Balancing Act, supra note 8, at 1362.
35. Armour et al., Balancing Act, supra note 8, at 1363.
36. Id.
37. While plaintiffs also could bring an options backdating case as a securities class action, the derivative claims—as opposed to the disclosure
duty claims in option backdating cases could have been litigated in Delaware but were not. Like mergers and LBOs, option backdating litigation moved increasingly away from Delaware.

B. Why Are Plaintiffs Fleeing Delaware?

Although both commentators and Delaware judges agree that there is a trend away from Delaware, it is not clear why the trend is occurring. This Note will examine several possible reasons for this trend.

The first theory is that the plaintiffs’ bar believes that Delaware, particularly the Court of Chancery, is too “pro-management” to provide a fair forum. They may base this belief on several opinions from the Court of Chancery, especially those from Chancellor Leo Strine and Vice Chancellor Laster. In his opinion In re Cox Communications Inc. Shareholder Litigation, Chancellor Strine expressed his doubts about the true utility of the plaintiffs’ bar, criticizing its members’ tendency to follow takeover announcements with “hastily-filed, first-day complaints that serve no purpose other than for a particular law firm and its client to get into the medal round of the filing speed (also formerly known as the lead counsel selection) Olympics.”

Vice Chancellor Laster has also expressed frustration with several shareholder suits brought by the plaintiffs’ bar. In Revlon, he took control of the case away from lead counsel, finding that claims under federal law—were more popular because a federal action under 17 C.F.R. § 240.10b-5 must allege a “material” misstatement. Armour et al., Losing Cases, supra note 8, at 15; see also Employment of Manipulative and Deceptive Practices, 17 C.F.R. § 240.10b-5 (1976).

38. Some plaintiffs’ attorneys who responded to Armour, Black and Cheffins admitted that they preferred to avoid Delaware, but these were counterbalanced by those who always filed in Delaware and those who preferred to take a case-by-case approach. See Armour et al., Balancing Act, supra note 8, at 1350.

39. In re Cox Comms., Inc. S’holders Litig., 879 A.2d 604, 642 (Del. Ch. 2005) (holding that the appropriate attorney fee award was $1.275 million rather than $4.95 million as agreed to in the settlement because the attorneys had not sufficiently contributed to the favorable settlement to merit the higher award and the hours worked were excessive).

40. Id. at 608.
the firm had not adequately represented its client. The Vice Chancellor took the opportunity to comment further:

The resulting system involves little real litigation activity, generates questionable benefits for class members, provides transaction-wide releases for defendants, and offers a good living for the traditional plaintiffs’ bar. In a legal system that values representative litigation as a positive force, the business model of filing and free-riding has nothing to commend it.

Laster has maintained this stance in other cases. During a conference for the case *In re Compellent Technologies*, the Vice Chancellor complained to the firms competing to be named lead counsel that “[t]he whole problem is the diversion of interests between entrepreneurial plaintiffs’ counsel and the class. You all maximize by getting the most fee for the least work.” On another occasion, he criticized typical takeover suits, saying “[a] lot of these sue-on-every-deal cases are . . . worthless, they’re simply we see the announcement, then we file,” and even appointed special counsel to investigate possible collusion between plaintiffs and defense counsel in the case.

In a system in which the plaintiffs’ bar depends on contingency fees from cases that the judge may dismiss and settlements that the judge must approve, the use of words such as “questionable benefits” and “free-riding” by the Vice Chancellor must worry the plaintiffs’ bar. It is unlikely, however, that words of frustration from two members of the Chancery would be enough to cause the widespread flight from Delaware, especially when the complaints are limited to takeover cases. While these opinions may be part of the reason why plaintiffs are more reluctant to file in Delaware, something more is at work.

---

41. *In re Revlon Inc. S’holders Litig.*, 990 A.2d 940, 958 (Del. Ch. 2010).
42. *Id.* at 959–60.
43. *In re Compellent Tech., Inc. S’holders Litig.*, No. 6084-VCL, 2011 WL 6382523, at *28 (Del. Ch. Dec. 9, 2011) (finding that plaintiffs’ lawyers gained significant results in settlement that resulted in modification of deal protection provisions and rescission of the Rights Plan and awarding $2.3 million in attorney’s fees, but finding that the supplemental disclosures settlement was not significant and awarding only $100,000 in attorney fees).
A second possibility explaining why plaintiffs are filing in forums other than Delaware is what Chancellor Strine derisively called the “filing speed . . . Olympics” and the selection of lead counsel. In most states, the first firm to file a lawsuit against the defendant corporation regarding the litigated transaction gained an advantage in the race to become lead counsel, with its attendant financial rewards. In Delaware, however, the Chancery has shifted to an approach similar to lead counsel selection in federal securities cases. Former Chancellor William Chandler quantified this approach by identifying three factors that Delaware courts should consider when appointing lead counsel: the quality of the pleadings, the energy and enthusiasm demonstrated by the various attorneys, and the size of the economic stake each plaintiff has in the litigation. Armour, Black, and Cheffins found that this approach resulted in giving priority to firms who have an elite reputation, have successfully represented clients in the past, and currently represent clients with a significant stake in the litigation.

This rule likely supports plaintiffs’ firms mainly based in Delaware at the expense of outside firms that might not be as specialized or experienced in takeover litigation. An outside firm that is in a position to win the race to file is much more likely to file in a state that follows the “filing speed Olympics” than in Delaware, where it would be much less certain—perhaps even unlikely—that the firm would have the reputation, quality, or client to compete with the specialized corporate firms.

A final theory on why plaintiffs’ attorneys are filing outside of Delaware is that Delaware is more likely than other

---

50. See Armour et al., Balancing Act, supra note 8, at 1374 (analyzing Chancellor Chandler’s discussion in TCW Tech for criteria for appointing lead counsel).
jurisdictions to closely scrutinize and slash attorneys' fees when the judge believes such action is appropriate. Delaware courts base attorneys' fees on the relief obtained rather than the hours worked, which is the primary test in most states.\footnote{See Jonathan R. Macey \& Geoffrey P. Miller, Toward an Interest Group Theory of Delaware Corporate Law, 65 Tex. L. Rev. 469, 497 (1987) (describing the approach taken by Delaware on attorneys' fees, as established in Dann v. Chrysler Corp., 215 A.2d 709 (Del. Ch. 1965), aff'd, 223 A.2d 384 (Del. 1966)); see also Sugarland Indus., Inc. v. Thomas, 420 A.2d 142 (Del. 1980) (defining Delaware's approach to attorneys' fees).} The Delaware approach was considered the more generous to the plaintiffs' bar,\footnote{Macey \& Miller, supra note 51, at 497.} but beginning in 2001, Delaware courts have routinely slashed plaintiff-side attorneys' fees in high-profile cases.

In 2001, former Chancellor Chandler cut a $24.75 million fee to $12.3 million after the firm gained a $180 million settlement for shareholders.\footnote{Armour et al., Balancing Act, supra note 8, at 1371.} In 2005, Vice Chancellor Strine in Cox Communications \textit{sua sponte} reduced the attorney fee by 75\%, even though the defendants had not objected to the fee during settlement.\footnote{See \textit{In re Cox Comms.}, Inc. S'holders Litig., 879 A.2d 604, 640–42 (Del. Ch. 2005) (finding that plaintiffs' attorneys took no appreciable risk and did little productive work, and therefore awarded a fee of $1.275 million instead of the requested $4.95 million).} During the same year, Vice Chancellor Stephen Lamb reduced the fee in a takeover case to 28\% of the proposed fee.\footnote{See \textit{In re Instinet Grp. Inc. S'holders Litig.}, No. Civ.A 1289-N, 2005 WL 3501708, at *1 (Del. Ch. Dec. 14, 2005) (reducing the attorneys' fees from $1.623 million to $450,000).} In 2010, Vice Chancellor Donald Parsons, during his decision \textit{In re Cox Radio},\footnote{In re Cox Radio Inc. S'holder Litig., No. 4461-VCP, 2010 WL 1806616, at *23 (Del. Ch. May 6, 2010).} reduced a proposed $3.6 million fee to $490,000,\footnote{\textit{Id.}} and Vice Chancellor Laster in \textit{Brinckerhoff v. Texas Eastern Products Pipeline Co.}\footnote{Brinckerhoff v. Tex. E. Prods. Pipeline Co., 986 A.2d 370, 396 (Del. Ch. 2010) (approving settlement as fair and reasonable, awarding $10 million in fees and expenses to plaintiffs' counsel, and awarding $80,000 in fees and expenses to objectors' counsel).} reduced a proposed $19.5 million fee to $10 million, even though he acknowledged that the firm had expended significant efforts in the litigation.\footnote{See \textit{id.} (acknowledging plaintiffs' counsel's substantial efforts and
Each of these interrelated theories has some merit, and the answer is probably a mixture of each. It is most likely, however, that the “flight from Delaware” is motivated primarily by the reduced attorney fees, especially in cases that the Chancellor and Vice Chancellors are likely to view as weak or frivolous, and therefore likely to award very limited attorney’s fees.

C. Is the “Flight from Delaware” Harmful?

At this point, the reader might naturally feel indifferent as to whether corporate law cases are decided in Delaware, another state, federal courts, or all of the above. Indeed, from a litigation viewpoint, the effect of forum selection seems only to touch the plaintiff and defendant, with the winner at best gaining a favorable jurisdiction in which to bring his case. This hardly seems something for a judge or the legislature to view with alarm. When considered from an economic perspective, however, the view becomes radically different. Most of the defending corporations are public and sell stock to the general public, which means that these shareholders—and thus the general public—profit or suffer losses according to the corporations’ fortunes. While the fact that corporations sell stock to the general public should not insinuate that corporations should be above the

finding that counsel’s efforts created an asset worth $100 million but reducing fee to $10 million). There are signs that the Court of Chancery has realized that this may be a reason why the plaintiffs’ bar has filed in other jurisdictions and is moving to counteract this trend. For example, Chancellor Strine recently awarded attorney fees of $304 million (15%) in a securities derivative suit. In re S. Peru Copper Corp., No. 961-CS, 2011 WL 6382006, at *1 (Del. Ch. Dec. 20, 2011), vacated, 2011 WL 6476919 (Del. Ch. Dec. 22, 2011). Commentators agree that Strine sought to send a message to the plaintiffs’ bar that Delaware is still willing to award large fees for plaintiffs’ lawyers who take large risks that pay off. See Alison Frankel, Record $285 ml Fee Is Strine’s Message to Plaintiffs’ Bar, THOMSON REUTERS NEWS & INSIGHT (Dec. 20, 2011), http://newsandinsight.thomsonreuters.com/Legal/News/ViewNews.aspx?id=35135&terms=@ReutersTopicCodes+CONTAINS'+ANV’ (last visited Nov, 14, 2012) (on file with the Washington and Lee Law Review).

60. This is especially true for Delaware companies, which make up 63% of the Fortune 500 and more than 50% of all U.S. publicly traded companies. Delaware Division of Corporations, Why Choose Delaware as Your Corporate Home?, http://corp.delaware.gov/ (last visited Nov. 14, 2012) (on file with the Washington and Lee Law Review).
law—this would create even greater harm to society—it does mean that courts and legislatures should take action to prevent a corporation from being forced to spend money to defend against duplicative litigation, which requires more money to be spent on litigation to obtain the same result in multiple jurisdictions. As Vice Chancellor Laster noted:

Litigation is costly. So if you could envision this totality of stockholders, they would not want to sue willy-nilly and impose on their company the costs of defending multiple actions in multiple fora, where the cost of briefing on just a motion to dismiss, when you have experienced counsel from big firms, can approach seven figures. It’s real money.61

The additional money spent on litigation would otherwise be available for dividends or improvements to grow the company. In addition, society must be taxed to pay for two courts to listen to the same dispute and perhaps issue divergent opinions, which in turn might produce additional opinions that attempt to consolidate the first two opinions. As a result, society suffers when litigation is not pursued efficiently in a single forum.

Once it is acknowledged that single-forum litigation is the most beneficial to society, the question then becomes: which forum? This Note will argue that the state of incorporation should provide the forum for cases involving its corporations.

First, each state has a competitive advantage over all other states in applying its law. Because of Delaware’s highly developed corporate precedent, Delaware is a good example of a state that has a competitive advantage in applying its own law.62 Ted Mirvis stated:

We can talk . . . in the Delaware courts about p2p and Footnote 10 of Caremark and how Revlon duties intersect with the Unocal scrutiny and the double helix of Siliconix . . . [but for a judge outside of Delaware] it can be a little unnerving. I mean, you say, “Here’s five recent decisions each of which are

94 pages long and you read those five decisions and you can sort of get a basic idea of the vocabulary.  

The Chancellor and Vice Chancellors in the Court of Chancery spend the vast majority of their time deciding corporate law cases, and thus have an overwhelming competitive advantage due to their familiarity with Delaware corporate law cases. Additionally, the court has streamlined procedures that help to provide both plaintiffs and defendants with relatively quick decisions. Most importantly, the court has the advantage of understanding the policies and purposes behind Delaware corporate law because of the court’s in-depth familiarity gained through constant use and application. Out-of-state courts, on the other hand, are not as familiar with Delaware’s corporate law, and, without the advantage of familiarity gained through constant exposure, they are less able to comprehensively grasp the policies and details of judicially defined doctrines that would cause a chancellor or vice chancellor to decide in favor of one party.

The same logic applies with much the same force to other states applying their own corporate law, especially those with specialized courts that gain comprehensive knowledge of the state’s corporate and business law. The nineteen states that

63. Mirvis, supra note 6, at 17.
64. See Allen, supra note 3, at 7 (noting that the Delaware courts have developed considerable experience in dealing with corporate law issues).
66. As one attorney said: “What I tell clients is that even though Delaware law is being applied, when it’s being applied by a bench that doesn’t have as much familiarity with these cases, the predictability goes down.” David Marcus, Did Chancery Fee Rulings Chase Away Plaintiffs Lawyers?, DEL. L. WKLY., Nov. 29, 2006, http://www.delawarerelawweekly.com/news.php?news_id=109, (last visited Nov. 14, 2012) (on file with the Washington and Lee Law Review).
have created business courts in the last two decades share this competitive advantage to a greater or lesser extent, based on the degree that their case law and statutes are highly developed. Even the remaining states where corporate cases are heard by courts of general jurisdiction retain a competitive advantage in deciding their own corporate law, simply by virtue of the fact that they are more likely to hear cases involving that state’s corporation law than out-of-state courts.

Because duplicative litigation wastes society’s resources and the best forum to decide cases involving the corporation’s internal affairs is the state of incorporation, the legislature and judiciary have provided two doctrines that attempt to address the situation. This Note will show, however, that the internal affairs and forum non conveniens doctrines do not adequately address the problem, and will propose that forum selection clauses be utilized by companies and be upheld by the judiciary and legislature as the device that best addresses this problem of waste.

**D. Internal Affairs Doctrine**

Two doctrines govern the choice of law and forum for cases involving the relationship between a corporation and its officers, directors, and shareholders. The internal affairs doctrine is a specific doctrine that applies only to the corporation’s internal affairs, governs the choice of law, and requires that, except in a very unusual case, the law of the state of incorporation applies. 68 Forum non conveniens, the second doctrine, is a broader doctrine that applies in all cases and affects the choice of forum by allowing courts to refuse to accept jurisdiction over a case that is in a proper forum but would be inconvenient, 69 although modern courts have generally regarded this choice as one of convenience and discretion. 70

---

68. See Restatement (Second) of Conflict of Laws § 302(2) (1971) (defining internal affairs doctrine).


70. See Restatement (Second) of Conflict of Laws § 313 (1971), Reporters Note (noting that the modern trend is that “jurisdiction will be exercised unless considerations of convenience or of efficiency or of justice point
The internal affairs doctrine (IAD) is a judicial doctrine that recognizes that ordinarily the state of incorporation's law must be applied to regulate the internal affairs of a corporation, which is defined as the relationship between the corporation and its current officers, directors, and shareholders. The policy behind the doctrine is to promote certainty in the law for a corporation that might otherwise be faced with contradictory demands from the federal government and different state governments. Thus, as a general matter, the law of the state of incorporation will govern cases that involve the internal affairs of a corporation.

The latest cases from the United States Supreme Court on the subject support the assertion that each state has clear and (at least implicitly) exclusive authority over its domestic corporations. For example, in *Cort v. Ash*, Justice Brennan stated: "Corporations are creatures of state law, and investors commit their funds to corporate directors on the understanding that, except where federal law expressly requires certain responsibilities of directors with respect to stockholders, state law will govern the internal affairs of the corporation." A decade to the courts of the state of incorporation as the appropriate tribunals.

71. See *Edgar v. MITE Corp.*, 457 U.S. 624, 645 (1982) (finding that Illinois has an interest in regulating the internal affairs of a domestic corporation, but finding that interest insufficient to allow the state to impose a substantial burden on interstate commerce); see also *Restatement (Second) of Conflict of Laws* § 302(2) (1971) ("The local law of the state of incorporation will be applied to determine such issues, except in the unusual case where, with respect to the particular issue, some other state has a more significant relationship to the occurrence and the parties . . . .").

72. See *Edgar*, 457 U.S. at 645 (noting that without the IAD, a corporation might otherwise be "faced with conflicting demands").

73. See, e.g., *Rogers v. Guar. Trust Co. of N.Y.*, 288 U.S. 123, 133 (1933) ("When, by acquisition of his stock, plaintiff became a member of the corporation he, like every other shareholder, impliedly agreed that, in respect of its internal affairs, the company was to be governed by the laws of the state in which it was organized.").


75. *Id.*
later, in *CTS Corp. v. Dynamics Corp. of America*, Justice Powell reiterated that “[n]o principle of corporate law and practice is more firmly established than a State’s authority to regulate domestic corporations.” Implicit in the Court’s statement is the assertion that it is *not* within other states’ authority to regulate corporations that have been incorporated in other states.

From the standpoint of a corporation seeking certainty in the application of state law, however, there are two problems with the IAD. First, a state court may refuse to follow the IAD in extreme circumstances. For example, the Wisconsin Supreme Court declared in *Beloit Liquidating Trust v. Grade* that it would base a decision whether to apply the IAD on “whether the contacts of one state to the facts of the case are so obviously limited and minimal that application of that state’s law constitutes officious intermeddling.” Courts, however, rarely disregard the IAD and will only do so when a corporation has little to no contact with the state of incorporation other than the actual act of incorporation.

More importantly, the IAD is a choice of law provision, and the doctrine is not applicable in determining forum. This means

76. See *CTS Corp. v. Dynamics Corp. of Am.*, 481 U.S. 69, 89 (1987) (holding the Indiana Control Share Acquisition Act was not preempted and did not violate the Commerce Clause).

77. Id.


79. See *Beloit*, 677 N.W.2d at 307 (Wis. 2004) (holding Wisconsin law applies even to a Delaware corporation due to an absence of contacts between the corporation and Delaware and because Wisconsin has not legislatively or judicially adopted the internal affairs doctrine).

80. Id. at 307.

81. See *Fletcher Cyclopedia of the Law of Corporations* § 4223.50 n.14 (citing only three cases from two jurisdictions for the proposition that not all jurisdictions follow the internal affairs doctrine).

82. This was the case in *Beloit*, for example, in which the court found that the corporation had no contacts with Delaware other than the act of incorporation and filling bankruptcy. *Beloit*, 677 N.W.2d at 307.

83. See *Fletcher Cyclopedia of the Law of Corporations* § 4223.50 (“[T]he internal affairs doctrine is a choice-of-law doctrine and not a bar to
that corporations may be required to defend cases in states other than the state of incorporation, even though that court will apply the law of the state of incorporation.

The United States Supreme Court has often considered the IAD in a forum non conveniens context. An early example is *Rogers v. Guaranty Trust Co. of New York*. The Court in *Rogers* upheld the district court’s dismissal of the case, explaining that the IAD applied and finding that “considerations of convenience, efficiency, and justice” warranted dismissal. The Court, however, did not mandate dismissal in all cases or establish a definitive rule. Instead, it analyzed the defendant company’s activities in New Jersey to ensure that the company had not “organized in that state as a mere matter of convenience for the purpose of carrying on all its business in another state.” In *Rogers*, the Court stressed the comity due to the state of incorporation in a forum non conveniens analysis.

In *Williams v. Green Bay & Western Railroad Co.*, the Court ruled that IAD jurisprudence informed part of the forum non conveniens analysis. Reversing the district court’s dismissal of the case, the Court ruled “the fact that the corporate law of another State is involved does not set the case apart for special treatment.” Instead, the Court noted that the federal court’s application of the law of a state outside its territorial jurisdiction was only one factor that could be considered in a forum non

---

84. See Rogers v. Guar. Trust Co. of New York, 288 U.S. 123, 133 (1933) (holding the district court properly dismissed the case “without prejudice to the enforcement of the rights of plaintiff, if any, in the courts of New Jersey”).

85. See id. (finding the facts of the case clearly fit within the internal affairs of the company and that dismissal was appropriate).

86. See id. (“Obviously, no definite rule of general application can be formulated by which it may be determined under what circumstances a court will assume jurisdiction of stockholders’ suits relating to the conduct of internal affairs of foreign corporations.”).

87. Id.

88. See Stevelman, supra note 19, at 78–79 (arguing the Court applied a conservative approach to comity in *Rogers*).

89. See Williams v. Green Bay & W. R.R. Co., 326 U.S. 549, 553 (1946) (ruling the doctrine of forum non conveniens was to be used as “an instrument of justice” and reversing the case’s dismissal under that doctrine).

90. Id.

91. Id.
THE POWER OF A SUGGESTION

conveniens analysis. In applying this case-by-case analysis designed “as an ‘instrument of justice,’” the Court stressed the ability of the federal court to decide the case and, in this instance, provide the requested relief.

The Delaware Supreme Court also follows the IAD. In McDermott, Inc. v. Lewis, the court concluded that application of the IAD to transfer of stock from a Panamanian corporation to a Delaware subsidiary—and thus application of Panamanian law—was required by due process, the Commerce Clause, and Delaware conflicts principles. In so doing, it rejected Norlin Corp. v. Rooney, Pace Inc., a Second Circuit decision that applied New York law to a Panamanian corporation due to substantial contacts with New York. The Delaware Supreme Court emphasized that the IAD involved “those matters which are peculiar to the relationships among or between the corporation and its current officers, directors, and shareholders” and suggested that the application of the IAD was supported by the Due Process Clause and Commerce Clause of the Constitution.

92. See id. at 554–57 (stating that forum non conveniens will turn on a case-by-case analysis and that there were no special circumstances that should have led the district court to decline to exercise jurisdiction).

93. See id. at 554, 557 (finding that the policy of fairness and justice behind the doctrine of forum non conveniens was not met in the present case).

94. McDermott, Inc. v. Lewis, 531, A.2d 206, 209 (Del. 1987) (holding that Panamanian law governs under the internal affairs doctrine and reaffirming that doctrine as “a major tenet of Delaware corporation law”).

95. Id. at 218.

96. Norlin Corp. v. Rooney, Pace Inc., 744 F.2d 255, 269 (2d Cir. 1984) (affirming the preliminary injunction of a transfer of 49% of stock to an Employee Stock Option Plan and Trust and a plan for the board of directors to vote those shares in anticipation of a hostile takeover attempt).

97. Id. at 261 (holding that New York law expressly applies to the Panamanian corporation and prohibiting the hostile takeover defense mechanism at issue).

98. McDermott, 531 A.2d at 214 (citing Edgar v. MITE Corp., 457 U.S. 624 (1982)).

99. Id. at 218 (finding that the Due Process Clause requires that the corporation’s agents “be given adequate notice of the jurisdiction whose laws will ultimately govern the corporation’s internal affairs,” and the Commerce Clause requires that Delaware law not apply to a Panamanian corporation).
While the doctrine may be important later in determining the enforceability of a forum selection clause, the IAD does not determine which court or courts will hear the case, a situation that leaves corporations vulnerable to identical actions in multiple jurisdictions, even if each court will apply the incorporating state’s law.

E. Forum Non Conveniens

Another option for a corporation seeking to defend against litigation involving its internal affairs in the state of incorporation is a motion for change of venue on the basis of forum non conveniens. A court may not grant a forum non conveniens motion if an adequate alternative forum is not available. The United States Supreme Court in *Gulf Oil Co. v. Gilbert* listed nine factors that are relevant in forum non conveniens analysis: (1) the private interest of the litigant; (2) the relative ease of access to sources of proof; (3) the availability of compulsory process and cost of attendance for witnesses; (4) possibility of view of premises if appropriate, (5) the pendency or nonpendency of a similar action in another jurisdiction; (6) the ability to exercise jurisdiction over all the individual defendants; (7) the relative advantages and obstacles to fair trial; (8) the enforceability of the judgment; and (9) any other practical

---

100. See infra Part III.C (noting that the IAD might be important because it mandates that the court apply the law of the incorporating state in determining whether the forum selection clause is enforceable).

101. Courts may also grant stays of the action or may transfer the case to another jurisdiction under 28 U.S.C. § 1404(a), but both methods incorporate forum non conveniens policy. See, e.g., *Goodman v. Fleischmann*, 364 F. Supp. 1172, 1175 (E.D. Pa. 1973) (ruling that when considering a transfer under 28 U.S.C. § 1404(a), a court must confirm that it has jurisdiction over the case, and must then use the *Gulf Oil Corp.* factors to weigh the parties’ interests to determine whether the transfer furthers the interest and convenience of the parties).

102. See *Dirienzo v. Philip Servs. Corp.*, 294 F.3d 21, 29 (2d Cir. 2002) (“A forum non conveniens motion cannot be granted absent an adequate alternative forum.”).

considerations that make trial of a case easy, expeditious, and inexpensive. The Court also noted a public interest in “having localized controversies decided at home.” In addition, the burden is on the defendant making the motion to show that the factors strongly weigh in favor of a change of venue, and it is within the sound discretion of the court in granting or denying a forum non conveniens motion, which may only be reviewed for abuse of discretion.

A forum non conveniens motion will only succeed if there is an adequate alternative forum and the court is satisfied that the defendant has met its burden in establishing that the balance of the factors weigh in favor of finding heavy hardship to defendants or the court. It would then grant a motion to dismiss, stay, or transfer venue. Although the governing law in cases involving internal affairs of the corporation will be that of the incorporating state under the IAD, other factors may not militate towards venue in the incorporating state. It is not certain, for example, that the proof in the second factor or the individual defendants in the fifth factor will be in the incorporating state. To the contrary, over 50% of all publicly traded companies are incorporated in

---

104. See id. (giving factors to be considered by district court in analysis of forum non conveniens motion).

105. Id.

106. See, e.g., id. ("[U]nless the balance is strongly in favor of the defendant, the plaintiff’s choice of forum should rarely be disturbed."). But see Koster v. Am. Lumbermens Mut. Cas. Co., 330 U.S. 518, 524 (giving less weight to plaintiff's choice of forum in a derivative action in which there may be many different potential plaintiffs in many different fora or in which plaintiff has only a small financial interest and is acting on behalf of a widely scattered group of plaintiffs).

107. See Rogers v. Guar. Trust Co. of N.Y., 288 U.S. 123, 130–31 (1933) (ruling that the district court was “free in the exercise of a sound discretion to decline to pass upon the merits of the controversy and to relegate the plaintiff to an appropriate forum”).

108. See, e.g., Sinochem Int'l Co. v. Malaysia Int'l Shipping Corp., 549 U.S. 422, 430 (2007) (finding that defendant faces a “heavy burden” when advocating dismissal under forum non conveniens); Piper Aircraft Co. v. Reyno, 454 U.S. 235, 249 (1981) (“Under Gilbert, dismissal will ordinarily be appropriate where trial in the plaintiff's chosen forum imposes a heavy burden on the defendant or the court, and where the plaintiff is unable to offer any specific reasons of convenience supporting his choice.”).

109. Id.
Delaware, and these national and worldwide corporations have the majority of their data and employees in states other than Delaware. This does not change the fact, however, that the corporation's internal affairs are governed by the law of the incorporating state, and that each state's courts are best able to interpret its own state law.

Turning specifically to Delaware state law, the Delaware Supreme Court has established two doctrines to aid Delaware courts in ruling on forum non conveniens and stay motions. In *McWane Cast Iron Pipe Corp. v. McDowell-Wellman Engineering Co.*, the court ruled that Delaware courts will not grant a stay “as a matter of right by reason of a prior action pending in another jurisdiction involving the same parties and the same issues.” The court, however, supported a general rule that “litigation should be confined to the forum in which it is first commenced” and that the courts should be wary of defendants seeking to defeat plaintiff's choice of forum by “commencing litigation involving the same cause of action in another jurisdiction of its own choosing.” The court noted that the


112. The Delaware Supreme Court has found no difference between a motion on grounds of forum non conveniens and a motion to stay. *See Gen. Foods Corp. v. Cryo-Maid, Inc.*, 198 A.2d 681, 683 (Del. 1964) (“In principle we can see no difference between a stay based upon similar grounds and an actual dismissal of the action itself.”), *overruled on other grounds by PepsiCo, Inc. v. Pepsi-Cola Bottling Co. of Asbury Park*, 261 A.2d 520 (Del. 1969).

113. *McWane Cast Iron Pipe Corp. v. McDowell-Wellman Eng’g Co.*, 263 A.2d 281, 283 (Del. 1970) (holding that the superior court abused its discretion in refusing to stay or dismiss the action pending in Delaware in light of an Alabama action between the same parties and involving the same issues).

114. *Id.*

115. *Id.*
policy underpinnings of comity and “the necessities of an orderly and efficient administration of justice” support this view.\textsuperscript{116}

If there is no action pending in another jurisdiction, Delaware courts apply \textit{General Foods Corp. v. Cryo-Maid, Inc.},\textsuperscript{117} which is a much tougher standard than \textit{McWane}.\textsuperscript{118} Under \textit{Cryo-Maid}, the moving party must prove “overwhelming hardship”\textsuperscript{119} through any or all of the \textit{Cryo-Maid} factors: “(1) the relative ease of access to proof; (2) the availability of compulsory process for witnesses; (3) the possibility of the view of the premises, if appropriate; (4) all other practical problems that would make the trial of the case easy, expeditious and inexpensive;” and (5) “whether or not the controversy is dependent upon the application of Delaware law which the courts of this state more properly should decide than those of another jurisdiction.”\textsuperscript{120}

The IAD and forum non conveniens motions mandate the use of the incorporating state's law and in some cases dismissal or stay of an action in another jurisdiction in favor of those of the incorporating state. While these are both positive developments, a defendant corporation may face actions in two courts, and must attempt to convince one or the other to grant a forum non conveniens motion to stay or dismiss.\textsuperscript{121} Even if one court is willing to dismiss or stay the case, the corporation must litigate a forum non conveniens motion and pay the associated expenses. Frequently, however, both courts are unwilling to dismiss or stay

\textsuperscript{116} Id.

\textsuperscript{117} See \textit{Cryo-Maid}, 198 A.2d at 684 (affirming the vice chancellor's stay of Delaware action in favor of later-filed Illinois action when factors established favor the later action).


\textsuperscript{120} Id.

\textsuperscript{121} The corporation may be forced to defend the same action in even more than two courts. In M&A transactions, for example, there are an average of four actions for each transaction, with one transaction resulting in forty-one separate suits. Robert B. Thompson & Randall S. Thomas, \textit{The Public and Private Faces of Derivative Lawsuits,} 57 \textit{Vand. L. Rev.} 1747, 1769 (2004).
in favor of the other.\textsuperscript{122} In this case, the corporation faces the unenviable position of defending the same litigation twice or more. The use of a forum selection clause in a corporation’s governing document—if effective—would put an end to this unnecessary litigation. As the United States Supreme Court noted in another context: “[A] clause establishing \textit{ex ante} the forum for dispute resolution has the salutary effect of dispelling any confusion about where suits . . . must be brought and defended.”\textsuperscript{123}

\textit{III. Use of Forum Selection Clauses in Governing Documents}

\textit{A. Policy Supporting Forum Selection Clauses}

The reasons why forum selection clauses are needed have been summarized by former Chancellor Chandler as “the multi-forum deal litigation” problem.\textsuperscript{124} He observed:

\textit{[T]he fallout of [multi-forum deal litigation] has become increasingly problematic in recent years as more and more of these cases are filed in multiple jurisdictions. Judges, defense counsel, and the plaintiffs’ bar are now routinely confronted with these sorts of disputes and have yet to come up with a workable solution. The potential problems, as one can imagine, are numerous. Defense counsel is forced to litigate the same case—often identical claims—in multiple courts. Judicial resources are wasted as judges in two or more jurisdictions review the same documents and at times are asked to decide the exact same motions. Worse still, if a case does not settle or consolidate in one forum, there is the possibility that two judges would apply the law differently or otherwise reach different outcomes, which would then leave the law in a confused state and pose full faith and credit problems for all involved . . . . The problems do not end there. In the event that

\textsuperscript{122} Compare, e.g., \textit{In re Topps Co. S’holders Litig.}, 924 A.2d 951, 954 (Del. Ch. 2007) (denying motion to stay or dismiss because Delaware had a particularly strong interest in addressing new issues in its own law), \textit{with In re Topps Co., Inc. S’holders Litig.}, No. 600715/07, 2007 WL 5018882, at *3 (N.Y. Sup. Ct. June 8, 2007) (denying motion to stay or dismiss because the New York action was the first filed).


defense counsel settles in Delaware over another jurisdiction, leaving one set of plaintiffs’ counsel out in the cold, the unfavored forum’s plaintiffs’ lawyers then often flock to Delaware to oppose the settlement (and vice versa). And there are the post-settlement or post litigation issues as well: class certification, approval of attorneys’ fees, and then dividing those attorneys’ fees between the various plaintiffs’ counsel.125

Chancellor Chandler identifies three main concerns: the litigation cost of defending the same case in multiple jurisdictions, the judicial resources wasted in redundant decisions on the same issue, and the possibility of conflicting decisions on the same issue.126 These concerns apply to all litigation against a corporation, but with different rationales depending on whether the claim is made directly against the company or by derivative action on behalf of the company.

These concerns are at their strongest in M&A litigation, which tends to be highly duplicative. In Cornerstone Research’s study on security class action filings, thirty-eight of the forty M&A actions brought in federal court were also brought in state court.127 The other two actions related to foreign companies.128 One attorney estimates that 50% of all M&A transactions—not just the transactions that result in litigation—generate duplicative litigation.129 In typical M&A litigation, the shareholder bringing the action will usually allege an unfair transaction price or claim that the directors gave shareholders inadequate or misleading information about the proposed transaction.130 In other words, the shareholder is bringing a claim on behalf of the corporation as its agent or representative against the directors who proposed the transaction. Because the corporation is required to indemnify these directors if they are successful in defending themselves and are likely to indemnify

125. Id.
126. Id.
128. Id.
129. Savitt, supra note 5.
130. See Cornerstone Research, supra note 127, at 33 (describing the typical M&A litigation).
them even if they are not, a corporation would never file duplicative litigation when it will likely bear the costs of defending multiple cases on the same issue. While a single fiduciary duty action may increase the transaction’s value to the company by requiring the board to prove the price was the best that it could obtain and that it fairly disclosed the advantages and risks of the transaction to the company, additional actions in other jurisdictions add no marginal value and even present possible holdout problems.

While M&A litigation is the largest trouble spot for duplicative litigation, a corporation will usually be required to indemnify its directors for other duplicative derivative litigation. The corporation will indemnify the successful director for reasonable expenses as long as the action was “by reason of the fact” that the person was a director, which would include

131. Directors and officers who are successful in defending themselves against derivative actions are usually entitled to indemnification as a matter of right in most jurisdictions. See 3 JAMES D. COX & THOMAS LEE HAZEN, TREATISE ON THE LAW OF CORPORATIONS 259 (3d ed. 2010). Even if the director or officer is not successful in defending herself, she will be indemnified for her expenses (but not any fines, judgments, or settlements) if a court or a committee of independent directors determines that the director's or officer's conduct met the standards set forth in the indemnification statute. See id. at 262–63 (citing MBCA § 8.51(a) (2008) and DGCL tit. 8 § 145(a) (2001)). Therefore, if shareholders initiate suits in multiple jurisdictions and the director or officer is successful in defending, the corporation is liable for the additional expense. Even if the shareholder is successful, the corporation is likely to pay the expenses in order to attract skilled directors and officers. The typical corporation will also provide director and officer (D&O) insurance. This insurance consists of two parts: coverage to reimburse the corporation for its costs in indemnifying its directors and officers and coverage that extends directly to the individual officer or director for whom indemnification from the corporation is not available or immediately forthcoming.

132. For example, if a corporation faces two actions regarding a proposed merger and plaintiff A settles with the corporation, plaintiff B may seek a larger settlement than A because B is the only thing standing between the corporation and the merger. The corporation is likely to acquiesce. As Jim Pittinger noted: "Defendants have been settling too many of these suits in order to not hold up the underlying transaction. It’s easy to lose a payoff in the hundreds of thousands, or even low millions, of dollars for plaintiffs’ attorneys’ fees when you have a deal in the billions or tens-of-billions with fees already in the multimillions for assorted lawyers, underwriters, etc." CORNERSTONE RESEARCH, supra note 127, at 33. The absurdity of this reality is apparent when it is noted that an additional action adds no marginal value to the company, even if successful.
fiduciary-duty claims. The unsuccessful director would be entitled to indemnification against a fiduciary duty claim if the court finds the defendant is “fairly and reasonably entitled to indemnification.” Therefore, whether the action is successful or not, the corporation is likely to be required to indemnify its directors and officers for their expenses in defending against litigation brought “by reason of the fact” that the defendant was a director or officer. Again, the corporation would never bring suit against a director or officer in multiple jurisdictions because it would, if unsuccessful, likely be forced to pay the defense’s expenses twice. It is illogical to allow a shareholder to bring the same claim on the corporation's behalf in multiple jurisdictions.

In contrast to derivative litigation, a shareholder pursuing a direct action against the company or its directors or officers acts on her own behalf, and the understanding that the shareholder is acting as a representative of the corporation does not apply. There are reasons, however, why shareholders bringing direct litigation should be bound by a forum selection clause. When the shareholder bought the stock he either had notice of the forum selection clause if it had been adopted or notice that the board of directors had the authority to adopt such a provision. Indeed, the board of directors can only adopt a bylaw including a forum selection clause if the charter gives them this power. All portions of the charter must be approved by the shareholders, either at the founding of the corporation or in an amendment. Thus, the shareholder should be bound by either the prior vote to give the board authority to amend the bylaws—and his decision to buy stock in that company with knowledge of the board’s authority to do so—or by the vote of the shareholders to include a forum selection provision in the charter.

133. Cox & Hazen, supra note 131, at 257–58, 262.
134. Id. at 264.
135. Id. at 257–58.
136. See Del. Code Ann. tit. 8, § 109(a) (2011) (“[A]ny corporation may, in its certificate of incorporation, confer the power to adopt, amend or repeal bylaws upon the directors . . . .”).
137. See id. §§ 102, 242 (defining adoption and amendment of the certificate of incorporation).
A shareholder plaintiff may also argue that he should be entitled to file his claim in any court with jurisdiction because he would, in effect, lose his day in court due to his inability to travel to the selected forum. A California plaintiff, for example, may argue that, if the court enforces a Delaware forum selection clause, he will be unable to litigate his case in Delaware because of an inability to travel to Delaware because of health or travel expenses. Putting aside the ease of interstate travel, the practical realities of corporate litigation ensure that the plaintiff will not be materially hindered in bringing his case. Unlike, for example, a tort claim, a shareholder’s claim is not likely to require testimony from eyewitnesses or the plaintiff himself, but will instead involve facts gleaned from discovery and testimony from experts, all of which are equally available in the state of incorporation as in the plaintiff’s selected forum.

B. Legal History of Forum Selection Clauses

1. Federal Courts

Cases on forum selection clauses in any area of law are sparse. The United States Supreme Court has decided two forum selection cases in admiralty law, while the Delaware Supreme Court’s headline case on forum selection clauses was decided in the limited liability company (LLC) context, and that decision was likely driven by policy considerations unique to that business arrangement.

The United States Supreme Court’s decision in *M/S Bremen v. Zapata Off-Shore Co.* signaled a shift in the federal courts’

("Bylaws ordinarily are binding on the stockholders or members whether they expressly consent to them or not.").


142. *M/S Bremen*, 407 U.S. at 16 (holding that forum selection clauses are valid in international admiralty contracts unless the opposing party can show that enforcement would be unreasonable, unjust, or invalid for fraud or
view on forum selection clauses from disfavor on public policy
grounds to acceptance and acknowledgement that such clauses
were to be enforced unless some reason other than the mere
existence of the clause was found. The Court ruled that forum
selection clauses are prima facie valid and would be enforced
unless the resisting party “could clearly show that enforcement
would be unreasonable and unjust, or that the clause was invalid
for such reasons as fraud or overreaching.”

Because there was no allegation of fraud or overreaching, the
Court only considered whether the clause was unreasonable or
unjust. It noted that some courts have found that a forum
selection clause may be unreasonable if it “is seriously
inconvenient for the trial of action.” The Court noted, however,
that the defendant who was attempting to escape the clause was
party to a freely negotiated contract that bound it to the London
forum as specified in the contract, and questioned how the
defendant could prove unreasonableness when it agreed in a
freely negotiated contract to what it was opposing in court.

In Carnival Cruise Lines, Inc. v. Shute, the Court applied
the M/S Bremen analysis to form contracts that had not been
negotiated. In Carnival Cruise, the company sent tickets that

overreaching).

143. See id. at 12 (finding that the old “provincial attitude” against enforcing
forum selection clauses is outdated and ruling that they should be upheld as a
useful and sometimes necessary addition to business contracts).
144. Id. at 15.
145. See id. (“The correct approach would have been to enforce the forum
clause specifically unless Zapata could clearly show that enforcement would be
unreasonable and unjust, or that the clause was invalid for such reasons as
fraud or overreaching.”).
146. Id. at 16.
147. See id. (“Of course, where it can be said with reasonable assurance that
at the time they entered the contract, the parties to a freely negotiated private
international commercial agreement contemplated the claimed inconvenience, it
is difficult to see why any such claim of inconvenience should be heard to render
the forum clause unenforceable.”).
forum selection clause incorporated into ticket contract enforceable despite lack
of bargaining and alleged inconvenience to the ticketholders).
149. See id. at 593 (noting that “common sense dictates that a ticket of this
kind would be a form contract the terms of which are not subject to
negotiation”).
incorporated by reference an attached notice that all litigation regarding disputes between the passengers and the cruise liner must be brought in the Florida courts. While recognizing that, unlike in *M/S Bremen*, the parties in this case had unequal bargaining power, the Court rejected this distinction as a dividing line for enforcement of forum selection clauses. The Court instead emphasized that forum selection clauses in form contracts should be scrutinized by the judiciary for fundamental fairness. The Court equated fundamental fairness with the lack of bad-faith motives, and cited such factors as Carnival’s principal place of business and that many of its cruises depart from and return to Florida ports as evidence of good faith.

The single case that squarely addresses forum selection clauses in a corporation’s governing documents is *Galaviz v. Berg*, a California case involving Oracle, one of the first corporations to utilize a forum selection clause. The *Galaviz* court applied federal common law, which requires a mutual agreement between the parties, even if the contract was one of adhesion, as in *Carnival Cruise*. In the case at hand, the court highlighted the unilateral amendment of the bylaws by the directors after most of the alleged wrongdoing had occurred.

150. See id. at 588 (quoting contract language).

151. See id. at 593 (“[W]e do not adopt the Court of Appeals’ determination that a nonnegotiated forum-selection clause in a form ticket contract is never enforceable simply because it is not the subject of bargaining.”).

152. Id. at 595.

153. See id. (finding that Carnival has its principal place of business in Florida and many of its cruises depart from and return to Florida ports as evidence of good faith in including a forum selection clause in the ticket contract).


157. Galaviz, 763 F. Supp. 2d at 1174 n.4 (noting that Oracle would not have been able to accomplish a unilateral amendment of the bylaws under contract
“[W]here, as here, the bylaw was adopted by the very individuals who are named as defendants, and after the alleged wrongdoing took place, there is no element of mutual consent to the forum choice at all, at least with respect to shareholders who purchased their shares prior to the time the bylaw was adopted.”158 It distinguished *Carnival Cruise* and its Ninth Circuit progeny, *Argueta v. Banco Mexicano S.A.*,159 by characterizing these cases as “merely giving effect to the bilateral agreement between the parties”160 in the case, while it emphasized that the Oracle forum selection clause had not been agreed to or ratified by the shareholders.161

The *Galaviz* court, however, was willing to enforce forum selection clauses when these clauses qualified as bilateral contracts.162 It noted that if the clause fell under the *Carnival Cruise* and *Argueta* analysis, “there would be little basis to

---

158. Id. at 1171.
159. *Argueta v. Banco Mexicano S.A.*, 87 F.3d 320, 327 (9th Cir. 1996) (upholding forum selection clause). The *Argueta* court distilled three factors from *Carnival Cruise* and *M/S Bremen* that would destroy the enforceability of a contractual forum selection clause:

1. Its incorporation into the contract was the result of fraud, undue influence, or overweening bargaining power; (2) the selected forum is so gravely difficult and inconvenient that the complaining party will for all practical purposes be deprived of its day in court; or (3) enforcement of the clause would contravene a strong public policy of the forum in which the suit is brought.

160. *Galaviz v. Berg*, 763 F. Supp. 2d 1170, 1174 (N.D. Cal. 2011). The court failed to take into account, however, the fact that shareholders authorized the board of directors to pass bylaws. If a forum selection clause in a contract of adhesion, such as that in *Carnival Cruise*, is permissible, it seems unreasonable to refuse to enforce such a clause in a bylaw that is authorized by the shareholders.

161. Id. at 1172. The *Galaviz* court found what it characterized as suspicious facts and, if it had not struck down the clause on a contract theory, may well have found that the facts meet the “fraud, undue influence, or overweening bargaining power” requirement of *Argueta*, 87 F.3d at 327. See *Galaviz*, 763 F. Supp. 2d at 1172 (finding that the board of directors passed the forum selection clause after the vast majority of alleged fraudulent overcharges had occurred).

162. See *Galaviz*, 763 F. Supp. 2d at 1174 (“Were the *Argueta* factors controlling here [because of a bilateral contract], there would be little basis for declining to enforce the venue provision of Oracle’s bylaws.”).
decline to enforce the venue provision.” The court found more persuasive value in a charter amendment—that requires the approval of a majority of shareholders—than a bylaw approved solely by the board of directors.

2. Delaware Courts

Like the federal courts, the Delaware Supreme Court has dealt with forum selection clauses, but not in a corporation's governing documents. The main case decided by the Delaware Supreme Court in this field is *Elf Atochem North America, Inc. v. Jaffari*, in which the court upheld a forum selection clause in a LLC agreement and found that the clause operated to strip jurisdiction from Delaware courts in favor of arbitration in California. While the court’s decision in *Elf Atochem* was largely based on the Delaware LLC Act and its emphasis on freedom of contract between the parties to the LLC agreement, the discussion in *Elf Atochem* on the Court of Chancery’s jurisdiction is relevant to a discussion of forum selection clauses in the corporation context.

In discussing the Delaware LLC Act’s vestment of jurisdiction in the Court of Chancery, the Delaware Supreme Court listed three purposes that this action achieved. The most

163. *Id.*  
164. See *id.* at 1175 (“Certainly were a majority of shareholders to approve such a charter amendment, the arguments for treating the venue provision like those in commercial contracts would be much stronger, even in the case of a plaintiff shareholder who had personally voted against the amendment.”). For in-depth discussion of this point, see *infra* Part III.E.3 (discussing the difference between charter and bylaw clauses).  
165. *Elf Atochem N. Am., Inc. v. Jaffari*, 727 A.2d 286, 291, 295 (Del. 1999) (holding that a forum selection clause mandating arbitration in California was enforceable due to the strong policies supporting “maximum effect to the principle of freedom of contract and to the enforceability of LLC agreements” and “in favor of arbitration”).  
166. *Id.* at 287.  
167. See *id.* at 291 (discussing the emphasis given to parties’ freedom to contract in LLC agreements and holding that parties may contract to avoid Delaware jurisdiction). The Court of Chancery followed *Elf Atochem* in a factually similar case. See *Douzinias v. Am. Bureau of Shipping, Inc.*, 888 A.2d 1146, 1149 (Del. Ch. 2006) (applying choice of law provision to require arbitration in Texas).
important of these for purposes of this Note is the third, which reads: “[Vesting of jurisdiction in the Court of Chancery] tends to center interpretive litigation in Delaware courts with the expectation of uniformity.” This observation illustrates the concern Delaware courts have about the interpretation of Delaware law by other courts, and would hold true in the corporate context as well.

Recent case law demonstrates that forum selection clauses would likely be upheld in corporations as well as LLCs. In *Baker v. Impact Holding, Inc.*, Vice Chancellor Parsons noted that Delaware’s legislature had enacted legislation in response to *Elf Atochem*, effectively banning forum selection clauses that would bind nonmanager members and limited partners in the LLC and limited partnership contexts, respectively. Vice Chancellor Parsons, however, also noted that the legislature did not amend the General Corporation Law in a similar way and upheld the forum selection clause against a former director of the corporation. In addition, as has already been discussed, Vice Chancellor Laster in *Revlon* agreed with Vice Chancellor Parsons in supporting the policy behind forum selection clauses.

---

168. *Elf Atochem*, 727 A.2d at 292. The court also listed as policy reasons favoring this forum selection:

1) it assured that the Court of Chancery has jurisdiction it might not otherwise have because it is a court of limited jurisdiction that requires traditional equitable relief or specific legislation to act, and
2) it established the Court of Chancery as the default forum in the event the members did not provide another choice of forum or dispute resolution mechanism.

*Id.*


170. *Id.* at *2* (citing amended versions of Delaware LLC Act §§ 17-109(d) and 18-109(d)). The legislature likely amended these sections to keep Delaware LLCs and LPs from selecting another state as the exclusive jurisdiction.

171. *Id.*

172. *Id.* at *15.

173. *See In re Revlon, Inc. S'holders Litig.*, 990 A.2d 940, 960–61 (Del. Ch. 2010) (acknowledging that one possible effect of greater judicial oversight of “frequent filers” is that they may file elsewhere, but supporting use of forum selection clause if board of directors and shareholders believe appropriate).
The Delaware Supreme Court has also ruled on forum selection clauses in contracts. In *Ingres Corp. v. CA, Inc.*, the court adopted the presumption of validity instituted by *M/S Bremen* and *Carnival Cruise*. In *Ingres*, the court defined the relationship between the McWane forum non conveniens doctrine and forum selection clauses. The court held that “where contracting parties have a legally enforceable forum selection clause, a court should honor the parties’ contract and enforce the clause, even if, absent any forum selection clause, the McWane principle might otherwise require a different result.” The court also incorporated *M/S Bremen*’s presumption of validity and found that forum selection clauses would be enforced unless the opposing party “clearly show[s] that enforcement would be unreasonable and unjust, or that the clause [is] invalid for such reasons as fraud and overreaching.”

In summary, while Delaware courts have only tangentially touched on forum selection clauses in corporate governing documents—and the Delaware Supreme Court has yet to address them at all—Delaware courts seem sympathetic to corporations utilizing these clauses to defend litigation in the state of incorporation and willing to enforce them.

**C. Enforcement of Forum Selection Clauses**

The arguments on the enforceability of forum selection clauses will vary depending on whether the court is a federal court sitting in diversity or a state court. In ruling on a forum selection clause, a federal court sitting in diversity can take two

---

174. *Ingres Corp. v. CA, Inc.*, 8 A.3d 1143, 1145 (Del. 2010) (holding that forum selection clauses trump the McWane principle).

175. *Id.*

176. *Id.*

approaches. First, it can analyze the clause under federal common law as articulated in *M/S Bremen* and *Carnival Cruise* and followed by *Galaviz*.\(^ {178}\) *Galaviz* declined to follow Delaware corporate law and held that federal common law applied to forum selection clauses because they are procedural in nature and thus governed by federal law.\(^ {179}\) In coming to this conclusion, *Galaviz* followed *Manetti–Farrow, Inc. v. Gucci American, Inc.*\(^ {180}\) a Ninth Circuit case, which held that, under an *Erie–Hanna* analysis,\(^ {181}\) the federal interest in upholding the federal venue rules outweighed the state interest.\(^ {182}\) The Second, Fifth, and Eleventh Circuits have also held that federal law governs forum selection.\(^ {183}\)

If a federal court does find that federal common law controls, a corporation defending a bylaw or charter forum selection clause can argue these cases are a close analogy to the forum selection clause in *Carnival Cruise*. In that case, the Court noted:

> First, a cruise line has a special interest in limiting the fora in which it potentially could be subject to suit. Because a cruise ship typically carries passengers from many locales, it is not unlikely that a mishap on a cruise could subject the cruise line to litigation in several different fora. Additionally, a clause


\(^{179}\) See *Galaviz v. Berg*, 763 F. Supp. 2d 1170, 1171 (N.D. Cal. 2011) (ruling that “the enforceability of a purported venue requirement is a matter of federal common law”).

\(^{180}\) *Manetti–Farrow, Inc. v. Gucci Am., Inc.*, 858 F.2d 509, 512–13 (9th Cir. 1988) (holding that forum selection clause in exclusive dealership contract governed dealer’s tort and contract claims).

\(^{181}\) Under *Erie Railroad Co. v. Tompkins*, 304 U.S. 64 (1938), and *Hanna v. Plumer*, 380 U.S. 460 (1965), courts presiding over a diversity case must decide whether the issue is substantive or procedural in nature. Generally, if the issue is procedural, federal law applies. If the issue is substantive, state law applies. See A. BENJAMIN SPENCER, CIVIL PROCEDURE: A CONTEMPORARY APPROACH 360 (2d ed. 2008).

\(^{182}\) *Manetti–Farrow*, 858 F.2d at 512–13.

establishing ex ante the forum for dispute resolution has the salutary effect of dispelling any confusion about where suits arising from the contract must be brought and defended, sparing litigants the time and expense of pretrial motions to determine the correct forum and conserving judicial resources that otherwise would be devoted to deciding those motions. Finally, it stands to reason that passengers who purchase tickets containing a forum clause like that at issue in this case benefit in the form of reduced fares reflecting the savings that the cruise line enjoys by limiting the fora in which it may be sued.184

Each of these three policy reasons cited by the Court would be satisfied by a typical forum selection clause in a corporation’s governing documents. First, a corporation, like a cruise line, has a special interest in limiting the fora in which it is potentially subject to suit because corporations typically operate in many different states and a single action or transaction has effects spread over many states. Second, because the forum is already specified, both the plaintiffs and the corporation can save litigation expenses and the states can conserve judicial resources. While less obvious, the corporation’s shareholders will benefit under the forum selection clause because the corporation will not be indemnifying or advancing litigation costs to defending directors or officers in multiple jurisdictions.185

Corporations defending a bylaw forum selection clause under federal common law face an uphill battle in proving a bilateral agreement similar to that which was found in *M/S Bremen* and *Carnival Cruise*, especially in bylaw forum selection clauses that are passed by the board of directors. Corporations should stress, however, that shareholders must give the board of directors authorization to pass bylaws,186 and this authorization arguably


185. *See supra* Part II.A (describing the additional, needless costs incurred by the corporation in indemnifying directors or officers for defending against the same claim in multiple jurisdictions).

186. *See, e.g.*, DEL. CODE ANN. § 109(a) (2011) (“[A]ny corporation may, in its certificate of incorporation, confer the power to adopt, amend or repeal bylaws upon the directors.”); MODEL BUS. CORP. ACT § 10.20(b) (2011) (“A corporation’s board of directors may amend or repeal the corporation’s bylaws, unless . . . the shareholders in amending, repealing, or adopting a bylaw expressly provide that the board of directors may not amend, repeal, or reinstate that bylaw.”).
equates to the shareholders’ half of a bilateral agreement that was required in Galaviz. Uncertainty at this point, however, renders this mutual agreement requirement under federal common law the most persuasive argument in favor of charter, as opposed to bylaw, forum selection clauses.

If a defending corporation decides to take a corporate, rather than contractual, approach, Manetti-Farrow and the cases following it have focused on the traditional forum selection clause in a contract, and can be distinguished in the bylaw or charter forum selection clause context. In these cases, a corporation can point to the state’s interest in enforcement of its substantive law governing the relationship between the shareholders, directors, officers, and the corporation itself. A corporation in this position should argue that the balance is shifted in the state’s favor and that state’s law should control. This argument is bolstered by the fact that the Third, Fourth, and Eighth Circuits have all held that state law governs forum selection.

If a federal court follows state law, it would apply the law of the state of incorporation due to the IAD. For Delaware corporations, this would mean Delaware corporate law. As discussed previously, it

187. See Galaviz v. Berg, 763 F. Supp. 2d 1170, 1171 (N.D. Cal. 2011) (finding that the bylaw forum selection clause was a unilateral amendment to the “contract” between the corporation and shareholders and declining to enforce it).

188. See infra Part III.E.3 (comparing the advantages and shortcomings of charter and bylaw forum selection clauses).

189. See, e.g., Manetti–Farrow, Inc. v. Gucci Am., Inc., 858 F.2d 509, 511 (9th Cir. 1988) (noting that the issue in the case was contracts that included identical forum selection clauses, which provided: “For any controversy regarding interpretation or fulfillment of the present contract, the Court of Florence has sole jurisdiction”).


191. See, e.g., CTS Corp. v. Dynamics Corp. of Am., 481 U.S. 69, 89 (1987) (“No principle of corporate law and practice is more firmly established than a State’s authority to regulate domestic corporations.”).
seems likely that Delaware law would enforce charter or bylaw forum selection clauses, although the case law admittedly remains undeveloped.192

If a corporation seeks to enforce a forum selection clause in state court, the decision is simpler. Under the IAD, the court must apply the law of the state of incorporation, which for a Delaware corporation would be Delaware. If a court finds that Delaware law requires dismissal, it will only decline to enforce a forum selection clause if it finds that the state in which the court sits requires that result as a matter of public policy.

A potential example of this scenario might occur under § 2115 of the California Corporations Code, which provides that a corporation, the securities of which are not traded on a national exchange, which transacts more than half of its business in California, and which has more than half of its voting stock held by California residents, is subject to California corporate law.193 Because California courts have applied § 2115 to corporations incorporated outside of California as long as they have sufficient contacts with the state,194 it is possible, even likely, that a California court may refuse to stay or dismiss the action because it believes that California has the largest stake in protecting resident shareholders and employees and that this interest trumps Delaware’s interest in regulating the internal affairs of its corporation.195 However, if California courts take this approach, it

192. See supra Part III.B.2 (discussing Delaware case law on forum selection clauses).


194. See, e.g., State Farm Mut. Auto. Ins. Co. v. Superior Court, 8 Cal. Rptr. 3d 56, 68–69 (recognizing the validity of the internal affairs doctrine, but applying § 2115 to the corporation because of sufficient contacts with California).

195. Id. at 217. Proponents of Section 2115 would point to RESTATEMENT (SECOND) OF CONFLICT OF LAWS § 302(2) (1971), which provides that any state may apply its own law in “the unusual case” in which it has a more significant relationship to the occurrence and to the parties than the incorporating state. These proponents would argue that this is exactly what California’s law accomplishes. The problem with this argument is that use of this exception is extremely rare. See FLETCHER CYCLOPEDIA OF THE LAW OF CORPORATIONS § 4223.50 n.14 (citing only two cases that utilized this exception).
arguably conflicts with Supreme Court precedent and may result in a certiorari grant by the Supreme Court.196

D. Previous Use of Forum Selection Clauses by Corporations

The history of the forum selection clause really begins with Revlon. For the 195 Delaware corporations that have currently adopted or are in the process of adopting choice of forum provisions, 189 (96.9% of all forum selection clauses either adopted or in the process of being adopted) adopted the provision after Revlon.197 By any standard, Revlon has had a substantial effect on corporations and corporate attorneys and heightens the need for clarity in this area.

Of the 192 Delaware corporations, 103 adopted forum selection clauses in connection with an initial public offering, making this by far the single most common scenario.198 Another forty-six clauses (23.6%) were adopted concurrently with other bylaw amendments, usually as part of annual bylaw reviews.199 Sixteen (8.2%) were adopted by board of directors on a stand-alone basis.200 Five (2.6%) adopted forum selection clauses while emerging from bankruptcy protection.201 Ten (5.1%) took advantage of forum selection clauses when reincorporating in Delaware.202

According to Joseph Grundfest, forum selection clauses have evolved in clusters, each based on an initial wording that was copied by “descendants” in that group.203 Grundfest has termed the first attempt at forum selection clauses the Gibson Dunn cluster.204

196. See CTS Corp., 481 U.S. at 89 (“No principle of corporate law and practice is more firmly established than a State’s authority to regulate domestic corporations.”); Grundfest, supra note 62, at 23.
198. Id.
199. Id.
200. Id.
201. Id.
202. Id.
203. See Grundfest, supra note 62, at 3 (noting that forum selection clauses tended to occur in clusters).
204. Id. at 4.
Any action brought by any stockholder against the Corporation or against any officer, director, employee, agent or advisor of the Corporation, including without limitation any such action brought on behalf of the Corporation, shall be brought solely in a court of competent jurisdiction located in the State of Delaware.

This cluster included Standard Pacific, Inc. in 1991, CKE Restaurants, Inc. in 1994, and Kennedy-Wilson, Inc. in 2009. Identical in wording, these early attempts at forum selection limited jurisdiction over shareholder suits against “any officer, director, employee, agent or advisor of the company” to “a court of competent jurisdiction located in the State of Delaware.”


| Oracle, Inc. | The sole and exclusive forum for any actual or purported derivative action brought on behalf of the Corporation shall be the Court of Chancery in the State of Delaware. |
| Netlist, Inc. | The Delaware Chancery Court shall be the sole forum and venue for any lawsuit or legal proceeding by the corporation against any of its directors or officers within the jurisdiction of that court. The state or federal courts located in the State of Delaware shall be the sole forum and venue for any lawsuit or legal proceeding by the corporation against any of its directors or officers not within the jurisdiction of the Delaware Chancery Court. |
| Netsuite, Inc. | Delaware shall be the sole and exclusive forum for (i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of a fiduciary duty owed by any director, officer or other employee of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim arising pursuant to any provision of the DGCL, (iv) or any action asserting a claim governed by the internal affairs doctrine. |

205. Id.

206. Id. (grouping Standard Pacific and CKE Restaurants); Allen 2011, supra note 65, at 7 n.42 (noting that Standard Pacific, CKE Restaurants, and Kennedy-Wilson Holdings have identical forum selection clauses).


Unless the corporation consents in writing to the selection of an alternative forum, the Court of Chancery of the State of Delaware shall be the sole and exclusive forum for (i) any derivative action or proceeding brought on behalf of the corporation, (ii) any action asserting claim of breach of a fiduciary duty owed by any director, officer, or other employee of the corporation to the corporation or the corporation’s stockholders, (iii) any action asserting a claim arising pursuant to any provision of the DGCL, or (iv) any action asserting a claim governed by the internal affairs doctrine. Any person or entity purchasing or otherwise acquiring any interest in shares of capital stock of the corporation shall be deemed to have notice of and consented to the provisions of this Article VII, Paragraph D.

The Oracle clause was the simplest. It encompassed all derivative actions and required these to be litigated in the Delaware Court of Chancery. Netlist was similar, but bound the corporation to bring all actions against its directors and officers in the Delaware Court of Chancery, or, if the Chancery did not have jurisdiction, in Delaware state or federal court. This seems to be a poorly worded option because it is not clear whether Netlist is attempting to follow Oracle in encompassing all derivative litigation or only circumstances in which the corporation itself sues a director or officer.

Netsuite’s and Financial Engines’ clauses were more complex and much broader than the language in either the Gibson Dunn cluster or the Oracle clause. Instead of covering only derivative actions as the previous clauses had done, the Netsuite/Financial Engines language covered both derivative actions and claims in which the Delaware courts would have a competitive advantage over other courts in applying Delaware law. The difference between Netsuite’s and Financial Engines’ language, however, was the choice made between mandatory forum selection in the Netsuite clause and elective forum selection in Financial Engines. The advantages and disadvantages of these options will be evaluated later.

211. See Grundfest, supra note 62, at 16 (noting that “each state has a competitive advantage over other state courts in interpreting its own state’s law”).
212. Compare Netsuite, Inc., Bylaw (Form 8-K, art. 8) (Dec. 26, 2009) (“Delaware shall be the sole and exclusive forum . . . .”), with Financial Engines, Inc., Charter (Form 10-Q, art. D) (Dec. 9, 2009) (“Unless the corporation consents in writing to the selection of an alternative forum, the Court of Chancery of the State of Delaware shall be the sole and exclusive forum . . . .”).
213. See infra Part III.E.2.
After *Revlon* was decided on March 16, 2010, forum selection clauses proliferated. Grundfest divides these clauses into the Skadden, Pillsbury, Simpson, K&E, and Grundfest clusters.

**Figure 3**

| Skadden cluster | Unless the Corporation (through approval of the Board of Directors) consents in writing to the selection of an alternative forum, the Court of Chancery of the State of Delaware shall be the sole and exclusive forum for (i) any actual or purported derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of fiduciary duty owed by any director or officer of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim arising pursuant to any provision of the GCL, (iv) any action asserting a claim governed by the internal affairs doctrine. Any person or entity purchasing or otherwise acquiring any interest in shares of capital stock of the Corporation shall be deemed to have notice of and to have consented to the provisions of this Article FOURTEENTH. |
| Skadden cluster–Swift Transportation exception | Unless the Corporation otherwise consents to an alternative forum in writing, the Court of Chancery of the State of Delaware shall be the sole and exclusive forum for (i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of a fiduciary duty owed by any director or officer of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim against the Corporation arising pursuant to any provision of the DGCL or the Corporation’s Certificate of Incorporation or By-Laws or (iv) any action asserting a claim against the Corporation governed by the internal affairs doctrine. Any person or entity purchasing or otherwise acquiring any interest in shares of capital stock of the Corporation shall be deemed to have notice of and consented to the provisions of this Article THIRTEENTH. |
| Pillsbury cluster | Unless the Corporation consents in writing to the selection of an alternative forum, the Court of Chancery of the State of Delaware shall be the sole and exclusive forum for (i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of a fiduciary duty owed by any director, officer, or other employee of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim arising pursuant to any provision of the Delaware General Corporation Law, or (iv) any action asserting a claim governed by the internal affairs doctrine. Any person or entity purchasing or otherwise acquiring any interest in shares of capital stock of the Corporation shall be deemed to have notice of and consented to the provisions of this Article VII. |

214. See *supra* note 172 and accompanying text.

215. Grundfest, *supra* note 62, at 5. The Grundfest cluster consists exclusively of limited partnerships and limited liability companies, and thus will not be discussed in this Note.
| Simpson cluster | Unless the Corporation consents in writing to the selection of an alternative forum, the Court of Chancery of the State of Delaware shall be the sole and exclusive forum for (i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of a fiduciary duty owed by any director, officer, employee or agent of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim arising pursuant to any provision of the DGCL, or (iv) any action asserting a claim governed by the internal affairs doctrine, in each such case subject to said Court of Chancery having personal jurisdiction over the indispensable parties named as defendants therein. Any person or entity purchasing or otherwise acquiring any interest in shares of capital stock of the corporation shall be deemed to have notice of and consented to the provisions of this Article X. |
| K&E cluster | The Court of Chancery of the State of Delaware shall be the sole and exclusive forum for (i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of a fiduciary duty owed by any director or officer of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim against the Corporation arising pursuant to any provision of the CGL or this Certificate of Incorporation or the Corporation’s Bylaws or (iv) any action asserting a claim against the Corporation governed by the internal affairs doctrine. |

The Skadden cluster mainly tracks the elective Financial Engine language, but with two notable exceptions. Swift Transportation chose to include claims arising under the corporation’s certificate of incorporation and bylaws as well as those arising under the Delaware General Corporation Law (DGCL). All three clauses in the Skadden cluster assert that holders of corporate stock are deemed to have notice of the forum selection provision.

Like the Skadden cluster, the Pillsbury cluster tracks the elective Financial Engines language and adds a notice clause, but unlike the Skadden cluster, does not include claims arising under the certificate of incorporation or bylaws.

---

217. Swift Transp. Co., Charter (Form 10-K, art. 13) (Mar. 29, 2011). This language was not followed by Primerica or Liberty Mutual’s clauses.
218. See, e.g., id.
The Simpson cluster also tracks the elective Financial Engines language and, like both Skadden and Pillsbury, adds a notice clause. In addition, it provides that the forum selection clause shall not be binding unless “the Court of Chancery has personal jurisdiction over the indispensable parties named as defendants therein.”

The K&E cluster follows the mandatory Netsuite language, except that—like the Skadden cluster—it governs actions asserting a claim arising from the corporation’s certification of incorporation or bylaws.

E. Content of Forum Selection Clauses

1. Claims Covered by a Forum Selection Clause

For a company interested in adopting (or a corporate attorney who will be advising clients about) a forum selection clause, the first consideration is the coverage of the clause. As seen in the Oracle, Netlist, Netsuite, and Financial Engines clauses, a forum selection clause can be written to cover directors, employees, shareholders, or any combination of these. Allen identified four categories of actions generally covered by forum selection clauses. These categories overlap to a considerable degree, allowing what Allen characterizes as a “belt and suspenders” approach. The first is any derivative action or proceeding brought on behalf of the corporation. While this would cover the majority of the claims against a corporation, it would arguably not cover nonfiduciary duty claims and would certainly not cover claims asserted against the corporations by

---

221. The Simpson cluster includes LPL Investment Holdings, Inc. and FXCM, Inc. Grundfest, supra note 62, at 4.
222. See, e.g., FXCM, Inc., Charter (Form S-1, § 10.1) (Dec. 1, 2010).
224. See, e.g., Charter Commc’ns, Inc., Charter (Form 8-K, art. 11) (Aug. 20, 2010) (following the Netsuite language, but adding “or this Certificate of Incorporation or the Corporation’s Bylaws” to the third subsection).
225. See supra Figure 2.
226. See Allen, supra note 3, at 4.
227. Id. at 3.
shareholders, directors, or employees individually. Allowing plaintiffs to escape the forum selection clause by characterizing the action as something other than a derivative action seems unwise for a corporation undergoing the trouble to adopt a forum selection clause. While covering derivative actions may be a good start, forum selection clauses should include something more.

The second category covers any action asserting a claim of breach of fiduciary duty owed by any director or officer to the corporation or its stockholders.\footnote{Id.} This is more tailored than the first category and can be expanded to fiduciary duties owed by employees\footnote{See id. (finding 59.5% (116) of forum selection clauses in this category have also addressed fiduciary duties owed by employees).} or agents or advisors.\footnote{See id. (finding that 15.4% (30) of forum selection clauses in this category have also addressed fiduciary duties owed by agents or advisors of the corporation).}

The third and fourth categories are in line with the policy advocated by this Note. The third category covers any action asserting a claim arising pursuant to any provision of the DGCL.\footnote{See id. at 4.} This category aligns with the reality that each state has a competitive advantage in deciding matters involving its own law.\footnote{See Grundfest, supra note 62, at 16 (noting a competitive advantage for each state court in deciding that state’s law).} The fourth category covers any claim governed by the internal affairs doctrine.\footnote{Allen, supra note 3, at 4.} This would have the blanket effect of requiring each state to decide claims governed by its own law and dismiss those that are not.

In addition, some companies have added other limiting factors designed to increase the equity of restricting forum and eliminating cases in which the court may be particularly willing to prevent an unfair forum restriction. This Note will discuss two of the most useful of these provisions.\footnote{The author is aware of three other special provisions, each of which simply follows existing law and is not particularly necessary or helpful. The first is an exception to the application of the Delaware forum selection clause when a federal court has assumed exclusive jurisdiction of a proceeding. Latham & Watkins, supra note 178, at 3. The second provision provides for the Delaware Chancery Court to be the exclusive forum “to the fullest extent” permitted by law. Allen, supra note 3, at 8. The third is a provision providing that any person}
exception to application of the forum selection provision when the court of the chosen forum has determined that an indispensable party is not subject to the jurisdiction of that court. Forty-eight (24.6%) corporations with forum selection clauses included this provision.\(^\text{235}\) This provision helps to prevent clearly unfair situations in which a plaintiff cannot obtain complete redress, and a court might be willing to decline to enforce the forum selection provision in that circumstance because it prevents the plaintiff from having his day in court.\(^\text{236}\) While useful in restricting forum selection clauses to their legal limit, this provision should only encompass those situations in which it is possible to join the indispensable party in another jurisdiction.\(^\text{237}\) Currently, only one corporation has adopted this language.\(^\text{238}\)

Another useful provision would be language disclaiming retroactive application of the forum selection clause to acts or omissions occurring before adoption. This language responds to the *Galaviz* court’s concern about the retroactive effect of Oracle’s

---

\(^{235}\) Allen, *supra* note 3, at 8.

\(^{236}\) See *Argueta v. Banco Mexicano S.A.*, 87 F.3d 320, 325 (9th Cir. 1996) (holding that a court should decline to enforce the forum selection clause if doing so would prevent the plaintiff from having his day in court); see also *Ingres Corp. v. CA, Inc.*, 8 A.3d 1143, 1145 (Del. 2010) (holding that the court will decline to enforce the forum selection clause if doing so would be “unreasonable and unjust”).

\(^{237}\) See Latham & Watkins, *supra* note 178, at 3 (advocating this limitation).

\(^{238}\) See Avid Technology, Inc., Bylaws (Form 8-K, art. 17, § 1) (Oct. 21, 2011) (including an exemption for case in which a indispensable party could not be joined in Delaware or when a federal court has exclusive jurisdiction). The Avid Technology clause reads:

[P]rovided that the foregoing provision shall not apply in the event that (a) the action could not be brought in the Court of Chancery of the State of Delaware because of the inability to join an indispensable party, which party could be joined in the action in another form, or (b) a United States federal court has properly assumed exclusive jurisdiction over the action.

*Id.*
bylaw forum selection clause. Currently, Boeing Company is the only corporation to include this provision, although one other company stated in its amendment proposal that adoption of the forum selection clause would have no effect on pending derivative proceedings.

2. Mandatory or Elective Provision

The second consideration is whether to choose a mandatory or elective provision. A mandatory forum selection clause requires the plaintiff to bring the action in the jurisdiction selected by the corporation, while an elective forum selection clause allows the corporation to consent in writing to the plaintiff’s choice if it sees some advantage to defending in that jurisdiction. The initial forum selection clauses in the Gibson Dunn cluster and by Oracle and Netlist were mandatory, but, starting with Netsuite and Financial Engines, the elective provision became an established option. Currently, 56.1% of Delaware corporations with a forum selection clause specify that the corporation may consent in writing to the selection of an alternative forum. To many corporations, elective forum selection clauses are an attractive option because they offer a “heads I win, tails you lose” alternative. If the selected forum is deemed to be in the company’s best interest, it will enforce the clause. If, however, the forum in which the plaintiff brings the lawsuit is perceived to be more favorable to the defendant

239. See Galaviz v. Berg, 763 F. Supp. 2d 1170, 1171 (N.D. Cal. 2011) (“A bylaw unilaterally adopted by directors, however, stands on a different footing. Particularly where, as here, the bylaw was adopted by the very individuals who are named as defendants, and after the alleged wrongdoing took place . . . .”).

240. The Boeing Co., Inc., Bylaws (Form 8-K, art. VII, § 5) (Oct. 4, 2011) (“With respect to any action arising out of any act or omission occurring after the adoption of this By-Law . . . .”).

241. Allen, supra note 3, at 8 (noting that InsWeb Corporation included this disclosure in its charter amendment proposal).


243. See id. at 5 (distinguishing between mandatory and elective forum selection clauses); see also Allen 2011, supra note 65, at vi (noting that the percentage of elective forum clauses has risen from 30.8% in July 2010 to 56.1% in August 2011).

244. Allen, supra note 3, at 8.
company than the jurisdiction selected by the forum selection clause, the company can consent to the alternative forum in writing and defend there.

While the upside is great for a corporation with an elective forum selection clause, there is some downside risk as well. With Galaviz being the only major case on point outside of Delaware, the enforceability of forum selection clauses in general is uncertain, and a court may be more willing take policy into account when deciding the enforceability of a forum selection clause. When a lawyer seeking to uphold a forum selection clause is required not only to argue the policy behind forum selection clause in general but also to support an elective provision, he may have an impossible task in convincing a judge who sees an elective forum selection clause as too advantageous to the corporation. The plaintiff will argue that an elective forum selection clause allows the corporation to forum shop while preventing the plaintiff–shareholder from doing the same.\textsuperscript{245} While this is not an exceedingly persuasive argument, the advantages of an elective forum selection clause\textsuperscript{246} do not seem to outweigh the risks.

3. Charter or Bylaws

The final decision that an advocate of a forum selection clause must make is whether to place the clause in the corporate charter or bylaws. Charters of Delaware corporations are governed by DGCL § 102. Section 102(b)(1) authorizes corporations to include in their charters:

\begin{quote}
(1) Any provision for the management of the business and for the conduct of the affairs of the corporation, and any provision creating, defining, limiting and regulating the powers of the corporation, the directors, and the stockholders, or any class of
\end{quote}

\textsuperscript{245} See Latham & Watkins, \textit{supra} note 178, at 3 (discussing this argument).

\textsuperscript{246} The advantages of an elective clause seem few. A corporation might be willing to allow the plaintiff to continue the action if the action is in a favorable jurisdiction to the defendant. If, however, the jurisdiction were favorable to the defendant, why would the plaintiff bring the action there? In an adversarial system in which any benefit to the plaintiff is a detriment to the defendant, the benefit of an elective forum selection clause seems minimal.
the stockholders, or the governing body, members, or any class
or group of members of a nonstock corporation; if such
provisions are not contrary to the laws of this State.247

Due to the broad nature of § 102(b)(1), corporations have wide-
ranging authority to insert in the charter of the corporation what
its directors and shareholders deem appropriate, and this would
include inserting a forum selection clause. Charter provisions are
binding on all shareholders, whether or not they voted for the
provisions or whether or not they owned stock prior to the
adoption of the provision.248

In deciding between charter and bylaw provisions, there are
several risks and rewards that a corporation must balance.
Courts agree that charter forum selection provisions, without
exceptional circumstances, are going to be upheld. Even when
striking the forum selection clause at issue, the Galaviz court
noted: “Certainly were a majority of shareholders to approve such
a charter amendment, the arguments for treating the venue
provision like those in commercial contracts would be much
stronger, even in the case of a plaintiff shareholder who had
personally voted against the amendment.”249 Combining this
statement with the Galaviz court’s analysis of the dispute as a
contract case under M/S Bremen and Carnival Cruise, one would
predict that that court would likely have enforced a charter
provision that was approved by the shareholders, even while it
refused to uphold a bylaw provision passed by the board of
directors.

Even Revlon does not expressly mention bylaw forum
selection clauses but speaks of “charter provisions selecting an
exclusive forum for intra-entity disputes.”250 It does, however,
mention Oracle’s bylaw forum selection clause and approve this
possibility, at least by implication.251

248. See Centaur Partners, IV v. Nat’l Intergroup, Inc., 582 A.2d 923, 928
(Del. 1990) ("Corporate charters and by-laws are contracts among the
shareholders of a corporation and the general rules of contract interpretation
are held to apply.").
251. Id. at 960 n.8.
With great reward, however, comes great risk. Shareholders may be unwilling to affirmatively vote in favor of a charter forum selection provision limiting their right to sue in the most advantageous forum available. For example, Allstate Corporation’s recent attempt to place a forum selection clause in its charter gained support from only 41.7% of shareholders.\(^2\) There are several ways that the corporation can persuade reluctant shareholders. The corporation can start by explaining the benefits to shareholders as a class. These include (1) avoiding the possibility of costly duplicative litigation and (2) the possibility of conflicting outcomes in different jurisdictions in the same case, both of which lead to money saved through decreased litigation expenses by the clause (and presumably distributed as dividends or invested back into the company which would increase the stock value). The corporation could also bundle the forum selection clause with shareholder-friendly proposals, as Life Technology Corporation did when it bundled a board declassification amendment with a forum selection clause.\(^3\)

Bylaw forum selection clauses, on the other hand, are relatively easy to adopt as long as the board of directors has the authority to pass bylaws under the charter.\(^4\) As is the case with charters, the DGCL has a sweeping bylaw statute. DGCL § 109(b) states: “The bylaws may contain any provision, not inconsistent with law or with the certificate of incorporation, relating to the business of the corporation, the conduct of its affairs, and its rights or powers or the rights or powers of its stockholders, directors, officers or employees.”\(^5\) In a typical corporation that provides notice that the bylaws may be amended at any time, bylaws can be amended by the board of directors without shareholder consent and are binding on the shareholders.\(^6\) This eliminates the possibility of shareholder rejection that is present in charter proposals, as occurred at

\(^2\) Allen, supra note 3, at 5.

\(^3\) Id.

\(^4\) See Del. Code Ann. tit. 8, § 109(a) (2011) (“[A]ny corporation may, in its certificate of incorporation, confer the power to adopt, amend or repeal bylaws upon the directors . . . .”).

\(^5\) Id.

\(^6\) See CA, Inc. v. AFSCME Emps. Pension Plan, 953 A.2d 227, 234 (Del. 2008) (“Bylaws, by their very nature, set down rules and procedures that bind a corporation’s board and its shareholders.”).
Allstate. As long as a majority of the board of directors agrees on the need to pass a forum selection clause, which would be required to pass a charter provision anyway, the clause will be adopted.

The upside of the bylaw provision is also its downside. In Galaviz, the only case directly on point, the court refused to enforce a board-adopted bylaw forum selection clause and supported its holding by pointing to the unilateral way in which the clause was adopted. Analogizing the clause to a commercial contract, the court noted: “Here, in contrast, the venue provision was unilaterally adopted by the directors who are defendants in this action, after the majority of the purported wrongdoing is alleged to have occurred, and without the consent of existing shareholders who acquired their shares when no such bylaw was in effect.” In distinguishing Carnival Cruise, the court made much of the bilateral agreement that contained the forum selection clause in Carnival Cruise, and emphasized the absence of a bilateral agreement in the case before it. If other courts follow the Galaviz court in applying a contractual analysis to bylaw clauses, they could strike down these clauses with regularity.

While bylaw clauses offer easy adoptability, their enforceability is uncertain with Galaviz as the only case squarely on point. If a court does decline to enforce a corporation’s bylaw clause, the only thing that a corporation gains from the adoption is debts from litigating the disputed clause and additional unfavorable precedent. In addition, the corporation is unlikely to be able to pass a charter forum selection clause when its bylaw clause has been struck down by the courts.

Charter clauses, on the other hand, are much more likely to be enforced by the courts under the contractual approach taken by Galaviz. They are bilateral agreements voted upon by the

257. Allen, supra note 3, at 5.
259. See id. (“Under contract law, a party's consent to a written agreement may serve as consent to all the terms therein, whether or not all of them were specifically negotiated or even read, but it does not follow that a contracting party may thereafter unilaterally add or modify contractual provisions.”).
260. See id.

Modern federal law plainly favors the enforcement of
board of directors (acting for the company) and the shareholders (acting for themselves). A corporation that can fit the adoption of its forum selection clause into a bilateral contract approach will be almost certain to have it upheld under a *Carnival Cruise* analysis.261

4. Proposed Forum Selection Clause

The best forum selection clause depends on the current situation of the corporation adopting one. If the corporation is in transition, such as being on the verge of an IPO, emerging out of bankruptcy, or reincorporating in Delaware, it should adopt a charter amendment that reads:

The Court of Chancery of the State of Delaware shall be262 the sole and exclusive forum for (i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of a fiduciary duty owed by any director or officer of the Corporation to the Corporation or to the Corporation’s stockholders, (iii) any action asserting a claim against the Corporation arising pursuant to any provision of the Delaware General Corporation Law or the Corporation’s Certificate of Incorporation or By-Laws, or (iv) any action asserting a claim against the Corporation governed by the internal affairs doctrine;263 provided that the foregoing provision shall not apply in the event that (a) the action could not be brought in the Court of Chancery of the State of Delaware because of the inability to join an indispensable party, which party could be joined in the action

contractual venue clauses, but even in the case of a form contract, a court merely gives effect to a bilateral agreement between the parties that any disputes they may have arising out of that agreement will be litigated in a particular forum.

261. See *Carnival Cruise Line, Inc. v. Shute*, 499 U.S. 585, 595 (1991). This is assuming, of course, that the corporation does not engage in actions that would cause a court to find the clause void for violating *Carnival Cruise’s* fundamental fairness test.

262. This Note advocates use of a mandatory forum selection clause for reasons discussed earlier in Part III.E.2.

263. This Note advocates use of all four categories discussed earlier in Part III.E.1, including the Swift Transportation addition of “or the Corporation’s Certificate of Incorporation or Bylaws.”
in another forum, or (b) a United States federal court has properly assumed exclusive jurisdiction over the action.264

For a public corporation, there are many possible routes, ranging from adoption by the board of directors of a bylaw forum selection clause to a charter amendment with a forum selection clause. As long as its board of directors has the authority to adopt bylaws,265 it could unilaterally adopt a bylaw forum selection clause, but the board must acknowledge the possibility that this provision may not be enforced by the courts, as happened to Oracle in *Galaviz*. If a corporation chooses this route, it should include the language in Boeing’s clause disclaiming retroactive effect in an effort to avoid a court following *Galaviz* in striking it down as a unilateral amendment to a contract.266

There are two other options for a corporation not interested in the arduous task of passing a charter amendment. First, a board of directors can pass a bylaw forum selection clause and, at the shareholders’ next meeting, propose that the shareholders ratify the directors’ actions. This is similar to a charter amendment, but has the benefit of only requiring a majority of the shareholders present at the meeting rather than a majority of all shares outstanding.267 A ratified bylaw clause would also address the *Galaviz* court’s concern about a bilateral contract.268

---

264. This proposed clause tracks the language of Avid Corporation’s forum selection clause, which is the most comprehensive language of any forum selection clause. See Avid Technology, Inc., Bylaws (Form 8-K, art. 17, § 1) (Oct. 21, 2011).

265. See Del. Code Ann. tit. 8, § 109 (2011) (“[A]ny corporation may, in its certificate of incorporation, confer the power to adopt, amend or repeal bylaws upon the directors.”).

266. See The Boeing Co., Inc., Bylaws (Form 8-K, art. VII, § 5) (Oct. 4, 2011) (“With respect to any action arising out of any act or omission occurring after the adoption of this By-Law . . . .”).

267. Compare Del. Code Ann. tit. 8, § 216 (2011) (stating that default quorum to amend corporation’s bylaws is a simple majority of those present at the shareholders’ meeting and a bylaw is passed with the approval of a majority of those present), with id. § 242 (requiring an affirmative vote of “a majority of the outstanding stock entitled to vote thereon” in order to pass a charter amendment).

268. See Galaviz v. Berg, 763 F. Supp. 2d 1170, 1171 (N.D. Cal. 2011) (“[A] court merely gives effect to a bilateral agreement between the parties that any disputes they may have arising out of that agreement will be litigated in a particular forum.”).
Alternatively, a board of directors can pass a bylaw forum selection clause while asking that the shareholders vote at the next shareholders’ meeting on a proposal to repeal this bylaw. This allows the board to claim the status quo and allows the bylaw clause to proceed without affirmative shareholder action. It is uncertain, however, how the courts would react to what is essentially affirmation by inaction.

A third alternative that the board of directors can take is to propose a shareholder bylaw with a forum selection clause. This takes advantage of the lesser voting requirements under Delaware law for passing a bylaw rather than a charter amendment, but has the contractual element required by Galaviz.

If a corporation is more concerned about enforcement of the forum selection clause than ease of adoptability, it should propose a charter amendment adding a forum selection clause. While the amendment must be approved by a majority of the shares outstanding, the corporation can bundle the forum selection clause with other provisions to encourage adoption. Even without bundling, companies have had a high success rate when they put the issue to the shareholders. If a corporation is able to pass a charter amendment, it is assured of enforcement.

IV. Conclusion

The need for forum selection clauses is driven by the proliferation of duplicative litigation over corporate transactions and alleged breaches of fiduciary duties. It has become increasingly common for a corporation to be required to defend

269. Compare Del. Code Ann. tit. 8, § 216 (2011) (stating that default quorum to amend corporation’s bylaws is a simple majority of those present at the shareholders’ meeting and a bylaw is passed with the approval of a majority of those present), with id. § 242 (requiring an affirmative vote of “a majority of the outstanding stock entitled to vote thereon” in order to pass a charter amendment).

270. See id. § 242 (requiring an affirmative vote of “a majority of the outstanding stock entitled to vote thereon” in order to pass a charter amendment).

271. See Allen, supra note 3, at 5 (noting that, of the six corporations that proposed charter amendments, five succeeded); see also supra notes 252–53 and accompanying text (discussing these six corporations).
several different suits in multiple forums. This costs corporations additional litigation expenses that only benefit the additional lawyers required to litigate the duplicative actions. Not only does this harm corporations, it also harms society in general which purchases stock in the corporation and must pay for the judicial resources required in each action. The IAD and forum non conveniens, while useful in controlling duplicative litigation, do not adequately prevent this problem because neither jurisdiction is required to defer to the other.

This Note has proposed that corporations adopt forum selection clauses in their governing documents. As the Supreme Court noted: forum selection clauses “establishing ex ante the forum for dispute resolution has the salutary effect of dispelling any confusion about where suits . . . must be brought and defended.”272 A corporation adopting a forum selection clause has many choices to make: which claims to include, mandatory or elective selection of forum, insertion into its charter or bylaw, and whether to include exceptions to the general rule in the absence of an indispensible party or for other specific circumstances. Through these choices, a corporation can craft a forum selection clause that best meets the company’s needs. A forum selection clause provides assurance to a corporation that, while it may be required to defend against litigation in one jurisdiction, it will not be forced to defend against the same claim in multiple forums.

Appendix 1: Characteristics of the Prominent Forum Selection Clauses

<table>
<thead>
<tr>
<th>Corporation</th>
<th>Claims Covered by Clause</th>
<th>Mandatory or Elective</th>
<th>Charter or Bylaw</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gibson Dunn cluster</td>
<td>“any action brought by any stockholder against the Corporation or against any office, director, employee, agent or advisor of the Corporation.”</td>
<td>Mandatory</td>
<td>Bylaw</td>
</tr>
<tr>
<td>Oracle, Inc.</td>
<td>“any actual or purported derivative action brought on behalf of the Corporation.”</td>
<td>Mandatory</td>
<td>Bylaw</td>
</tr>
<tr>
<td>Netlist, Inc.</td>
<td>“any lawsuit or legal proceeding by the corporation against any of its directors or officers within the jurisdiction of [the Delaware Chancery Court].”</td>
<td>Mandatory</td>
<td>Bylaw</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corporation</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netsuite, Inc.</td>
<td>&quot;(i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of a fiduciary duty owed by any director, officer or employee of the Corporation to the Corporation or the Corporation's stockholders (iii) any action asserting a claim arising pursuant to any provision of the DGCL, (iv) or any action asserting a claim governed by the internal affairs doctrine.&quot;</td>
</tr>
<tr>
<td>Financial Engines, Inc.</td>
<td>&quot;(i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of a fiduciary duty owed by any director, officer or employee of the Corporation to the Corporation or the Corporation's stockholders (iii) any action asserting a claim arising pursuant to any provision of the DGCL, (iv) or any action asserting a claim governed by the internal affairs doctrine.&quot;</td>
</tr>
<tr>
<td>Skadden cluster</td>
<td>&quot;(i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of a fiduciary duty owed by any director or officer of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim against the Corporation arising pursuant to any provision of the DGCL or the Corporation's Certificate of Incorporation or By-Laws or (iv) any action asserting a claim against the Corporation governed by the internal affairs doctrine.&quot;</td>
</tr>
<tr>
<td>Skadden cluster—Swift Transportation option</td>
<td>&quot;(i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting claim of breach of a fiduciary duty owed by any director or officer of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim against the Corporation arising pursuant to any provision of the DGCL or (iv) any action asserting a claim against the Corporation governed by the internal affairs doctrine.&quot;</td>
</tr>
<tr>
<td>Pillsbury cluster</td>
<td>&quot;(i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting claim of breach of a fiduciary duty owed by any director or officer of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim against the Corporation arising pursuant to any provision of the DGCL or (iv) any action asserting a claim against the Corporation governed by the internal affairs doctrine.&quot;</td>
</tr>
<tr>
<td>Cluster</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Simpson</td>
<td>“(i) any derivative action or proceeding brought on behalf of the Corporation, any action asserting a claim of breach of a fiduciary duty owed by any director, officer or other employee of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim against the Corporation arising pursuant to any provision of the DGCL or the Corporation’s Amendment and Restated Certificate of Incorporation or bylaws or (iv) any action asserting a claim against the Corporation governed by the internal affairs doctrine.”</td>
</tr>
<tr>
<td>K&amp;E</td>
<td>“(i) any derivative action or proceeding brought on behalf of the Corporation, (ii) any action asserting a claim of breach of a fiduciary duty owed by any director or officer of the Corporation to the Corporation or the Corporation’s stockholders, (iii) any action asserting a claim against the Corporation or the Corporation’s stockholders, (iv) any action asserting a claim against the Corporation arising pursuant to any provision of the GCL or this Certificate of Incorporation or the Corporation’s Bylaws or (v) any action asserting a claim against the Corporation governed by the internal affairs doctrine.”</td>
</tr>
</tbody>
</table>
Limiting the Legislative Privilege: Analyzing the Scope of the Speech or Debate Clause

Kelly M. McGuire*

Table of Contents

I. Introduction ............................................................2126

II. Supreme Court Decisions: Defining the Scope of the Speech or Debate Clause .............................................2127
   A. Defining “Speech or Debate” .....................................2128
   B. Defining “shall not be questioned in any other Place” ..............................................................2137

III. The Circuit Split: The Speech or Debate Clause and Its Application to Document Review ......................2140
   A. The D.C. Circuit: United States v. Rayburn
      House Office Building ..........................................2141
   B. The Ninth Circuit: United States v. Renzi ..............2143

IV. Resolving the Split: Narrowing the Scope of the Legislative Privilege ............................................2146
   A. A Look at the Text of the Speech or Debate Clause ..............................................................2146
   B. The Legislative Privilege and the Framers’ Intent ..............................................................2148
   C. The Speech or Debate Clause and the Need for Legislative Accountability ..................................2160

V. Proposed Solution: Limiting the Speech or Debate Clause and Increasing Legislative Accountability ......2166

---

* Candidate for J.D., Washington and Lee University School of Law, May 2013. I would like to thank Professor Russell Miller for his advice and guidance in selecting this topic and throughout the writing process. Also, thank you to the members of the Washington and Lee Law Review Editorial Board for their assistance.
A. Document Nondisclosure and the Scope of the Phrase “shall not be questioned in any other Place” ................................................................. 2166
B. Narrowing the Meaning of Legislative Acts ........ 2167
VI. Conclusion ................................................................. 2168

I. Introduction

Article I, Section Six, Clause One of the American Constitution provides that Senators and Representatives:

shall in all Cases, except Treason, Felony and Breach of the Peace, be privileged from Arrest during their Attendance at the Session of their respective Houses, and in going to and returning from the same; and for any Speech or Debate in either House, they shall not be questioned in any other Place.1

This language, known as the Speech or Debate Clause (the Clause), sets forth the legislative privilege.2 Like the President's executive privilege, the legislative privilege permits legislators to refuse to disclose information protected by the Clause. Since the Clause's inception, the Supreme Court has interpreted it broadly and has continually expanded the breadth of the privilege given to legislators.3 Such interpretations have enabled self-interested legislators to abuse the privilege by shielding their misconduct from the Judiciary, the Executive, and the public.4 Moreover, this improper broadening of the legislative privilege has precluded needed inquiry into legislators' actions.5

The Supreme Court has not indicated whether the privilege granted by the Speech or Debate Clause permits legislators to refuse to disclose documents relating to legislative actions.6 The

2. Id.
3. See infra Part II (discussing the scope of the Speech or Debate Clause as set forth by the Supreme Court).
4. See infra Part IV (discussing legislators' ability to hide misconduct because of the broadening of the Speech or Debate Clause).
5. See infra Part IV.C (discussing the need for legislative accountability).
6. See infra Part III (indicating that the Supreme Court has not determined whether the Speech or Debate Clause includes a document
LIMITING THE LEGISLATIVE PRIVILEGE

Ninth and D.C. Circuits, however, have considered this question.⁷ The D.C. Circuit broadly interpreted the Clause and determined that it privileges legislators’ documents.⁸ In contrast, the Ninth Circuit decided that when the legislative action at issue is not protected by the privilege, the Clause does not prohibit the review of relevant documents referencing legislative acts.⁹

In Part II, this Note explains the current interpretation of the Speech or Debate Clause as set forth by Supreme Court decisions. Part III discusses the D.C. and Ninth Circuit cases creating the circuit split regarding the privilege’s application to document disclosure. Part IV analyzes the Speech or Debate Clause using textual, historical, and ethical constitutional interpretive methods. Finally, Part V proposes a new test for applying the Speech or Debate Clause that will answer the question of document disclosure and narrow the scope of the privilege.

II. Supreme Court Decisions: Defining the Scope of the Speech or Debate Clause

To narrow the scope of the legislative privilege and thereby attempt to curb its abuse, it is important to understand the current interpretation of the Clause as set forth by the Supreme Court. In its numerous decisions addressing the scope of the Speech or Debate Clause,¹⁰ the Court has provided legislators with a very powerful protection.¹¹ For instance, it has indicated

⁷ See infra Part III (discussing the circuit split between the D.C. and Ninth Circuits regarding the scope of the Speech or Debate Clause).
⁸ See United States v. Rayburn House Office Bldg., 497 F.3d 654, 663 (D.C. Cir. 2007) (“Accordingly, we hold that a search that allows agents of the Executive to review privileged materials without the Member’s consent violates the Clause.”).
⁹ See United States v. Renzi, 651 F.3d 1012, 1039 (9th Cir. 2011) (stating that “the alleged choices and actions for which [Renzi was] prosecuted [lay] beyond [the] limits” of the Speech or Debate Clause and declining to find a document “non-disclosure privilege”).
¹⁰ See infra Part II.A–B (discussing Supreme Court decisions defining the Speech or Debate Clause’s scope).
¹¹ See, e.g., Eastland v. U.S. Servicemen’s Fund, 421 U.S. 491, 501–03
that, when the privilege applies, it applies absolutely. The Court has also stated that the Clause must be “read broadly to effectuate its purposes.”

In addition to giving legislators a robust privilege, throughout its decisions, the Supreme Court has interpreted the two distinct portions of the Speech or Debate Clause: (1) “Speech or Debate in either House” and (2) “shall not be questioned in any other Place.”

A. Defining “Speech or Debate”

In Kilbourn v. Thompson, the first Supreme Court case considering the Speech or Debate Clause, the Court rejected the narrowest interpretation of the phrase “speech or debate.” Rather than limiting the Clause to its literal meaning, the Court concluded that “speech or debate” also includes “things generally done in a session of the House by one of its members in relation to the business before it.” For example, the Court explained that written reports, resolutions, and voting all constitute privileged “speech or debate.”

(1975) (discussing the powerful privilege granted by the Speech or Debate Clause); United States v. Johnson, 383 U.S. 169, 180 (1966) (same).

12. See, e.g., Eastland, 421 U.S. at 501 (indicating that “the prohibitions of the Speech or Debate Clause are absolute” (citations omitted)).


15. See Kilbourn v. Thompson, 103 U.S. 168, 204–05 (1880) (holding that a resolution by defendant members of the House of Representatives was privileged by the Speech or Debate Clause).

16. See id. at 204 (explaining that a narrow construction of the Clause, which would “limit [the privilege] to words spoken in debate,” would not be adopted).


18. See Kilbourn, 103 U.S. at 204 (“The reason of the [Clause] is as forcible in its application to written reports presented in that body by its committees, to resolutions offered, which, though in writing, must be reproduced in speech, and to the act of voting, whether it is done vocally or by passing between the tellers.”).
In *Tenney v. Brandhove*, the Court followed the expansive interpretation of the phrase “speech or debate” set forth in *Kilbourn*. The *Tenney* Court determined that an investigation by a legislative committee was within “the sphere of legitimate legislative activity” covered by the privilege. The Court explained that a broad interpretation of the Clause is necessary to enable legislators to fulfill their roles as lawmakers. In reaching its conclusion the Court asserted that to “exceed[] the bounds of legislative power it must be obvious that there was a usurpation of functions exclusively vested in the Judiciary or the

---

19. See *Tenney v. Brandhove*, 341 U.S. 367, 378–79 (1951) (holding that a committee investigation was protected by the legislative privilege because it did not clearly impinge upon the powers of another branch of the federal government). The Court considered whether defendant legislators could be subject to civil liability for requiring plaintiff Brandhove to appear at a committee hearing. *Id.* at 370–72. A committee of California legislators, “the Senate Fact-Finding Committee on Unamerican Activities,” summoned Brandhove to appear before it at a committee hearing. *Id.* at 369–70. Brandhove appeared, but refused to testify. *Id.* at 370. Consequently, the state courts held him in contempt. *Id.* at 371. Brandhove then brought a claim seeking damages in which he asserted that the committee hearing “was not held for a legislative purpose.” *Id.* He argued that the hearing was instead held to prevent him from “effectively exercising his constitutional rights.” *Id.* The Court first explained that “acts done within the sphere of legislative activity” could not subject legislators to civil liability. *Id.* at 376. The Court then considered whether the committee action constituted a privileged legislative act. *Id.* The Court stated that in order for legislators to fulfill their role as lawmakers, “the cost and inconvenience and distractions of a trial” must be avoided. *Id.* at 377. The Court concluded that a broad privilege is necessary to accomplish this goal. *Id.* Additionally, the Court explained that the voting process rather than the courts should be used to correct legislative abuse. *Id.* at 378. Finally, the Court stated that for legislative action to be outside the privilege, it must clearly encroach upon the power of the Executive or the Judiciary. *Id.* The Court found that the committee hearing did not reach this high standard, and the action was protected from civil suit by the legislative privilege. *Id.* at 378–79.

20. See *id.* at 378 (stating that “[t]o find that a [legislative action] has exceeded the bounds of legislative power it must be obvious that there was a usurpation of functions exclusively vested in the Judiciary or the Executive”); see also *Kilbourn*, 103 U.S. at 204 (stating that the legislative privilege extends to “things generally done in a session of the House by one of its members in relation to the business before it”).


22. See *id.* at 377 (“Legislators are immune from deterrents to the uninhibited discharge of their legislative duty, not for their private indulgence but for the public good.”).
Executive.” Moreover, the case made clear that a mere allegation of a legislator's improper motive does not terminate the privilege.

In spite of the expansive definition of “speech or debate,” the Court has indicated that there is a limit to which actions by legislators will qualify as privileged legislative acts. For example, in *Gravel v. United States*, although the Court agreed with prior cases stating that the privilege's scope is not limited to literal speech and debate, it emphasized that “[l]egislative acts are not all-encompassing.” *Gravel* provided

---

23. Id. at 378.
24. See id. at 377 (“The claim of an unworthy purpose does not destroy the privilege.”).
25. See, e.g., *Gravel v. United States*, 408 U.S. 606, 625 (1972) (“Legislative acts are not all-encompassing.”); *Tenney v. Brandhove*, 341 U.S. 367, 377 (1951) (“This Court has not hesitated to sustain the rights of private individuals when it found Congress was acting outside its legislative role.” (citations omitted)).
26. See *Gravel v. United States*, 408 U.S. 606, 618, 624–25 (1972) (holding that although the legislative privilege “applies not only to a Member but also to his aides,” the privilege does not extend to actions that do not fall within the legitimate legislative sphere). The Court considered whether a subpoena requiring Senator Gravel's aide, Rodberg, to testify in an investigation into possible criminal conduct relating to the disclosure and publication of top-secret national defense information, known as the Pentagon Papers, violated the legislative privilege. Id. at 608–09. Senator Gravel, as chairman, called an evening meeting of the Subcommittee on Buildings and Grounds of the Senate Public Works Committee. Id. at 609. At the meeting, he read large portions of the Pentagon Papers to the subcommittee and had the entire forty-seven volumes of the papers placed into the public record. Id. Rodberg helped Senator Gravel prepare for and hold the subcommittee meeting. Id. A few weeks later, Senator Gravel had the papers published by Beacon Press. Id. at 609–10. First, the Court explained that for legislators to be adequately protected from the Executive and Judiciary, the legislative privilege must shield legislators' aides from questioning. Id. at 616–18. The Court then determined the extent to which the Clause protected Senator Gravel from inquiry regarding the crime. Id. at 622. The Court discussed the history of the Clause and noted that “the English legislative privilege was not viewed as protecting republication of an otherwise immune libel on the floor of the House.” Id. Similarly, the Court concluded that although the Senator could not be questioned about the subcommittee meeting itself, his involvement in the publication of the papers was not a privileged legislative act. Id. at 616, 626. Thus, the Court held that the Speech or Debate Clause did not protect Rodberg from testifying in the criminal investigation about his own involvement and that of Senator Gravel in arranging the publication of the Pentagon Papers. Id. at 626–27.
27. See id. at 624–25.
LIMITING THE LEGISLATIVE PRIVILEGE

a test for determining which actions beyond literal speech and debate can be classified as immune legislative acts. To be privileged under the Gravel standard, a legislator's action must be “an integral part of the deliberative and communicative processes by which Members participate in committee and House proceedings with respect to the consideration and passage or rejection of proposed legislation or with respect to other matters which the Constitution places within the jurisdiction of either House.” This standard vastly decreases the number of activities that will be deemed immune legislative acts under the Clause.

Additionally, Gravel emphasized that a senator who violates criminal law, in his role as a legislator, cannot avoid liability by simply asserting the legislative privilege. In the case of Senator Gravel, this meant he could be asked about his involvement with the publication of confidential national defense papers because this was a nonprivileged violation of criminal law. He could not, however, be forced to answer questions regarding the subcommittee meeting at which he had portions of the confidential papers read into the public record. The Court ultimately held that the Speech or Debate Clause could not legitimize Senator Gravel's aide's refusal to testify before a grand jury about his or the Senator's criminal action.

28. See id. at 625 (giving a standard for determining which acts are privileged by the Speech or Debate Clause).
29. Id.
30. See Robert J. Reinstein & Harvey A. Silverglate, Legislative Privilege and the Separation of Powers, 86 HARV. L. REV. 1113, 1118 (1973) (stating that in Gravel, "a split Court held that the scope of activities protected by the clause is very narrow and does not include publication of the record or receipt of the material for use in committee").
31. See Gravel v. United States, 408 U.S. 606, 626 (1972) (stating that the Speech or Debate Clause "does not privilege either Senator or aide to violate an otherwise valid criminal law in preparing for or implementing legislative acts").
32. See id. at 626 (“If republication of these classified papers would be a crime under an Act of Congress, it would not be entitled to immunity under the Speech or Debate Clause.”).
33. See id. at 616 (“We have no doubt that Senator Gravel may not be made to answer—either in terms of questions or in terms of defending himself from prosecution—for the events that occurred at the subcommittee meeting.”).
34. See id. at 626–27 (“The Speech or Debate Clause does not in our view
On the same day that *Gravel* was decided, the Court, in *Brewster v. United States*, applied a similar standard for interpreting legislative acts under the Clause. The Court explicitly stated that the *Kilbourn* standard should not be expanded to include in the privilege “everything . . . ‘related’ to the office of a Member.” If such a standard were implemented, legislators could characterize nearly every action as somehow related to the legislative process. To avoid this impermissible broadening, the Court restricted the privilege to questioning about legislators’ actions and motivations that are “clearly a part of the legislative process.” Like in *Gravel*, when applying this extend immunity to Rodberg, as a Senator's aide, from testifying before the grand jury about the arrangement between Senator Gravel and Beacon Press or about his own participation, if any, in the alleged transaction . . . .

35. See generally id. (indicating that the case was decided on June 29, 1972); *Brewster v. United States*, 408 U.S. 501 (1972) (same).

36. See *Brewster v. United States*, 408 U.S. 501, 528–29 (1972) (holding that “under these statutes and this indictment, prosecution of appellee is not prohibited by the Speech or Debate Clause”). The Court considered which actions by legislators qualify as immune legislative acts under the Speech or Debate Clause. *Id.* at 503. Senator Brewster was charged with accepting a bribe in return for promises regarding postage rate legislation. *Id.* at 502. Senator Brewster argued that his actions were privileged under the Speech or Debate Clause and the indictment must be dismissed. *Id.* at 503. The Court disagreed with Senator Brewster. *Id.* at 528–29. It explained that although past cases have defined a legislative act “as an act generally done in Congress in relation to the business before it,” the privilege is not unlimited. *Id.* at 512. The Court noted that there are a number of activities that are in some sense related to legislative activity but are not protected under the Clause because they are not “clearly a part of the legislative process.” *Id.* at 512–13, 516. The Court emphasized that a strong link to the legislative process is necessary to prevent legislators from avoiding criminal prosecution. *Id.* at 520. The Court concluded that Senator Brewster's actions were not privileged because accepting a bribe was in no way "part of the legislative process or function." *Id.* at 525, 528–29.

37. See *id.* at 515–16 (“In every case thus for [sic] before this Court, the Speech or Debate Clause has been limited to an act which was clearly a part of the legislative process—the *due* functioning of the process.”).

38. *Id.* at 513–14.

39. See *id.* at 516 (stating that if the Speech or Debate Clause was construed to extend the legislative privilege "to include all things in any way related to the legislative process," there would be "few activities in which a legislator engages that he would be unable somehow to 'relate' to the legislative process").

40. *Id.*
standard to the facts, the Court determined that accepting a bribe does not meet the legislative act standard and was therefore not privileged.41

Although it appears that the Supreme Court has attempted to narrow the legislative act definition, 42 a number of cases have permitted legislators’ actions to be privileged in disconcerting circumstances.43 In Doe v. McMillan, 44 school children’s parents

41. See id. at 526 (stating that “[t]aking a bribe is, obviously, no part of the legislative process or function” and is therefore not a privileged legislative action).

42. See id. at 515–16 (“[T]he Speech or Debate Clause has been limited to an act which was clearly a part of the legislative process—the due functioning of the process.”); United States v. Gravel, 408 U.S. 606, 625 (1972) (stating that not every action taken by a legislator will be privileged, and to be privileged under the Clause, an action must be essential to carrying out the individual’s duties as a legislator); see also Eastland v. U.S. Servicemen’s Fund, 421 U.S. 491, 504 (1975) (applying the standard set forth in Gravel for determining whether activities qualify as immune legislative acts (citing Gravel, 408 U.S. at 625)).

43. See Eastland, 421 U.S. at 507 (“We conclude that the Speech or Debate Clause provides complete immunity for the Members for issuance of this subpoena.”); Doe v. McMillan, 412 U.S. 306, 312–13 (1973) (“Congressmen and their aides are immune from liability for their actions within the ‘legislative sphere . . . .’” (quoting Gravel, 408 U.S. at 624–25)).

44. See Doe v. McMillan, 412 U.S. 306, 324 (1973) (holding that “the Court of Appeals applied the immunities of the Speech or Debate Clause and of the doctrine of official immunity too broadly”). The Court considered the scope of the Speech or Debate Clause. Id. at 307. The House of Representatives directed the Committee on the District of Columbia to carry out an investigation. Id. As part of the investigation, the Committee collected academic information about specific students. Id. at 308. The report was presented at Committee hearings, given to the Speaker of the House, and voted on for publication. Id. at 312. Additionally, the report was printed and given to members of Congress for “legislative purposes.” Id. Parents of children who were identified in the report brought suit for a violation of theirs “and their children’s statutory, constitutional, and common-law rights to privacy.” Id. at 309. The Court considered whether the investigation and use of the children’s information should be protected by the legislative privilege. Id. at 309–11. The Court stated that to be immune under the privilege, the actions must fall “within the ‘legislative sphere.’” Id. at 312 (citation omitted). The Court noted that actions that qualify as legislative acts can be privileged even though in other situations they would violate the law. Id. at 312–13. Additionally, the Court stated that it is not the role of the Judiciary to question the necessity of congressional action taken within the legislative sphere. Id. at 313. The Court held that the authorization of the investigation, the actual investigation, the disclosure of the information at the hearings, and the preparation, publication, and distribution.
sued legislators and their aides who collected information from schools about the children. The parents claimed a violation of theirs “and their children’s statutory, constitutional, and common-law rights to privacy.” The Court stated that the investigation, the presentation of the information at Committee hearings, and the referral of the report to the Speaker of the House were all privileged legislative acts. The information remained privileged when it was “distributed to and used for legislative purposes by Members of Congress, congressional committees, and institutional or individual legislative functionaries.” The Court indicated that acts within the “legislative sphere” are privileged even if in other situations they would be considered unconstitutional or a violation of local law. The Court’s only constraint on this extremely broad interpretation of the privilege was its statement that it did not extend to those who, by congressional authorization, provided the materials to the public in violation of the Constitution and other laws.

of the report to Committee members were privileged legislative acts. Id. at 313. The Court also held that, despite congressional authorization, defendants who made the information available to the public at large were not acting within the legislative sphere. Id. at 316. Thus, their actions were not privileged. Id.

45. See id. at 309 (stating that petitioners brought suit on behalf of their children against various members of the House Committee on the District of Columbia).

46. Id.

47. See id. at 312 (stating the Clause protected Congressmen-Committee members, the Committee staff, the consultant, and the investigator who “introduce[ed] material at Committee hearings that identified particular individuals, . . . referr[ed] the report that included the material to the Speaker of the House, and [who] vot[ed] for publication of the report”).

48. Id.

49. See id. at 312–13 (“Congressmen and their aides are immune from liability for their actions within the ‘legislative sphere’ even though their conduct, if performed in other than legislative contexts, would in itself be unconstitutional or otherwise contrary to criminal or civil statutes.” (citation omitted)).

50. See id. at 316 (stating that defendants who provided the information to the public, pursuant to congressional authorization, were not protected by the privilege).
Similarly, in *Eastland v. United States Servicemen’s Fund*, 51 the Court determined that a senator’s questionable actions fell within the scope of the privilege, despite the attempted narrowing of the legislative act definition. 52 In this case, Senator Eastland signed a subpoena in his role as chairman of a Senate subcommittee. 53 The organization affected by the subpoena claimed it violated its constitutional rights and sought an injunction that would prevent its enforcement. 54

The Court applied the test created in *Gravel*. 55 It determined that investigation and inquiry qualify as legislative acts because they are “an integral part of the legislative process.” 56 The Court emphasized that the investigation was protected by the privilege.

---

51. See *Eastland v. U.S. Servicemen’s Fund*, 421 U.S. 491, 507 (1975) (“We conclude that the *Speech or Debate Clause* provides complete immunity for the Members for issuance of this subpoena.”). The Court considered whether a subpoena issued by Senator Eastland, in his role as chairman of a Senate subcommittee, fell within the “sphere of legitimate legislative activity” protected by the *Speech or Debate Clause*. *Id.* at 494, 501. In 1970, the Senate authorized the Senate Subcommittee on Internal Security to conduct an investigation of the Internal Security Act of 1950. *Id.* at 493. Acting under this authority, the Subcommittee initiated an investigation of the United States Servicemen’s Fund, Inc. (USSF). *Id.* While carrying out the examination, Senator Eastland signed a *subpoena duces tecum*, which was issued to USSF’s bank. *Id.* at 494. Consequently, USSF brought an action seeking an injunction that would prevent the enforcement of the subpoena. *Id.* at 496. USSF claimed that the subpoena was aimed at silencing the organization in violation of the First Amendment. *Id.* at 495. In determining whether the act of issuing the subpoena should be privileged, the Court reaffirmed the test for legislative action set forth in prior cases. *Id.* at 501–04. The Court then determined that subpoenas have “long been held to be a legitimate use by Congress of its power to investigate.” *Id.* at 504 (citation omitted). Thus, the Court concluded that the legislative privilege protected the legislators who authorized the subpoena. *Id.* at 507.

52. See *id.* at 507 (“We conclude that the *Speech or Debate Clause* provides complete immunity for the Members for issuance of this subpoena.”).

53. See *id.* at 494 (stating that Senator Eastland approved a *subpoena duces tecum*).

54. See *id.* at 495–96 (stating that the United States Servicemen’s Fund, Inc. sought an injunction claiming that the subpoena violated the First Amendment).

55. See *id.* at 504 (quoting the *Gravel* standard for determining whether actions qualify as immune legislative acts).

56. *Id.* at 504–05.
because it was connected to a permissible function of Congress.\textsuperscript{57} In reaching this conclusion, the Court distinguished \textit{Eastland}'s facts from prior cases in which illegal actions by legislators were not privileged.\textsuperscript{58} For example, the Court explained that in \textit{Kilbourn}, the individual who carried out an illegal arrest was not immune because his action was not “essential to legislating.”\textsuperscript{59} In contrast, according to the Court, the \textit{Eastland} subpoena was privileged because it was made in an investigation relating to a topic that could be the subject of lawmaking.\textsuperscript{60} This was found to be true despite the alleged violation of the Constitution.

Finally, it is important to note that the legislative privilege only applies to legislative actions that have been completed.\textsuperscript{61} As explained in \textit{United States v. Helstoski},\textsuperscript{62} “[p]romises by a

\begin{itemize}
\item \textsuperscript{57} See \textit{id.} at 505–06 (“The particular investigation at issue here is related to and in furtherance of a legitimate task of Congress.” (citing \textit{Watkins v. United States}, 354 U.S. 178, 187 (1957))).
\item \textsuperscript{58} See \textit{id.} at 507–08 (explaining that respondents had pointed to language in \textit{Gravel} to argue that the subpoena should not be privileged because it violated their right to privacy (citing \textit{Gravel v. United States}, 408 U.S. 606, 621 (1972))). The Court, however, disagreed and distinguished the cases. \textit{id.} at 508.
\item \textsuperscript{59} \textit{id.} at 508 (citation omitted).
\item \textsuperscript{60} See \textit{id.} (stating that the action in this case was privileged because it was “a routine subpoena intended to gather information about a subject on which legislation may be had” (citing \textit{Quinn v. United States}, 349 U.S. 155, 161 (1955))).
\item \textsuperscript{61} See Jay Rothrock, \textit{Striking a Balance: The Speech or Debate Clause’s Testimonial Privilege and Policing Government Corruption}, 24 TOURO L. REV. 739, 752 (2008) (“[T]he Court in \textit{United States v. Helstoski} held that the term ‘legislative acts’ extends only to those past acts which have already taken place.”(citation omitted)).
\item \textsuperscript{62} See \textit{United States v. Helstoski}, 442 U.S. 477, 489 (1979) (holding that the Speech or Debate Clause protects “references to past legislative acts of a Member” from required disclosure). The Court considered the effect of the Speech or Debate Clause on the admissibility of legislative act evidence in a former congressman’s criminal trial. \textit{id.} at 479. Respondent Helstoski, a former member of the House, was on trial for allegedly receiving payment in return for promises of legislative action, which he eventually carried out. \textit{id.} Specifically, Helstoski accepted money and in return introduced immigration bills that were favorable to aliens. \textit{id.} Helstoski appeared before a grand jury on numerous occasions and eventually claimed his indictment must be dismissed under the Speech or Debate Clause. \textit{id.} at 484. The district court did not dismiss the indictment. \textit{id.} It did, however, state that the Speech or Debate Clause prohibited the Government from introducing evidence of completed legislative acts at trial. \textit{id.} The Government appealed, arguing that completed legislative
Member to perform an act in the future are not legislative acts."63 Thus, they are not immune under the Speech or Debate Clause.64

B. Defining “shall not be questioned in any other Place”

In determining the scope of the legislative privilege, the Supreme Court has also interpreted the meaning of the phrase “shall not be questioned in any other Place.”65 The Supreme Court has determined that this language of the Speech or Debate Clause provides legislators with testimonial66 and liability privileges.67

The suggestion of a testimonial privilege first arose in *United States v. Johnson*.68 In *Johnson*, the lower court concluded that acts should be admissible in order to show the respondent’s motive for accepting the bribe. *Id.* at 485. The Supreme Court disagreed, stating that the case law “leave[s] no doubt that evidence of a legislative act of a Member may not be introduced by the Government.” *Id.* at 487 (citations omitted). The Court indicated that the fact that this holding may make prosecutions more difficult is irrelevant. *Id.* at 488. Finally, the Court stated that promises of future legislative acts will not be privileged under the Clause. *Id.* at 489.

63. *Id.* at 489.

64. See *id.* (indicating that the Clause does not protect legislators from inquiry about legislative acts that may occur in the future).


66. See, e.g., *United States v. Gravel*, 408 U.S. 606, 616 (1972) (stating that a senator cannot “be made to answer . . . in terms of questions” about a legislative act); *United States v. Johnson*, 383 U.S. 169, 177 (1966) (“We see no escape from the conclusion that such an intensive judicial inquiry, made in the course of a prosecution by the Executive Branch under a general conspiracy statute, violates the express language of the Constitution and the policies which underlie it.”).

67. See, e.g., *Eastland v. U.S. Servicemen’s Fund*, 421 U.S. 491, 502–03 (1975) (explaining that the Speech or Debate Clause provides legislators with “protection against civil as well as criminal actions, and against actions brought by private individuals as well as those initiated by the Executive Branch”).

68. See *United States v. Johnson*, 383 U.S. 169, 184–85 (1966) (holding that a conviction based on a criminal statute that is obtained by questioning a legislator about legislative action is a violation of the Speech or Debate Clause). The Court considered whether a former congressman’s conspiracy conviction could be overturned because of a violation of the Speech or Debate Clause. *Id.* at 171. The lower court found former Congressman Johnson guilty of conspiracy. *Id.* Johnson had agreed to persuade the Department of Justice to dismiss charges against a loan company in exchange for funds from the company. *Id.* at
former Congressman Johnson had agreed to persuade the Department of Justice to dismiss charges against a loan company in exchange for funds from the company. Additionally, the lower court found that Johnson had agreed to make a speech in the House that would support the loan company. The Supreme Court, however, held that Johnson’s legislative privilege had been violated because the conviction was obtained through inquiry into the circumstances and motives of a speech made in the House. Although the Court found in favor of Johnson, it made clear that its decision is limited to the facts of the case. The Clause cannot be used to prohibit a criminal conviction that is not based on inquiry into privileged legislative action.

The testimonial privilege initially suggested in Johnson was specifically recognized in Gravel. Gravel indicated that the

---

171–72. Additionally, Johnson had agreed to make a speech in the House that would support the loan company. Id. The conspiracy conviction, however, was set aside by the court of appeals due to a claim that questioning Johnson about his agreement to make the speech in return for payment violated the Speech or Debate Clause. Id. at 171. The Court first explained that the Speech or Debate Clause clearly does not cover “conduct, such as was involved in the attempt to influence the Department of Justice, that is in no wise related to the due functioning of the legislative process.” Id. at 172. The Court next considered the Government’s inquiry about the speech. Id. at 173–76. The Government had questioned Johnson about who wrote the speech, the reasons for certain statements contained in the speech, and the motives for its delivery. Id. The Court subsequently held that this type of questioning violated the Speech or Debate Clause. Id. at 184–85; see also Rothrock, supra note 61, at 751 n.45 (stating that “Johnson implies the creation of the testimonial privilege”).

69. See Johnson, 383 U.S. at 171–72 (explaining the charges against former Congressman Johnson).

70. See id. (explaining the bribe accepted by Johnson).

71. See id. at 184–85 (holding that a criminal conviction based on questioning about legislative actions violated the Speech or Debate Clause).

72. See id. at 185 (“We emphasize that our holding is limited to prosecutions involving circumstances such as those presented in the case before us.”).

73. See id. (“Our decision does not touch a prosecution which, though as here founded on a criminal statute of general application, does not draw in question the legislative acts of the defendant member of Congress or his motives for performing them.”).

74. See Rothrock, supra note 61, at 751 n.45 (“[M]any courts cite [the testimonial privilege’s] creation to Gravel v. United States, 408 U.S. 606, 615, 616 (1972), which first explicitly articulated and applied the privilege impliedly created in Johnson.”).
testimonial privilege permits legislators and their aides to refuse to testify about legislative acts.\textsuperscript{75} Additionally, the testimonial privilege can be used to immunize qualifying legislative acts from being admitted into evidence at trial.\textsuperscript{76} This will not, however, prevent a legislator from being convicted through questioning about actions not protected by the Clause.\textsuperscript{77}

In addition to the testimonial privilege, the Clause provides a powerful immunity against liability.\textsuperscript{78} Thus, when a legislator is acting within the legislative sphere,\textsuperscript{79} his actions are immune from suit.\textsuperscript{80} The immunity against liability protects congressmen’s legislative acts from being the subject of civil and criminal actions.\textsuperscript{81}

In \textit{Eastland}, the Court cited numerous cases confirming that the Clause provides a very broad immunity privilege.\textsuperscript{82} \textit{Eastland} explained that this expansive interpretation is necessary to ensure that legislators can carry out their legislative tasks without the threat of suit.\textsuperscript{83} The Court reasoned that both civil and criminal actions “create[] a distraction and force[] Members to divert their time, energy, and attention from their legislative

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{75} See Gravel v. United States, 408 U.S. 606, 616 (1972) (stating that the Senator and his aide could not be forced to testify about the Subcommittee meeting).
\item \textsuperscript{76} See Rothrock, supra note 61, at 751 (“[A] testimonial privilege may be asserted to prevent the admission of legislative acts into evidence . . . .”).
\item \textsuperscript{77} See id. (“[T]he legislator can still be prosecuted based upon the unprotected evidence.”).
\item \textsuperscript{78} See Eastland v. U.S. Servicemen’s Fund, 421 U.S. 491, 502–03 (1975) (discussing the Clause’s protection against civil and criminal actions and those brought by the Executive).
\item \textsuperscript{79} See supra Part II.A (defining which actions by legislators qualify as privileged legislative acts).
\item \textsuperscript{80} See \textit{Eastland}, 421 U.S. at 503 (“We reaffirm that once it is determined that Members are acting within the ‘legitimate legislative sphere’ the \textit{Speech or Debate Clause} is an absolute bar to interference.” (citing Doe v. McMillan, 412 U.S. 366, 314 (1973))).
\item \textsuperscript{81} See id. at 502 (stating that the Clause “protect[s] against civil as well as criminal actions”).
\item \textsuperscript{82} See id. at 502–03 (citing Kilbourn v. Thompson, 103 U.S. 168 (1881), Tenney v. Brandhove, 341 U.S. 367 (1951), Doe v. McMillan, 412 U.S. 306 (1973), and Dombrowski v. Eastland, 387 U.S. 82 (1967)).
\item \textsuperscript{83} See id. at 503 (explaining the negative implications of lawsuits on legislators’ ability to perform their legislative duties).
\end{itemize}
\end{footnotesize}
tasks to defend the litigation.”84 This unnecessary “distraction” can impinge the legislative process.85 Additionally, the Judiciary’s involvement threatens “legislative independence,” which the Clause was designed to protect.86

III. The Circuit Split: The Speech or Debate Clause and Its Application to Document Review

An issue that has not been addressed by the Supreme Court is the Speech or Debate Clause’s application to governmental review of a legislator’s documents in an investigation into the legislator’s alleged misconduct.87 The resolution of this issue depends on whether document review falls within the purview of the phrase “shall not be questioned in any other Place.”88 Although the Supreme Court has not spoken on this issue,89 the Ninth and D.C. Circuits have each drawn a different conclusion regarding the breadth of the privilege, creating a split among the circuits.90

84. Id.
85. Id.
86. Id.
87. See United States v. Rayburn House Office Bldg., 497 F.3d 654, 659 (D.C. Cir. 2007) (indicating that the Supreme Court has not addressed the Speech or Debate Clause’s application to document disclosure).
88. See id. at 659–60 (considering whether the testimonial privilege “includes a non-disclosure privilege”).
89. See id. at 659 (“The Supreme Court has not spoken to the precise issue at hand.”).
90. Compare United States v. Renzi, 651 F.3d 1012, 1039 (9th Cir. 2011) (“T]he Clause does not incorporate a non-disclosure privilege as to any branch,” (citations omitted)), with Rayburn, 497 F.3d at 663 (“[W]e hold that a search that allows agents of the Executive to review privileged materials without the Member’s consent violates the Clause.”).
A. The D.C. Circuit: United States v. Rayburn House Office Building

In United States v. Rayburn House Office Building,91 the United States Court of Appeals for the District of Columbia Circuit considered the propriety of a search of former Congressman Jefferson’s office in light of the privilege conferred by the Speech or Debate Clause.92 This case is significant because the search was the first time that the Executive had ever ordered a search of a sitting congressman’s office.93 Not surprisingly, Jefferson challenged the search, asserting a violation of the Speech or Debate Clause.94

91. See United States v. Rayburn House Office Bldg., 497 F.3d 654, 663 (D.C. Cir. 2007) (holding that “a search that allows agents of the Executive to review privileged materials without the Member’s consent violates the Clause”). The court considered the application of the Speech or Debate Clause to searches performed by the Executive Branch. Id. at 655. Additionally, the court determined the proper remedy for a violation of the Clause. Id. at 663–66. The Department of Justice issued a search warrant for former Representative William J. Jefferson’s office. Id. at 656. The warrant was issued in an investigation surrounding Jefferson’s alleged involvement in a bribery scheme. Id. It was believed that Jefferson had accepted payments from businesses and in return promised to use his position in Congress to promote legislation benefitting those paying him. Id. The warrant permitted FBI agents to collect information from Jefferson’s office. Id. Then, a “filter team” reviewed the information and was instructed to remove documents that “were not responsive to the search warrant” or “were subject to the Speech or Debate Clause privilege.” Id. at 656–57. The circuit court was given documents labeled “potentially privileged” to determine if they were in fact covered by the Clause. Id. at 657. Jefferson argued that the search violated his legislative privilege. Id. The court pointed to Brown & Williamson Tobacco Corp. v. Williams, 62 F.3d 408, 415 (D.C. Cir. 1995), in which it held that “the Clause includes a non-disclosure privilege” in civil suits. Id. at 660. The court reasoned that a nondisclosure privilege applies in criminal actions as well. Id. The court determined that the privilege was violated because the Executive had access to privileged documents before Jefferson had the opportunity to claim that they were protected by the Clause. Id. at 662. Finally, the court ordered that the privileged documents be returned and prohibited those who saw them from revealing their contents. Id. at 666.

92. See id. at 655 (considering whether the search of former representative Jefferson’s office violated the Speech or Debate Clause).

93. See id. at 659 (“May 20–21, 2006 was the first time a sitting Member’s congressional office has been searched by the Executive.”).

94. See id. at 657 (stating that Jefferson “argued . . . that the issuance and execution of the search warrant violated the Speech or Debate Clause”).
In determining that the legislative privilege was violated, the court followed its holding in *Brown & Williamson Tobacco Corp. v. Williams*.95 The D.C. District Court, in *Brown*, created an extremely broad testimonial privilege in civil suits.96 Specifically, the *Brown* court protected not only direct questioning about legislative action but also documents referring to such acts.97

In following *Brown*, the *Rayburn* court explained that the holding in *Brown* was rooted in a concern that document disclosure would “distract[]” legislators.98 *Rayburn* warned that without a document privilege, “the possibility of compelled disclosure may . . . chill the exchange of views with respect to legislative activity.”99 Based on this reasoning, the *Rayburn* court concluded that discovery of legislative act documents should similarly be prohibited in criminal suits.100

---

95. See id. at 660, 663 (discussing the District Court of Appeals for the D.C. Circuit’s precedent on this issue and finding the legislative privilege violated); *Brown & Williamson Tobacco Corp. v. Williams*, 62 F.3d 408, 423 (D.C. Cir. 1995) (“B & W’s claim at bottom, is to a right to engage in a broad scale discovery of documents in a congressional file that comes from third parties. The Speech or Debate Clause bars that Claim.”). The court considered whether the Speech or Debate Clause permits congressmen to refuse to respond to subpoenas requiring them to produce documents in a suit among third parties. *Id.* at 412. The underlying suit pertained to documents stolen from an attorney’s office by a former paralegal. *Id.* at 411. A House of Representatives subcommittee received the documents at issue. *Id.* at 412. Subpoenas ordering copies of the stolen documents were then issued to two congressmen on the committee that had allegedly received them. *Id.* The congressmen claimed the Speech or Debate Clause exempted them from compliance with the subpoenas. *Id.* The court determined that the Clause privileges congressmen from being forced to produce “documentary evidence” as well as being required to testify. *Id.* at 420. The court reasoned that such a privilege is vital to preventing congressional “distractions” that “divert [legislators’] time, energy, and attention from their legislative tasks.” *Id.* at 421 (citation omitted). The court determined that the subpoena violated the legislative privilege. *Id.* at 423.

96. See *Brown*, 62 F.3d at 420 (discussing the Speech or Debate Clause’s application to documents).

97. See id. (stating that the Speech or Debate Clause’s testimonial privilege applies to “[d]ocumentary evidence”).

98. See *United States v. Rayburn House Office Bldg.*, 497 F.3d 654, 660 (D.C. Cir. 2007) (stating that the *Brown* decision was based on the concern that document disclosure requirements would cause legislative “distraction[]”).

99. *Id.* at 661.

100. See *id.* at 660 (stating that a nondisclosure privilege must apply in both civil and criminal suits).
prevented from removing privileged documents before his office was searched, the D.C. Circuit held that his rights granted by the Speech or Debate Clause were violated. The court assured that this does not mean courts cannot review a legislator’s claim of privilege. Pursuant to its holding, the court ordered that the privileged documents be returned to Jefferson. The Government, however, was permitted to retain the nonprivileged documents.

B. The Ninth Circuit: United States v. Renzi

The question of the Speech or Debate Clause's application to document disclosure was also considered in United States v. Renzi. In Renzi, former Arizona Congressman Richard Renzi

101. See id. at 662 (stating that the search violated the Clause because it “denied the Congressman any opportunity to identify and assert the privilege with respect to legislative materials before their compelled disclosure to Executive agents”).

102. See id. (indicating that courts can review a congressman’s assertion of privilege).

103. See id. at 666 (“[W]e hold that the Congressman is entitled to the return of all legislative materials . . . that are protected by the Speech or Debate Clause seized [during the search].”).

104. See id. at 665 (“[I]t is unnecessary to order the return of non-privileged materials as a further remedy for the violation of the Clause.”).

105. See United States v. Renzi, 651 F.3d 1012, 1039 (9th Cir. 2011) (holding that the “actions for which [Renzi was] prosecuted [lay] beyond [the] limits” of the Speech or Debate Clause). The court considered the role of the Speech or Debate Clause in the criminal investigation of former Arizona Congressman Richard G. Renzi. Id. at 1016. Renzi was indicted for “public corruption charges of extortion, mail fraud, wire fraud, money laundering, and conspiracy.” Id. at 1018. Specifically, Renzi was accused of promising legislative action that would benefit parties who agreed to purchase a particular piece of land. Id. at 1016. Purchase of the land would allow its owner to repay a debt owed to Renzi. Id. Renzi contended that the land negotiations were legislative acts. Id. He argued that his actions were immune from suit under the Speech or Debate Clause. Id. Additionally, he contested the Government’s use of what he believed to be legislative act evidence to obtain a grand jury indictment. Id. Renzi requested a hearing at which the Government would have to show that it had not used evidence that he believed was protected by the privilege. Id. He argued that all the evidence relating to the land negotiations should have been suppressed. Id. First, the court determined that the land negotiations were not legislative acts. Id. at 1023. Next, the court concluded that the indictment was not "caused" by
was indicted for promising legislative action that would benefit parties who agreed to purchase a particular piece of land.\textsuperscript{106} Purchase of the land would allow its owner to repay a debt owed to Renzi.\textsuperscript{107} Throughout this discussion, this incident will be referred to as the land negotiations. Renzi first attempted to obtain the Clause’s liability, testimonial, and evidentiary privileges.\textsuperscript{108} In doing so, Renzi argued that his land negotiations were legislative acts.\textsuperscript{109} The court disagreed for two reasons.\textsuperscript{110} First, the land negotiations did not qualify because they were not completed legislative actions.\textsuperscript{111} The court refused to broaden the meaning of legislative acts to include discussions leading up to legislative action or promises of future legislative action.\textsuperscript{112} Second, the land negotiations were not privileged because accepting a bribe does not have the requisite relationship to the legislative process.\textsuperscript{113}

the use of privileged legislative act evidence. \textit{Id.} at 1031. Although the prosecution conceded that privileged documents had been shown to the jury, they were not necessary to obtain the indictment. \textit{Id.} at 1031–32. Therefore, the indictment was permitted to stand. \textit{Id.} Finally, the court expressly disagreed with the D.C. Circuit by permitting review of documentary evidence referencing legislative acts when the underlying action is not privileged. \textit{Id.} at 1036. The court upheld the district court judgment and dismissed each of Renzi’s causes of action. \textit{Id.}

\begin{itemize}
\item \textsuperscript{106} \textit{See id.} at 1016 (explaining the basis for Renzi’s indictment).
\item \textsuperscript{107} \textit{See id.} (explaining the basis for Renzi’s indictment).
\item \textsuperscript{108} \textit{See id.} at 1020 (explaining that if the charged actions qualified as legislative acts, the Government could not prosecute Renzi for them, the Government could not force him or his aides to testify about them, and “evidence of those acts could not be introduced to any jury, grand or petit” (citation omitted)).
\item \textsuperscript{109} \textit{See id.} at 1016 (“[H]e claims that the public corruption charges against him amount to prosecution on account of his privileged ‘legislative acts’ . . .”).
\item \textsuperscript{110} \textit{See id.} at 1022–27 (giving two reasons why Renzi’s action did not qualify as a legislative act).
\item \textsuperscript{111} \textit{See id.} at 1022 (“Completed ‘legislative acts’ are protected; promises of future acts are not.” (citation omitted)).
\item \textsuperscript{112} \textit{See id.} at 1023 (stating that “negotiating with and ultimately promising private individuals [to] perform future legislative acts” is not a legislative act (citation omitted) (emphasis removed)).
\item \textsuperscript{113} \textit{See id.} (“Taking a bribe is, obviously, no part of the legislative process or function; it is not a legislative act. It is not, by any conceivable interpretation, an act performed as part of or even incident to the role of legislator.” (citation omitted)).
\end{itemize}
Renzi also argued that the indictment must be thrown out because the grand jury was presented with privileged evidence.114 This argument also failed.115 Despite the fact that the jury was exposed to legislative act evidence, the jury could have made the indictment without it.116 There was sufficient evidence regarding the land negotiations, which were uncompleted nonprivileged actions, for the indictment to stand.117

Finally, the court considered the issue of document disclosure.118 Renzi sought an extremely broad privilege, like that in *Rayburn*,119 which would prevent the Executive Branch from forcing the congressman to produce documents relating to legislative acts.120 The court, however, refused to follow the D.C. Circuit.121 The court criticized *Rayburn*, stating that the D.C. Circuit erroneously permitted the threat of “distraction alone” to cause the privilege to attach.122 Although Renzi acknowledged that potential distraction has a role in the inquiry, it concluded that distraction can only trigger the privilege in cases in which

---

114. See id. at 1027 (considering “whether the district court erred by declining to dismiss the indictment in its entirety for, as Renzi alleges, the pervasive presentment of ‘legislative act’ evidence to the grand jury”).

115. See id. at 1031–32 (“We therefore have no cause to grant Renzi the relief he seeks.”).

116. See id. at 1031 (stating that although the privileged documents “should not have been presented, we cannot conclude that they were ‘essential elements of proof’ that caused the jury to indict” (citations omitted)).

117. See id. (explaining that to obtain the indictment, the Government only needed to “introduce evidence of Renzi’s promise to support legislation and the circumstances surrounding that promise”).

118. See id. at 1032–39 (considering whether the Speech or Debate Clause confers a document nondisclosure privilege).

119. See United States v. Rayburn House Office Bldg., 497 F.3d 654, 660 (D.C. Cir. 2007) (finding that the Speech or Debate Clause provides a document nondisclosure privilege).

120. See United States v. Renzi, 651 F.3d 1012, 1033 (9th Cir. 2011) (explaining that Renzi sought to “preclude the Executive from obtaining and reviewing ‘legislative act’ evidence”).

121. See id. at 1034 (“[W]e cannot agree with our esteemed colleagues on the D.C. Circuit.”).

122. See id. (“*Rayburn* rests on the notion that ‘distraction’ of Members and their staffs from their legislative tasks is a principal concern of the Clause, and that distraction alone can therefore serve as a touchstone for application of the Clause’s testimonial privilege.” (citation omitted)).
the action at issue is privileged. The court held that when evidence is sought to prosecute nonprivileged action, here Renzi’s illegal land negotiations, legislators can be forced to disclose documents containing references to legislative acts.

IV. Resolving the Split: Narrowing the Scope of the Legislative Privilege

In order to resolve this issue, it is important to carefully consider the text and history of the Speech or Debate Clause. Additionally, it is essential that one study the effect that this very powerful privilege has on the values, such as self-government, upon which the United States was founded.

A. A Look at the Text of the Speech or Debate Clause

In both the Renzi and Rayburn opinions, the courts reference the current scope of the Speech or Debate Clause as set forth in Supreme Court decisions. A textual analysis of the Clause, however, does not support the extremely broad reach it is given by courts today.

123. See id. at 1035 (stating the “concern for distraction alone precludes inquiry only when the underling action is itself precluded”).

124. See id. at 1036 (indicating that “documentary ‘legislative act’ evidence” can be reviewed as “part of an investigation into unprotected activity”).

125. See PHILIP BOBBITT, CONSTITUTIONAL FATE 7 (1982) (stating that two methods of constitutional interpretation are the “textual argument” and the “historical argument”).

126. See id. at 94 (“It is the character, or ethos, of the American polity that is advanced in ethical argument as the source from which particular decisions derive.”).

127. See United States v. Renzi, 651 F.3d 1012, 1021–23 (9th Cir. 2011) (discussing Supreme Court decisions interpreting the Speech or Debate Clause); United States v. Rayburn House Office Bldg., 497 F.3d 654, 659 (D.C. Cir. 2007) (same).

128. See Bobbitt, supra note 125, at 25–38 (discussing the “textual argument” as a method of constitutional interpretation). In this chapter, Bobbitt explains that the “textual argument” is based “on a sort of ongoing social contract, whose terms are given their contemporary meanings continually reaffirmed by the refusal of the People to amend the instrument.” Id. at 26.
Rayburn applied Brewster’s standard for interpreting the “speech or debate” portion of the Clause, stating that “the Supreme Court has limited the scope to conduct that is an integral part of ‘the due functioning of the legislative process.” 129 Renzi quoted similar language from Brewster.130 Additionally, Renzi cited Kilbourn for the assertion that the Clause protects more than just “words spoken in debate.”131 Such interpretations, however, impermissibly ignore the contemporary meaning of the Clause’s actual language.132

According to Webster’s dictionary, “speech” is defined as “the communication or expression of thoughts in spoken words.”133 “Debate” is defined as “a contention by words or arguments.”134 Each of these definitions requires words. Including legislators’ actions within the phrase “speech or debate” simply because they have a strong connection to legislating greatly deviates from the modern understanding of the Clause’s language.135

To resolve the issue of whether the Clause provides a nondisclosure privilege, the phrase “they shall not be questioned in any other Place” should also be assessed through a textual interpretation.136 The operative word in this phrase is

Thus, a textual constitutional analysis applies the modern meanings of words in order to interpret portions of the Constitution. Id. at 33.

130. Renzi, 651 F.3d at 1021–22.
131. Id. at 1021 (quoting Kilbourn v. Thompson, 103 U.S. 168, 204 (1880)).
132. See BOBBITT, supra note 125, at 30 (stating that the textual approach to constitutional analysis “give[s] absolute affect to the words of the Constitution”).
135. See, e.g., Brewster v. United States, 408 U.S. 501, 515–16 (1972) (stating that the Clause applies to “an act which was clearly a part of the legislative process—the due functioning of the process” (emphasis added)); see also BOBBITT, supra note 125, at 26 (stating that the textual constitutional analysis uses the current meanings of the words in the text).
136. See BOBBITT, supra note 125, at 25–38 (discussing the “textual argument” as a method of constitutional interpretation).
“questioned,” which is defined as “to ask a question of or about.” The word “question” is defined as “a form of words addressed to a person in order to elicit information or evoke a response; interrogative sentence.” It has also been defined as “an interrogative expression often used to test knowledge.” The plain meaning of the word “question” suggests that an individual be present in order to supply an answer to the questioner. Additionally, the word “they” in the Clause refers to congressmen who are the subjects of the questioning. Thus, the explicit language of the Clause supports a testimonial privilege, which protects congressmen from “words addressed to [them] in order to elicit information” about qualifying “speech or debate.” In contrast, an inanimate object, such as a document, cannot be “questioned” in such a manner. A document nondisclosure privilege would improperly expand the legislative privilege beyond the text’s contemporary meaning.

**B. The Legislative Privilege and the Framers’ Intent**

In addition to studying the Clause’s text, its history should be considered to determine the Clause’s application to a document nondisclosure privilege. As explained by John

---


140. See supra notes 138–39 and accompanying text (providing definitions for the word “question”).

141. See U.S. Const. art. I, § 6, cl. 1 (showing that the word “they” in the Speech or Debate Clause is referring to “Senators and Representatives”).

142. See id.; see also supra notes 138–39 and accompanying text (defining the word “question”).

143. See BOBBITT, supra note 125, at 33 (stating that the textual analysis interprets the Constitution by using modern definitions of its terms).

144. See id. at 9–24 (discussing the “historical argument” of constitutional
LIMITING THE LEGISLATIVE PRIVILEGE

Adams, the Constitution must be interpreted according to its original meaning “to preserve the advantages of liberty and to maintain a free government.”145 The Speech or Debate Clause should be applied today in accordance with the constitutional Framers’ intentions.146

The “original understanding” of the Clause can be defined by considering the circumstances of its inclusion in the American Constitution.147 The language of the Speech or Debate Clause is nearly identical to the parliamentary privilege granted by the 1689 English Bill of Rights.148 The Framers included the Clause at the Constitutional Convention without extensive discussion.149 Only minor changes were made to the Clause throughout the drafting period, and the final draft from the Committee of Detail had almost the same language as the Speech or Debate Clause in the Articles of Confederation.150 The Committee members unanimously accepted this final draft.151 Similarly, the Clause was not contested at states’ ratification debates.152 This nearly uncontroverted adoption of the Clause can be explained by the

145. Id. at 9 (quoting John Adams as support for the “historical argument” of constitutional interpretation).
146. See id. at 9–10 (stating that the historical argument considers the Framers’ intentions to determine the meaning of the Constitution).
147. See id. at 9–11 (“[T]he Supreme Court announced that construction of the Constitution must rely on ‘the meaning and intention of the convention which framed and proposed it for adoption and ratification.’” (citation omitted)).
148. See United States v. Rayburn House Office Bldg., 497 F.3d 654, 659 (D.C. Cir. 2007) (stating that the Speech or Debate Clause very closely matches the language of the 1689 English Bill of Rights).
149. See JOSH CHAFETZ, DEMOCRACY’S PRIVILEGED FEW 87 (2007) (“The Speech or Debate Clause seems to have been little discussed at the Philadelphia Convention.”).
150. See id. at 87–88 (explaining the minimal changes to the Clause throughout its drafting).
151. See id. at 88 (stating that the final draft by the Committee of Detail was “approved by the entire Convention without dissent or even recorded debate”).
152. See id. (“[T]he Clause was completely uncontroversial at the states’ ratification debates and in the debates in the press.”).
historical acceptance of a legislative speech or debate privilege in American government.\textsuperscript{153}

Because of the lack of debate surrounding the adoption of the Speech or Debate Clause at the Constitutional Convention, records from the Convention itself provide little insight into the Framers' understanding of the legislative privilege.\textsuperscript{154} After the Constitution's ratification, however, statements made by the Framers and other individuals who were influential in early America provide information regarding the Speech or Debate Clause's purpose.\textsuperscript{155}

James Wilson, a constitutional Framer, discussed the Speech or Debate Clause, stating:

\begin{quote}
In order to enable and encourage a representative of the publick to discharge his publick trust with firmness and success, it is indisputably necessary, that he should enjoy the fullest liberty of speech, and that he should be protected from the resentment of every one, however powerful, to whom the exercise of that liberty may occasion offence.\textsuperscript{156}
\end{quote}

Additionally, Wilson commended the Clause for supporting legislative independence.\textsuperscript{157} He said that it was "a very considerable improvement in the science and the practice of

\begin{itemize}
\item \textsuperscript{153} See \textit{id.} at 87 (stating that the Articles of Confederation and three state constitutions contained clauses providing a speech or debate privilege); \textit{see also} Reinstein & Silverglate, supra note 30, at 1136 ("Presumably because the principle was so firmly rooted, there was little discussion of the clause during the debates of the Constitutional Convention and virtually none at all in the ratification debates.").
\item \textsuperscript{154} See Reinstein & Silverglate, supra note 30, at 1136 (explaining that there was minimal discussion of the Clause at the Constitutional Convention); \textit{see also} Bobbitt, supra note 125, at 9 (explaining that "historical arguments depend on a determination of the original understanding of the constitutional provision to be construed").
\item \textsuperscript{155} See \textit{Chafetz}, supra note 149, at 88–89 (discussing statements made after the Constitutional Convention by the Founders about the Speech or Debate Clause's meaning).
\item \textsuperscript{156} \textit{Id.} at 88 (citation omitted).
\item \textsuperscript{157} See Laura Krugman Ray, \textit{Discipline Through Delegation: Solving the Problem of Congressional Housecleaning}, 55 U. Pitt. L. Rev. 389, 403 (1994) (stating that Wilson commended the Speech or Debate Clause for "its role in securing legislative independence").
\end{itemize}
government.”

From these statements, it is clear that the Framers included the Clause to facilitate legislators’ ability to carry out their role as lawmakers by shielding them from liability for their speech.

Thomas Jefferson wrote about the legislative privilege in a petition to the Virginia House of Delegates. Although Jefferson did not attend the Constitutional Convention or sign the Constitution, as one of the drafters of the Declaration of Independence and a “founding father,” his understanding of the Speech or Debate Clause can provide useful insight into the Framers’ intent. Jefferson made the petition in support of Samuel Cabell, who was a member of the House of Representatives. Cabell had been charged with seditious libel based on a letter he had written that spoke unfavorably of the Adams administration. In the petition, Jefferson wrote:

[I]n order to give to the will of the people the influence it ought to have, and the information which may enable them to exercise it usefully, it was a part of the common law, adopted as the law of this land, that their representatives, in the discharge of their functions, should be free from the cognizance or coercion of the coordinate branches, Judiciary and Executive; and that their communications with their constituents should of right, as of duty also, be free, full, and unawed by any.

First, the circumstances of the petition provide information on the original understanding of the Clause. Jefferson discussed
the legislative privilege in a petition to aid a congressman who had been charged because of a letter voicing his opinion.\textsuperscript{165} This fact supports the assertion that the Clause was aimed at preventing the other branches of government from punishing congressmen who express unpopular views.\textsuperscript{166} Second, in this statement, Jefferson, like Wilson, indicates that the legislative privilege is aimed at enabling legislators to fulfill their roles as lawmakers.\textsuperscript{167} According to Jefferson, in order for the people to actively participate in government through their legislators, they must be adequately informed.\textsuperscript{168} Consequently, the legislative privilege was included to encourage legislators to provide voters with essential information by reducing their fear that other branches of government will punish them.\textsuperscript{169}

In addition to the quoted language above, other portions of the petition emphasized “the link between freedom of speech and debate on the floor of the legislature, freedom of communication with constituents, and popular sovereignty.”\textsuperscript{170} Jefferson discussed the need for free communication between legislators and the people in representative government.\textsuperscript{171} He asserted that if legislators are put “into jeopardy of criminal prosecution, of vexation, expense, and punishment before the Judiciary” for statements that “do not exactly square with [its] ideas of fact or right, or with [its] designs of wrong,” legislative independence

\begin{itemize}
\item \textsuperscript{165} See id. (stating that Representative Cabell had been charged with seditious libel for a letter “denouncing the Adams administration”).
\item \textsuperscript{166} See id. (discussing Jefferson’s petition in support of Representative Cabell who was charged with seditious libel because of his controversial statements about the Adams administration).
\item \textsuperscript{167} See supra text accompanying note 163 (quoting Jefferson’s petition to the Virginia House of Delegates); see also supra notes 156–59 and accompanying text (discussing Wilson’s statements about the Speech or Debate Clause).
\item \textsuperscript{168} See supra text accompanying note 163 (quoting Jefferson’s petition to the Virginia House of Delegates).
\item \textsuperscript{169} See supra text accompanying note 163 (quoting Jefferson’s petition to the Virginia House of Delegates).
\item \textsuperscript{170} CHAFETZ, supra note 149, at 89.
\item \textsuperscript{171} See id. (“[F]or the Judiciary to . . . overawe the free correspondence which exists and ought to exist between [representatives and the people] . . . is to put the legislative department under the feet of the Judiciary . . . .”) (citation omitted)).
\end{itemize}
will be impinged and representative government destroyed.172

These statements provide further support for the argument that
the Speech or Debate Clause was designed to promote
communication between legislators and their constituents by
protecting legislators from liability for such communication.173

Lastly, James Madison stated that interpretation of the
Clause must be guided by “the reason and necessity of the
privilege.”174 He warned that “[i]t is certain that the privilege
has been abused in British precedents, and may have been in
American also.”175 These statements by James Madison indicate
that the Clause should be interpreted only as broadly as required
to accomplish its purposes.176 Additionally, courts must be careful
when applying the Clause in order to prevent legislators from
taking advantage of the protection for personal gain.177

In addition to considering statements made by the original
interpreters of the Constitution to determine the intended
meaning of the privilege, it is also significant that the Clause’s
language was largely taken from the English Bill of Rights.178
Thus, the Framers likely intended the American Speech or
Debate Clause to have a meaning similar to the English article
from which it originated.179 Consequently, the English
parliamentary privilege can be used to understand the
interpretation the Framers envisioned when they used nearly
identical language in the American Constitution.180

172. Id. (citation omitted).

173. See supra notes 170–72 and accompanying text (discussing Thomas
Jefferson’s statements about the Speech or Debate Clause’s purpose).

174. Chafetz, supra note 149, at 89 (citation omitted).

175. Id. (citation omitted).

176. See supra notes 174–75 and accompanying text (providing quotes from
James Madison about the Speech or Debate Clause).

177. See supra notes 174–75 (providing guidance from James Madison for
interpreting and applying the Speech or Debate Clause).

178. See United States v. Rayburn House Office Bldg., 497 F.3d 654, 659
(D.C. Cir. 2007) (stating that the Speech or Debate Clause very closely matches
the language of the 1689 English Bill of Rights).

179. See Kilbourn v. Thompson, 103 U.S. 168, 202 (1880) (“[I]t may be
reasonably inferred that the framers of the Constitution meant the same thing
by the use of language borrowed from that source.”).

180. See Rayburn, 497 F.3d at 659 (stating that the American Speech or
The concept of privileging parliamentary speech and debate is deeply rooted in English history. The English privilege emerged in the early fourteenth and fifteenth centuries. In the beginning, its scope was limited to shielding parliamentary speech and debate from civil suits. Over time, the privilege was expanded to include protection from Executive prosecution. This expansion brought a period of lengthy controversy with the Crown regarding the privilege’s reach. During this era, monarchs, displeased with criticism of their policies by members of Parliament, asserted, “[T]he privilege ended where [monarchial] prerogatives began.” Conversely, members of Parliament argued that the privilege applied to any action related to their role in Parliament.

The contention regarding the reach of the parliamentary privilege was present in the case of Sir William Williams. Sir William Williams, in his parliamentary role as Speaker of the House of Commons, permitted the publication of Dangerfield’s *Narrative of the Late Popish Designs*. He was subsequently

---

Debate Clause has nearly the same language as the 1689 English Bill of Rights granting the parliamentary privilege).

181. See *Chafetz*, supra note 149, at 69 (“The privilege of Members of Parliament to be free from questioning in any other place for their speeches in Parliament is undoubtedly an ancient one . . . .”).

182. See Reinstein & Silverglate, supra note 30, at 1122 (“Parliament’s privileges originated in the fourteenth and fifteenth centuries . . . .”).

183. See id. (stating that the parliamentary privilege originally only protected members of Parliament from civil suit).

184. See id. at 1123 (“The free speech privilege evolved gradually and painfully into a practical instrument for security against the executive, an evolution triggered by basic changes in the functions of the legislature.”).

185. See id. at 1126 (explaining that parliamentary “intrusions into the Crown’s prerogatives led to a century-long battle over Parliament’s freedom of speech or debate”).

186. Id. at 1127.

187. See id. (“[T]he House declar[ed] that the privilege was absolute for any matter touching parliamentary business.”).

188. See *Chafetz*, supra note 149, at 74 (explaining that Sir William Williams was “hauling before the King’s Bench on a charge of seditious libel” for an action carried out in his role as the Speaker of the House of Commons).

189. See id. (stating that Sir William Williams, while acting as the Speaker of the House of Commons, “ordered the printing of a pamphlet (Dangerfield’s *Narrative of the Late Popish Designs*) that was libelous of the Duke of York”).
charged with seditious libel. He raised the defense that his action was privileged because it was “done in [the] time of parliament, and ordered to be done by the House of Commons.” Despite his assertion of privilege, Sir William Williams was convicted.

Shortly after Sir William Williams’s conviction, Parliament passed the 1689 English Bill of Rights in response to incidents of monarchial aggression. The Bill of Rights included Article 9, which stated, “[T]he freedom of speech, and debates or proceedings in parliament, ought not to be impeached or questioned in any court or place out of parliament.” The legislative privilege guarantee was included for parliamentarians, like Sir William Williams, who were punished by the Crown for their parliamentary actions. Consequently, shortly after the Bill of Rights was passed, the House of Commons passed a resolution that pronounced the judgment against Williams “an illegal Judgment, and against the Freedom of Parliament.”

Although Sir William Williams was convicted, the 1689 English Bill of Rights was created to protect individuals in similar situations. This fact indicates that the English

---

190. See id. (stating that Sir William Williams was charged with seditious libel).
191. Id. (citation omitted).
192. See id. (stating that despite Sir William Williams’s assertion of parliamentary privilege, “[t]he judges disagreed”).
193. See Reinstein & Silvergate, supra note 30, at 1134 (“[S]hortly after the indictment of Sir William Williams, James II published a Declaration of Indulgence declaring it to be his ‘royal will and pleasure that . . . the execution of all and all manners of penal laws in matters ecclesiastical . . . be immediately suspended.’” (citation omitted)). This Declaration gave the “executive power to nullify statutes passed by Parliament.” Id. In response, the 1689 English Bill of Rights “both abolished the suspending power and guaranteed the speech or debate privilege.” Id. at 1135.
194. Chafetz, supra note 149, at 74 (citation omitted).
195. See id. at 74–75 (explaining that one drafter of the English Bill of Rights said that the Article 9 legislative privilege “was put in for the sake of one . . . Sir William Williams, who was punished out of Parliament for what he had done in Parliament” (citation omitted)).
196. Id. at 75 (citation omitted).
197. See id. (stating that the 1689 English Bill of Rights was created for
parliamentary privilege was aimed at shielding parliamentarians from suits based on the content of their speech. Consequently, it can be inferred that the Framers of the American Constitution had a similar protection in mind when they adopted nearly the exact same language in the American Speech or Debate Clause.

In summary, the statements by James Wilson, Thomas Jefferson, and James Madison reveal two related purposes of the Clause. First, the Framers sought to protect legislative independence by shielding legislators from speech-based litigation. Second, the Framers aimed to help legislators carry out their duty of informing the public without the fear of punishment. These purposes must be interpreted with the Clause’s roots in the English parliamentary privilege in mind. Rather than allowing the Clause to protect congressional action simply because legislators happened to discuss it aloud or put it in writing, the Clause must be understood in a much more limited sense. Instead, the Clause should be interpreted in a fashion that only protects as legislative action statements that could become the basis of a libel suit or speech-based crime.

people like “Sir William Williams, who was punished out of Parliament for what he had done in Parliament” (citation omitted)).

198. See id. at 74–75 (explaining that the Speaker of the House of Commons, Sir William Williams, was convicted of seditious libel, and the parliamentary privilege in the 1689 English Bill of Rights was created to prevent such injustice in the future).

199. See Kilbourn v. Thompson, 103 U.S. 168, 202 (1880) (indicating that because the Speech or Debate Clause has nearly the same language as that of the English Bill of Rights, the Framers likely intended a similar meaning).

200. See supra notes 156–77 and accompanying text (discussing statements by James Wilson, Thomas Jefferson, and James Madison regarding the intended purposes of the Speech or Debate Clause).

201. See supra notes 159, 166 and accompanying text (indicating that the Framers sought to protect legislators from suit based on the content of their speech).

202. See supra notes 163–73 and accompanying text (providing statements from Jefferson about the role of the Speech or Debate Clause in encouraging communication between legislators and their constituents).

203. See supra notes 178–99 and accompanying text (discussing the parliamentary privilege).

204. Supra notes 178–99 and accompanying text.

205. Supra notes 178–99 and accompanying text.
The Rayburn and Renzi decisions show that the current interpretation of the legislative privilege has deviated from the Framers’ “original understanding” of the Clause.\(206\) In finding a document nondisclosure privilege, the Rayburn court stated that “the purpose of the Speech or Debate Clause is ‘to insure that the legislative function the Constitution allocates to Congress may be performed independently,’ without regard to the distractions of private civil litigation or the periods of criminal prosecution.”\(207\) This focus on avoiding distraction to promote legislative independence is misplaced.\(208\) The legislative independence that the Framers sought to protect was grounded in the goal of encouraging legislators to fulfill their duty of open communication with their constituents.\(209\) This type of legislative independence necessitates a privilege that protects congressmen from speech-based suits, such as libel.\(210\) The Rayburn privilege, in contrast, allows a legislator who has misused his position of power to potentially use the privilege to avoid liability.\(211\) Rayburn gives legislators an opportunity to shield from review documents that may contain evidence of wrongdoing merely because the documents mention a “legislative act.”\(212\)

\(206\) See Bobbitt, supra note 125, at 9 (stating that “[h]istorical arguments depend on a determination of the original understanding of the constitutional provision to be construed”).

\(207\) See United States v. Rayburn House Office Bldg., 497 F.3d 654, 660 (D.C. Cir. 2007) (citations omitted).

\(208\) See id. (stating that the purpose of the privilege is to promote legislative independence by protecting legislators from the “the distractions of private civil litigation or the periods of criminal prosecution” (citations omitted)).

\(209\) See supra notes 163–73 and accompanying text (providing statements from Jefferson about the role of the Speech or Debate Clause in encouraging communication between legislators and their constituents).

\(210\) See supra notes 159, 166 and accompanying text (indicating that the Framers sought to protect legislators from suit based on the content of their speech).

\(211\) See Rayburn, 497 F.3d at 662 (indicating that in an investigation into misconduct, legislators must be given a chance to review documents for privilege before their forced disclosure to the Executive).

\(212\) See id. (stating that the Speech or Debate Clause requires a congressman to be given an “opportunity to identify and assert the privilege with respect to legislative materials before their compelled disclosure to Executive agents”).
Allowing legislators to refuse to disclose documents in an investigation of their misconduct is a far broader interpretation of the Clause than is necessary to fulfill the Framers’ goals. A privilege that could allow legislators to hide documents linked to crime goes beyond what is needed to protect congressmen from punishment for the content of speech that may be deemed unpopular or a misstatement of fact. Moreover, this opportunity for legislators to conceal their wrongdoing is the antithesis of the Framers’ aim of facilitating voters’ access to information.

Unlike the Rayburn court, the Renzi court permitted discovery of documents containing information about legislative acts in an investigation of nonprivileged action. The court, citing Brewster, stated that the purpose of the Clause is to “preserve the independence and thereby the integrity of the legislative process” and that “financial abuses by way of bribes . . . would gravely undermine legislative integrity and defeat the right of the public to honest representation.” The court stated that preventing investigation by the Executive and punishment by the Judiciary in such cases is “unlikely to enhance legislative independence.”

As already mentioned in a quote by Madison, the Clause must be interpreted in light of its

213. See supra notes 200–05 and accompanying text (discussing the Framers’ goals for the Speech or Debate Clause).

214. See supra note 174 and accompanying text (quoting James Madison for the proposition that the Speech or Debate Clause should be interpreted according to its “reason and necessity” (citation omitted)); supra note 172 and accompanying text (providing quotations by Jefferson about the intended purpose of the Speech or Debate Clause).

215. See supra notes 163–73 and accompanying text (discussing Jefferson’s statements indicating that the Speech or Debate Clause was created to encourage legislators to fulfill their duty of communication with their constituents without the fear of punishment based on their statements).

216. See United States v. Renzi, 651 F.3d 1012, 1036 (9th Cir. 2011) (indicating that the Speech or Debate Clause does not prohibit review of “documentary ‘legislative act’ evidence” when it is done “as part of an investigation into unprotected activity” (citation omitted)).

217. Id. (quoting Brewster v. United States, 408 U.S. 501, 524–25 (1972)).

“reason and necessity.”\textsuperscript{219} The Renzi court correctly realized that applying the Clause in a manner that would enable congressmen to hide misconduct, such as bribery, is not necessary to fulfill the Clause’s purposes.\textsuperscript{220} A nondisclosure privilege for documents in an investigation into legislative transgression is not needed to protect legitimate legislative speech or to encourage properly motivated legislative communication with the people.\textsuperscript{221} The Renzi court’s refusal to grant an unlimited nondisclosure privilege indicates that it also correctly considered James Madison’s warning about the possibility of congressional abuse of the privilege.\textsuperscript{222}

Renzi, however, in dicta, cited Eastland for its statement that “[w]hen the Clause bars the underlying action, any investigation and litigation serve only as wasted exercises that unnecessarily distract Members from their legislative tasks.”\textsuperscript{223} This statement suggests that legislators can refuse to disclose documents by simply asserting that the underlying investigation relates to an action protected by the legislative privilege.\textsuperscript{224} Because of the expansive definition of legislative acts,\textsuperscript{225} it will not be difficult for legislators to obtain this classification for their actions and prevent disclosure of documents.\textsuperscript{226} Such an

\textsuperscript{219}. CHAFETZ, supra note 149, at 88 (citation omitted).

\textsuperscript{220}. See Renzi, 651 F.3d at 1036 (indicating that the legislative privilege does not include document nondisclosure when the underlying action is not privileged).

\textsuperscript{221}. See supra notes 200–02 and accompanying text (giving the Framers’ intended purposes of the Speech or Debate Clause).

\textsuperscript{222}. See United States v. Renzi, 651 F.3d 1012, 1036 (9th Cir. 2011) (indicating that documents containing references to legislative acts can be reviewed when the underlying action is not privileged); see also CHAFETZ, supra note 149, at 88 (giving James Madison’s warning that the legislative privilege was abused in English precedent and the American privilege could similarly be abused).

\textsuperscript{223}. Renzi, 651 F.3d at 1036 (citations omitted).

\textsuperscript{224}. See id. (stating that when the underlying action is privileged “any investigation and litigation” is prohibited as “[t]hey work only as tools by which the Executive and Judiciary might harass their Legislative brother”).

\textsuperscript{225}. See supra Part II.A (discussing the broad interpretation of legislative acts).

\textsuperscript{226}. See Renzi, 651 F.3d at 1036 (“When the Clause bars the underlying action, any investigation and litigation serve only as wasted exercises that
interpretation of the Clause extends beyond the Framers' intended protection against speech-based suits and is the opposite of the Framers' aim of supporting legislative communication with the public.227

C. The Speech or Debate Clause and the Need for Legislative Accountability

Lastly, the scope of the Speech or Debate Clause and the permissibility of a document nondisclosure privilege should be considered pursuant to the “ethical argument” of constitutional analysis.228 This type of argument resolves constitutional issues by choosing solutions that “comport[] with the sort of people we are and the means we have chosen to solve political and customary constitutional problems.”229 Stated differently, this argument “uses the character, or ethos, of the American polity... as the source from which particular decisions derive.”230

It is beyond dispute that one such “ethos” or “ethical commitment[]” of the American polity is the notion of self-government.231 This is made clear in the Declaration of Independence, which is a reliable source for determining the values on which America was founded.232 The second paragraph of the Declaration states “[t]hat to secure these rights,

unnecessarily distract Members from their legislative tasks.” (citations omitted)).

227. See supra notes 200–02 and accompanying text (giving the Framers' intended purposes of the Speech or Debate Clause).
228. See BOBBITT, supra note 125, at 93–99 (discussing the “ethical argument” of constitutional analysis).
229. Id. at 95.
230. Id. at 94.
231. See Phillip Bobbitt, Constitutional Law and Interpretation, in A COMPANION TO PHILOSOPHY OF LAW AND LEGAL THEORY 126, 135 (Dennis Patterson ed., 1996) (“A corollary to this idea is reflected in the ethos of self-government and representative institutions.”).
232. See id. (“The proto-constitutional document, the Declaration of Independence, manifests these ethical commitments.’); see also James R. Stoner, Jr., Is There a Political Philosophy in the Declaration of Independence?, INTERCOLLEGIATE REV. 3, 8 (2005) (stating that the Declaration of Independence “assert[s] a right to self-government”).
Governments are instituted among Men, deriving their just powers from the consent of the governed.”233 In the very next sentence, the Declaration indicates that the people have the power to destroy the government if it fails to protect their rights.234 These statements clearly indicate a commitment to self-government.235

Additionally, the portion of the Declaration enumerating the wrongs of the English King shows the importance of self-government.236 For example, the Declaration admonished the King because he “refused to pass other Laws for the accommodation of large districts of people, unless those people would relinquish the right of Representation in the Legislature.”237 The Declaration then compares this refusal by the King to an act of a tyrant.238 The Declaration also indicates that the King “dissolved Representative Houses repeatedly” and he “called together legislative bodies at places unusual, uncomfortable, and distant from the depository of their Public Records, for the sole purpose of fatiguing them into compliance with his measures.”239 By taking away the people’s representation, the King obliterated self-government.240 Because the Declaration of Independence gives the King’s infringement on the Colonies’ self-government as a basis for the American Revolution, it is clear that the concept of self-government is an extremely important “ethos[] of the

233. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776) (emphasis added).
234. Id. (giving the people the right to destroy a government that is not serving their interests).
235. See Stoner, supra note 232, at 8–9 (indicating that the Declaration of Independence stands for individuals’ right to self-government); cf. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776) (explaining the rights of Man and the role of government in enforcing those rights).
236. Cf. THE DECLARATION OF INDEPENDENCE paras. 3–28 (U.S. 1776) (enumerating the wrongs of the English King).
237. Id. para. 5 (emphasis added).
238. See id. (comparing the King’s action to that of a tyrant).
239. Id. paras. 6–7.
240. Cf. id. paras. 3–28 (discussing the King’s infringement on Colonial America’s self-government).
American polity. Thus, it must be considered when making constitutional determinations.

Today, American self-government is achieved through a representative system in which the people elect legislators who will promote their interests. The integrity of the legislature is essential to sustaining legitimate self-government. It is, therefore, imperative that legislators be held accountable for misconduct. Because the Speech or Debate Clause provides legislators with a privilege that could potentially be used to hide legislative transgressions, it must be interpreted in a manner that eliminates, or at least reduces, the chances of such a result.

By holding that, pursuant to the Speech or Debate Clause, legislators must be permitted to “assert the privilege prior to disclosure of privileged materials to the Executive,” Rayburn’s interpretation fails to adequately account for the need for legislative accountability in a system of self-government. This holding gives the legislator who is accused of wrongdoing vast amounts of control over which documents the government is able to obtain. Even though the court states that the legislators’ assertions of privilege can be reviewed by the courts, savvy legislators will be able to avoid disclosure of potentially

241. See BOBBITT, supra note 125, at 94.
242. See id. at 93–99 (discussing the “ethical argument” of constitutional analysis).
243. See Ray, supra note 157, at 428 (stating the very important interest of “constituents in a fair and honest legislature”).
244. See id. (explaining that because “voters alone cannot purge their Congress of its malefactors,” legislative accountability must be obtained through other avenues).
245. See id. at 389 (explaining that the Speech or Debate Clause “provide[s] sweeping protection for members of Congress from investigation by the other branches of government”).
247. See Rothrock, supra note 61, at 759 (“By placing the initial, potentially unlimited control of Speech or Debate Clause privilege in the hands of the legislator being investigated, the Rayburn Court radically changed the distribution of the most essential investigatory element—information.”).
248. See Rayburn, 497 F.3d at 662 (stating that Jefferson did not argue “that his assertions of privilege could not be judicially reviewed”).
incriminating documents by simply inserting references to legislative material protected by the privilege.249 Rayburn’s extension of the phrase “shall not be questioned” will make investigations into the misconduct of legislators much more difficult.250 This will enhance corrupt legislators’ ability to escape liability for their betrayal of the public trust. If legislators are permitted to use their position in government for personal gain without reproach, they are no longer working toward the interests of their constituents. Like when the King interfered with representative government in Colonial America, self-government will be impinged if legislators are no longer working toward the goals of the people they have been chosen to represent.251

Unlike Rayburn, the Renzi decision does not provide a document nondisclosure privilege when the underlying action is not privileged.252 The court, however, suggested that legislators cannot be forced to disclose documents when the underlying action is a privileged act.253 This interpretation is much narrower than the Rayburn holding.254 Yet it still provides an avenue for legislators to avoid accountability for impermissible actions. Because of the broad understanding of what qualifies as a protected legislative act, legislators can refuse to disclose documents revealing unacceptable practices if the underlying

249. See id. (stating that legislators must be given an “opportunity to identify and assert the privilege with respect to legislative materials before their compelled disclosure to Executive agents”).

250. See Rothrock, supra note 61, at 759 (indicating that a document nondisclosure privilege gives legislators control over which information the government can obtain in an investigation).

251. See supra notes 236–42 and accompanying text (discussing the King’s wrongdoing as set forth in the Declaration of Independence).

252. See United States v. Renzi, 651 F.3d 1012, 1036 (9th Cir. 2011) (indicating that “documentary ‘legislative act’ evidence” can be reviewed as “part of an investigation into unprotected activity” (citation omitted)).

253. See id. (stating that “[w]hen the Clause bars the underlying action, any investigation and litigation serve only as wasted exercises that unnecessarily distract Members from their Legislative tasks” (citations omitted)).

254. See United States v. Rayburn House Office Bldg., 497 F.3d 654, 663 (D.C. Cir. 2007). (holding “that a search that allows agents of the Executive to review privileged materials without the Member’s consent violates the Clause”).
action is sufficiently linked to a permissible legislative function.\textsuperscript{255}

There are a number of disturbing Supreme Court cases wherein questionable legislative conduct was privileged because it qualified as a legislative act.\textsuperscript{256} For example, in Doe \textit{v. McMillan}, the Committee on the District of Columbia compiled a report containing “copies of absence sheets, lists of absentees, copies of test papers, and documents relating to disciplinary problems of certain specifically named students.”\textsuperscript{257} The children’s parents brought suit for a violation of their “children’s statutory, constitutional and common-law rights to privacy.”\textsuperscript{258} The court found liability immunity for the legislators because “their actions [were] within the ’legislative sphere.’”\textsuperscript{259} This finding was based on the argument that such an investigation is an “’integral part of the deliberative and communicative processes by which Members participate in committee and House proceedings with respect to the consideration and passage or rejection of proposed legislation.’”\textsuperscript{260} Thus, the Court prohibited investigation into a potential violation of the Constitution and criminal and civil statutes simply because the conduct could be classified as a legislative act.\textsuperscript{261}

Similarly, in \textit{Eastland}, questionable congressional conduct could not be investigated because it was found to be within “the

\begin{itemize}
\item \textsuperscript{255} See Renzi, 651 F.3d at 1036 (indicating that when the underlying action is privileged, legislative act documents cannot be reviewed by the Executive).
\item \textsuperscript{256} See e.g., Eastland \textit{v. U.S. Servicemen's Fund}, 421 U.S. 491, 495, 507 (1975) (finding that a subpoena alleged to violate complainants’ First Amendment rights was privileged by the Speech or Debate Clause); Doe \textit{v. McMillan}, 412 U.S. 306, 309, 312–13 (1973) (finding an investigation alleged to have violated the complainants’ “statutory, constitutional, and common-law rights to privacy” was privileged because it was “within the ‘legislative sphere’” (citation omitted)).
\item \textsuperscript{257} Doe, 412 U.S. at 308.
\item \textsuperscript{258} \textit{Id.} at 309.
\item \textsuperscript{259} \textit{Id.} at 312 (citation omitted).
\item \textsuperscript{260} \textit{Id.} at 313 (citation omitted).
\item \textsuperscript{261} See \textit{id.} at 312–13 (stating that “acts within the ‘legislative sphere’” are privileged “even though [the] conduct, if performed in other than legislative contexts would in itself be unconstitutional or otherwise contrary to criminal or civil statutes” (citation omitted)).
\end{itemize}
‘legitimate legislative sphere.’”262 In Eastland, a subpoena was issued to a bank that demanded the bank produce “any and all records appertaining to or involving the account or accounts of [United States Servicemen’s Fund].”263 United States Servicemen’s Fund (USSF) brought suit, stating that the subpoena was issued for an improper purpose.264 USSF also claimed the subpoena violated its members’ First Amendment rights.265 The Court, however, found that the issuers of the subpoena were protected by the privilege because it was issued according to an “authorized investigation” and was “in furtherance of a legitimate task of Congress.”266 Consequently, like in Doe, an alleged violation of the Constitution by congressmen could not be examined because it was labeled a legislative act.267

Permitting legislative immunity and document nondisclosure in circumstances such as those in Doe and Eastland diminishes legislative accountability in two ways.268 First, the immunity privilege prevents civil and criminal suits for legislative wrongdoing. Second, document nondisclosure limits investigations into wrongful legislative conduct, which restricts voters’ access to such information. Each of these consequences impair the legitimacy of self-government by allowing legislators to commit actions that are contrary to the interests of the people they represent.

263. Id. at 494.
264. See id. at 495 (stating that the complainants argued that the subpoena was issued for an improper purpose).
265. See id. at 492–93 (“We granted certiorari to decide whether a federal court may enjoin the issuance by Congress of a subpoena duces tecum that directs a bank to produce the bank records of an organization which claims a First Amendment privilege status . . . .”).
266. Id. at 505–06 (citation omitted).
267. See id. at 507 (“We conclude that the Speech or Debate Clause provides complete immunity for the Members for issuance of this subpoena.”).
V. Proposed Solution: Limiting the Speech or Debate Clause and Increasing Legislative Accountability

To narrow the scope of the Speech or Debate Clause so that it more closely reflects its text, the intent of the Framers, and the ethos of self-government, this Note proposes a two-part solution. First, the issue of whether “shall not be questioned” includes a nondisclosure privilege must be resolved. Second, the legislative acts that are protected under the phrase “Speech or Debate” must be narrowed.

A. Document Nondisclosure and the Scope of the Phrase “shall not be questioned in any other Place”

This Note proposes that the Renzi holding regarding document disclosure be adopted. Specifically, when the basis of the suit is not protected by the privilege, then the Clause should not permit document nondisclosure.269

First, this will bring the Speech or Debate Clause more in line with the actual meaning of the words used in the text.271 As discussed in Part IV.A, the contemporary meaning of the word “question” requires that it be directed at an individual.272 This definition forecloses the application of the Clause to documents.273 This test will greatly reduce legislative document nondisclosure and the privilege will more closely conform to the Clause’s actual language.274

269. See United States v. Renzi, 651 F.3d 1012, 1036 (9th Cir. 2011) (providing the case’s holding).
270. See id. (stating that “documentary ‘legislative act’ evidence” can be reviewed “as part of an investigation into unprotected activity” (citation omitted)).
271. See supra Part IV.A (interpreting the Speech or Debate Clause using a textual approach to constitutional interpretation).
272. See supra notes 138–39 and accompanying text (providing definitions of the word “question”).
273. See supra notes 136–43 and accompanying text (discussing the meaning of the word “question”).
274. See supra Part IV.A (discussing the Speech or Debate Clause’s text).
LIMITING THE LEGISLATIVE PRIVILEGE

Second, this standard adequately protects the ethos of self-government. By reducing legislators' ability to hide documents containing information about misconduct, legislative accountability will be bolstered. This standard should increase criminal and civil prosecution for wrongdoing and improve voters' access to information about corrupt legislators. As a result, the integrity of the legislature will be improved and self-government through representatives acting in good faith can be achieved.

B. Narrowing the Meaning of Legislative Acts

To obtain the protection of the privilege, a legislator’s action must be within the meaning of the phrase “Speech or Debate.” The Supreme Court has indicated that qualifying legislative acts will be privileged. This Note proposes a new test for the meaning of legislative acts that will greatly reduce the number of privileged activities. In order to be classified as “speech or debate,” legislative action should be conduct that could be the cause of a speech-based suit, such as libel. For example, under this test protected action would include legislators’ statements of opinion, policy, or belief that could later become the basis of a libel suit. If an action does not meet this test, it should not receive the benefit of the Speech or Debate privilege.

First, such a test adheres to the plain meaning of the words “speech” and “debate” in the Clause. Second, this complies with the intended meaning of the Clause as set forth by the Framers. The Framers included the Clause to promote legislators’ communication with their constituents and to protect legislators from punishment for their statements, views, and

275. See supra Part IV.C (discussing the ethos of self-government).

276. See supra Part II.A (discussing actions that are protected by the phrase “Speech or Debate”).

277. See supra Part II.A (discussing actions that are protected by the phrase “Speech or Debate”).

278. See supra Part IV.A (interpreting the Speech or Debate Clause through a textual analysis of constitutional interpretation).

279. See supra Part IV.B (discussing the Framers’ understanding of the Speech or Debate Clause).
opinions that could later be alleged untrue, libelous, or contrary to the law in some other way.\textsuperscript{280} This test accurately reflects these aims of the Framers.\textsuperscript{281}

Finally, the test will promote the interest in self-government, which requires that representatives be held accountable for wrongdoing.\textsuperscript{282} By requiring conduct to be linked to a threat of speech-based punishment, the circumstances in which legislators will be able to avoid document nondisclosure, based on a privileged underlying action, will be greatly reduced. This will help prevent the troubling results in cases such as \textit{Doe} and \textit{Eastland}.\textsuperscript{283} For example, in \textit{Doe} the investigation and report of children’s school records, which potentially violated the complainants' right to privacy, would not have qualified as a legislative act.\textsuperscript{284} Such an \textit{investigation} could not lead to a speech-based suit. Similarly, the \textit{Eastland} subpoena would not have qualified as a privileged legislative act under the proposed standard.\textsuperscript{285} Reducing the types of legislative acts protected by the privilege will thereby increase legislative accountability and improve the legitimacy of self-government.

\textbf{VI. Conclusion}

In \textit{Rayburn}, Congressman William J. Jefferson was charged with using his position as a legislator for personal gain through the acceptance of bribes.\textsuperscript{286} In spite of his alleged abuse of power, he was permitted to have documents returned to him because of a

\begin{itemize}
\item \textsuperscript{280} See supra Part IV.B (discussing the reasons the Framers included the Speech or Debate Clause).
\item \textsuperscript{281} See supra Part IV.B (discussing the aims of the Framers when they included the Speech or Debate Clause).
\item \textsuperscript{282} See supra Part IV.C (discussing the ethos of self-government).
\item \textsuperscript{283} See supra notes 256–68 and accompanying text (discussing \textit{Doe} and \textit{Eastland}).
\item \textsuperscript{284} See supra notes 256–61 and accompanying text (discussing the disturbing result in \textit{Doe}).
\item \textsuperscript{285} See supra notes 262–67 and accompanying text (discussing the disturbing result in \textit{Eastland}).
\item \textsuperscript{286} See United States v. Rayburn House Office Bldg., 497 F.3d 654, 656 (D.C. Cir. 2007) (discussing Jefferson’s involvement in a bribery scheme).
\end{itemize}
legislative privilege violation. Thus, the D.C. Circuit created a privilege that facilitates legislators’ ability to escape civil and criminal liability. Renzi attempted to eliminate this counterintuitive result by stating that when an investigation relates to nonprivileged action, forced document disclosure is permissible. Its decision, however, does not go far enough to prevent legislators’ use of the privilege to hide questionable conduct. By leaving the possibility of document nondisclosure in cases where the underlying action is privileged, Renzi failed to remove a powerful tool that corrupt legislators can use to avoid liability. Because of the overly expansive definition of legislative acts, dishonest legislators will have little trouble linking an action that could become the subject of investigation to a permissible legislative function. This will prevent punishment of legislative wrongdoing and leave the voters uninformed.

The proposal set forth in this Note should be adopted to help combat legislative malfeasance. By limiting the privilege to action that could be the cause of a speech-based suit, the quantity of privileged legislative action will be reduced. Additionally, a privilege that does not include document disclosure in misconduct investigations will enhance legislative accountability. This will

287. See id. at 666 (holding “that the Congressman is entitled to the return of all legislative materials . . . that are protected by the Speech or Debate Clause”).
288. See id. at 663 (holding that the legislative privilege includes document nondisclosure).
289. See United States v. Renzi, 651 F.3d 1012, 1036 (9th Cir. 2011) (permitting document nondisclosure when the underlying action is not privileged).
290. See id. (permitting document nondisclosure when the underlying action is not privileged).
291. See id. (indicating that when the underlying action is privileged, legislative act documents cannot be reviewed by the Executive).
292. See supra Part IV.C (discussing the disturbing results in Doe and Eastland because of the overly broad legislative act definition).
293. See supra Part V (proposing a new method of applying the Speech or Debate Clause).
294. See supra Part V (discussing a new method of interpreting the Speech or Debate Clause).
295. See supra Part IV.C (discussing the need for legislative accountability).
increase the integrity of representative government and fulfill the *ethos* of self-government upon which the United States was founded.\textsuperscript{296}

\textsuperscript{296} See *supra* Part IV.C (discussing the *ethos* of self-government).
The Real Error in *Citizens United*

Joanna M. Meyer*

*Table of Contents*

I. Introduction ................................................................. 2172

II. Corporate Form Pre-*Citizens United* .................. 2178
   A. Entity and Aggregate Theories of Corporate Form ....... 2178
   B. The Circularity of Defining a Corporation by
      Inherent Characteristics .................................. 2182
   C. The Stance of the Courts ..................................... 2186

III. *Citizens United* .......................................................... 2190
   A. Background ......................................................... 2190
   B. The Opinion ...................................................... 2194
   C. Public Impact .................................................... 2196
   D. Legal Impact ..................................................... 2201

IV. Corporate Theory After *Citizens United* .......... 2205
   A. The Public’s Misuse of “Person” ......................... 2206
   B. Looking for Theoretical Language in
      *Citizens United* ............................................. 2210
       1. Justice Kennedy’s Majority Opinion ............... 2210
       2. Justice Scalia’s Concurring Opinion ............. 2216
       3. Justice Stevens’s Dissenting Opinion .......... 2218
   C. What *Citizens United* Makes Clear ................. 2220
   D. The Real Source of the Corporate Right in
      *Citizens United* ............................................ 2222

V. Proper Grounds for Criticizing *Citizens United* ... 2225
   A. Attacking the Court’s View of Policy ............... 2225

* Candidate for J.D., Washington and Lee University School of Law, May 2013. I thank Professor Christopher Bruner and Professor David Millon for their great help as advisors for this Note. I dedicate this, my first publication, to John Williamson and the faculty at Highlands High School, the Philosophy Department at the University of Virginia, and Professor Adam Scales and the faculty at Washington and Lee School of Law for being influential teachers.
Corporations typically retain statutory personal rights with respect to commercial activity, such as the rights to own property, to sue and be sued, and to enter into contracts. But the extent to which corporations have constitutional personal rights, such as the right to free speech, is the source of debate among courts, legislatures, citizens, the legal academy, and, of course, corporations.

The academic community has analyzed corporations—particularly their essence, rights, and purposes—throughout their evolution. Corporate law encompasses many theories, and virtually every theory either aims to answer any of three basic questions or presumes a position on those questions in order to tackle a more nuanced issue. These three questions, therefore, are a defining feature of corporate theory as an area of study.

First, what is a corporation’s essence? Most theories describe a corporation’s essence in terms of one of two metaphors: an entity that exists separately from its shareholders or a mere aggregation of the corporation’s constituents. This Note will refer to this distinction as the corporate-essence dichotomy.

1. See, e.g., Del. Code Ann. tit. 8, § 122(4) (2012) (providing a corporation’s right to own property); id. § 122(2) (stating that a corporation may sue and be sued); id. § 122(13) (providing a corporation’s right to contract); see also 1 U.S.C. § 1 (2012) (stating that, in congressional acts, the word “person” presumptively includes corporations).

2. See U.S. Const. amend. I (“Congress shall make no law . . . abridging the freedom of speech, or of the press . . .”).

3. Although this Note categorizes the theories by how they answer certain questions, scholars conceptualize corporate theories in many different ways because the theories and their interrelation are complicated. See, e.g., Susanna K. Ripken, Corporations Are People Too: A Multi-Dimensional Approach to the Corporate Personhood Puzzle, 15 Fordham J. Corp. & Fin. L. 97, 97 (2009) (providing a different organization of theories).

4. See David Millon, Theories of the Corporation, 1990 Duke L.J. 201, 201
Second, what are the rights of a corporation? Many theories presume or posit that corporations represent either artificial creations of state law, thereby only possessing such rights as are explicitly granted by statute, or natural products of private initiative, thereby possessing rights beyond those granted by statute. This Note will refer to this distinction as the corporate-rights dichotomy.

Third, what are the purposes and goals of corporations? Some theories argue that corporations are mechanisms purely for private investment, aiming to maximize shareholder wealth. Other theories assert that corporations are vehicles both for private investment and for broader, public purposes, considering also the betterment of the economy and society. This Note will refer to this distinction as the corporate-purpose dichotomy. Often, intimately tied to this final question is an analysis of the balance of power between shareholders and the board of directors because that balance seems to indicate how corporations ought to be governed. This Note, however, will not discuss particular governance theories because they implicitly answer one of the three main questions common to virtually all corporate-law
Because of the importance of personal rights and the public attention to corporate personal rights, this Note focuses on the second of the three questions—the corporate-rights dichotomy—but also heavily implicates the first question, the corporate-essence dichotomy.

In 2010, *Citizens United v. FEC* stirred debate of corporations’ personal rights under the U.S. Constitution. The U.S. Supreme Court concluded, first, that “the Government may not suppress political speech on the basis of the speaker’s corporate identity.” Second, the Court overruled a law that prohibited corporate independent campaign expenditures on the grounds that the law was a ban on speech violating the First Amendment. The Court had already held, in a previous case, that speech could not be restricted solely because the speaker was a corporation, but the effect of *Citizens United*—that corporations may now spend unlimited funds advocating the success or defeat of candidates under the protection of the First

---

8. For example, the nexus-of-contracts theory, which views the corporate power structure as an aggregation of private interests, implicitly answers the corporate-essence question: the corporation must be an aggregation, not an entity. See Stefan J. Padfield, *The Dodd-Frank Corporation: More Than a Nexus-of-Contracts*, 114 W. Va. L. Rev. 209, 215 (2011) (“The aggregate theory is generally understood to capture the nexus-of-contracts view, the artificial-entity theory captures concession theory, and the real-entity theory arguably captures the director-primacy view of the corporation.” (citation omitted)); Reuven S. Avi-Yonah, *Citizens United and the Corporate Form*, in *ACCOUNTING, ECONOMICS, AND LAW* 1, 25 n.142 (2011) (“The point that the nexus-of-contracts theory is a reinvention of the aggregate view has been made repeatedly.”).


10. *See id.* at 913–16 (stating that the government could not suppress political speech on the basis of the speaker’s corporate identity and that a federal statute barring corporate expenditures for campaigns violated the First Amendment).

11. *Id.* at 913.

12. *Id.* at 896–98.

13. *See id.* at 913 (relying on previous U.S. Supreme Court holdings that restricting campaign expenditures amounts to a restriction on speech and that the government may not restrict corporate speech); *First Nat’l Bank of Bos. v. Bellotti*, 435 U.S. 765, 776 (1978) (holding that a Massachusetts statute prohibiting certain corporate political expenditures violates the First Amendment).
Amendment—has itself sparked considerable discourse among the media and the general public.\textsuperscript{14}

One source of public concern is that corporate entities can overwhelm the voices of individual citizens by pouring their amassed wealth into lobbying efforts and media ownership.\textsuperscript{15} Indeed, if corporations dominate both the media and elected offices, they can essentially control policy and law. Even private business owners or wealthy individuals typically cannot rival a large corporation’s resources.\textsuperscript{16} Such power is self-perpetuating because corporations can then use their influence to enhance their own wealth-seeking interests.\textsuperscript{17} Thus, members of the public often fear that corporations will have a corruptive influence on government officials.\textsuperscript{18}


\textsuperscript{15} See, e.g., Richard L. Hasen, Campaign Finance Laws and the Rupert Murdoch Problem, 77 TEX. L. REV. 1627, 1628 (1999) (listing major corporations that own corporate media subsidiaries such as ABC, NBC, and FOX News).


\textsuperscript{17} See Supplemental Brief Amicus Curiae of the Reporters Committee for Freedom of the Press in Support of Appellant, Citizens United v. FEC, 130 S. Ct. 876, 904 (2010) (No. 08-205), 2009 WL 2219299 at *2–3 (“The original, laudable intent of Congress presumably was to limit speech by corporations that seek to promote their own interests by influencing elections, while continuing to allow all other commentary (either non-corporate entities or by the news media) on political issues.”).

\textsuperscript{18} See Citizens United, 130 S. Ct. at 908 (addressing and rejecting the argument that the government has a sufficiently important interest in preventing corruption, based on corporate independent expenditures, of elected officials).
In *Citizens United*, however, the Federal Judiciary—the only Branch structurally insulated from political influence—defended corporations’ right to freely spend their funds on electioneering communications under the highest protection of the law. The U.S. Supreme Court rejected its earlier policy-based position against the distorted corporate influence and political corruption that would likely accompany unlimited corporate campaign expenditures. The decision leaves individual citizens and the legal community with the question: What kind of person is a corporation, such that it can speak under protection of the Constitution?

Central to the debate, and the focus of this Note, is how *Citizens United* has affected corporate personhood in theory and in practice. In particular, the public’s criticism of the 2010 case amplifies the flaw in the idea that we can deduce traits from a corporate person or essence—a notion that is fundamental to corporate-rights theories. Additionally, *Citizens United* indicates the U.S. Supreme Court’s stance on justifying corporate rights.

This Note focuses exclusively on publicly held businesses because the corporate rights debate primarily raises issues with the influence and standard characteristics of large-scale enterprises. The idea of corporate personhood does not raise the same questions with closely held companies whose owners also retain control over decisionmaking.

Part II of this Note discusses theories about characteristics of the corporate form prior to *Citizens United*, arising out of the entity and aggregate theories of corporate form (the corporate-

---


20. See *Citizens United*, 130 S. Ct. at 913 (applying constitutional protections and a strict-scrutiny analysis to state that corporate independent expenditures are permissible).

21. See id. at 902–11 (overruling precedent that permits a ban on political speech under the “antidistortion” and “anticorruption” rationales).

22. See infra Part II (discussing corporate-rights theories).

23. See Bruner, supra note 7, at 1386 n.1 (implying that issues arise from the separation of ownership and control, which is characteristic of publicly held businesses rather than those that are closely held).
THE REAL ERROR IN CITIZENS UNITED

essence dichotomy). 24 This Part then raises two important critiques of these theories that resurface in later discussion. 25 Part III analyzes the Citizens United opinion as it relates to corporate-rights theories and its repercussions in law and society. 26 Part IV revives the two critiques from Part II to reject the notion that rights can be deduced from any theoretical conception of the corporation. 27 Because no corporate-law theories can provide an appropriate source from which the U.S. Supreme Court can ascribe rights to the corporation, this Part then identifies the proper source. 28 Specifically, the opinion’s language suggests that the Court grounds the right to corporate political speech in public policy. 29 Finally, based on this refined reading of Citizens United, Part V provides more accurate bases for criticizing the opinion or otherwise addressing the public’s concerns. 30

Part VI concludes by reiterating the overarching argument of this Note: (1) Citizens United is unclear about the source of the corporate political speech right; (2) because the Court fails to make the source clear, public criticism of Citizens United is misguided; (3) the source for the constitutional corporate speech right is public policy, not some existential truth about the corporate form; and (4) real, functional claims about corporate rights only relate to corporate-rights theories in the sense that these theories are metaphors to discuss the real rights. 31

24. See infra Part II.A (discussing entity and aggregate theories).
25. See infra Part II.B (discussing the circularity problem).
26. See infra Part III (discussing Citizens United).
27. See infra Part IV (finding a fatal flaw in using entity and aggregate theories to deduce corporate rights).
28. See infra Part IV (finding the real source of the corporate speech right).
29. See infra Part IV (dissecting Citizens United to find the source of the corporate speech right).
30. See infra Part V (describing a proper analysis).
31. See infra Part VI (concluding that the Court failed in its duty to justify the right of corporate political speech by not clearly explaining its source).
By the nineteenth century, corporate America began to develop from joint stock trading companies and partnerships into quasi-public entities. As legislators and courts unequivocally granted corporate entities rights that, by statute, explicitly applied to persons, complex corporate-law questions led theorists to contemplate the ontology of the corporate legal personality. Generally, nineteenth-century theories conceptualized the corporate form either as an entity separate from its shareholders or as an aggregate of its members, and theorists attempted to deduce inherent rights or characteristics based on the corporate form.

Under the artificial-entity theory of corporate personhood, which “character[ized] . . . legal discourse for much of the 19th century,” the corporation is a mere “creature of the law, whose rights consist[] only of those conferred by the state.” The
corporate entity, under this view, is a legal fiction that exists only to the extent that statutory law prescribes.\textsuperscript{39} When courts have used language indicating an artificial-entity view, they typically have distinguished the corporate entity from its members to justify state regulation, rather than to imbue the corporation with rights.\textsuperscript{40}

In the second half of the nineteenth century, the advent of general incorporation laws facilitated widespread incorporation by simplifying the state-law requirements, making the corporation more universally available as a business form.\textsuperscript{41} At the same time, these laws signified a shift away from the idea of an artificial entity heavily regulated by the state.\textsuperscript{42} Theorists recognized corporations as “nothing but aggregations of private individuals,” as opposed to entities distinct from their constituents.\textsuperscript{43} Under this aggregation theory, corporations have rights derived from the rights of the natural persons behind the corporate veil.\textsuperscript{44} Although not consistently adopting the aggregate theory over entity theories, courts have applied constitutional protections to corporations on the grounds that “the members do not, because of such association, lose their rights to protection.”\textsuperscript{45}

\begin{thebibliography}{99}
\bibitem{39} See Trs. of Dartmouth Coll. v. Woodward, 17 U.S. 518, 636 (1819) (“Being the mere creature of law, [the corporation] possesses only those properties which the charter of its creation confers upon it . . . .”); Millon, \textit{supra} note 4, at 206 (“[T]he corporation owed its existence to the positive law of the state rather than to the private initiative of individual incorporators.”).
\bibitem{40} See, e.g., Hale v. Henkel, 201 U.S. 43, 74 (1906) (“[A corporation] has no right to refuse to submit its books [for state review, but an] individual may stand upon his constitutional rights as a citizen.”).
\bibitem{41} See Krannich, \textit{supra} note 36, at 75 (discussing how general incorporation laws “allowed individuals to incorporate their businesses without first seeking a special charter from the state legislature”); Avi-Yonah, \textit{supra} note 35, at 13 & n.75 (providing an example of general incorporation law).
\bibitem{42} See Krannich, \textit{supra} note 36, at 75 (“The immediate effect of the general incorporation acts was to ‘move[] the predominant role in corporate organization from the state to the incorporators and shareholders.’” (citation omitted)).
\bibitem{43} Millon, \textit{supra} note 4, at 202.
\bibitem{44} See Krannich, \textit{supra} note 36, at 77 (discussing how courts “imputed the corporation’s constitutional personhood from that of the individuals who had formed the corporation”).
\end{thebibliography}
The turn of the twentieth century saw significant developments in the corporate form, including corporations' acquisition of several legal features that make the public corporation an attractive entity form, as we know it today. For example, limited shareholder liability and free transferability of shares promote economic efficiency and investment by encouraging risk taking and reducing transaction costs. Similarly, corporations' perpetual existence incentivizes investment. Perhaps most significantly, companies with widely dispersed ownership vest control in boards of directors—rather than in their expanding base of shareholders, whose interest in the company becomes so insignificant that they, rationally, are apathetic about business decisions.

The ownership–control dichotomy, which divides the corporate identity between its directors and shareholders, coincided with the growing popularity of a third theory in legislative, judicial, and academic discourse about the corporate form. The idea of the corporation as a state-created artificial
entity gave way to a description of a corporation as an entity arising naturally by private initiative. Coupled with the separation of ownership and control, these theoretical developments “transformed shareholders from entrepreneurs into passive investors who placed their economic interests in the hands of professional managers.”

This advent of the natural-entity theory (or real-entity theory) accompanied corporations’ growth in size and pervasiveness, leading to a corporate structure that has persisted as “the dominant form of organization and production.” Professor Reuven Avi-Yonah has deemed this theory the most persistent of corporate-rights theories, partly because it is the “most congruent with business realities [such as the business judgment rule] as well as the [theory] most suited to a corporation–state balance.”

The artificial-entity theory, the aggregate theory, and the natural-entity theory each have sought to affirmatively describe the corporate form, and this debate has in turn prompted normative discussion among the academic community as to what corporations ought to be. For example, the political-speech right might appear to attach to corporations pursuant to a natural-entity perspective: corporations are entities whose speech the U.S. Supreme Court has deemed protected, so they have the

---

51. See Krannich, supra note 36, at 80 (describing the corporation as an autonomous being with existence independent of its shareholders and beyond state grant of authority); Millon, supra note 4, at 202 (“The aggregate characterization did not prove to be persuasive, but the notion of the corporation as a natural creation of private initiative and market forces replaced the idea that the corporation was artificial.”).

52. Millon, supra note 4, at 214–15.

53. Berle & Means, supra note 32, at xxv.

54. Avi-Yonah, supra note 8, at 19.

55. See Millon, supra note 4, at 204 (recognizing that “normative implications then are said to follow from the positive assertion [of what corporations are]”).
capacity for speech.\textsuperscript{56} Natural-entity theorists (and the general public) would then debate whether it is appropriate that corporations have the capacity for speech already found to be constitutionally protected.\textsuperscript{57}

The vast majority of corporate-theory scholars, in discussing corporations’ rights and the corporate form, adopt (at least implicitly) the aggregate theory or one of the two entity theories,\textsuperscript{58} despite each theory’s enduring inconsistencies.\textsuperscript{59}

\textbf{B. The Circularity of Defining a Corporation by Inherent Characteristics}

All three theories face a problem of circular reasoning. Because the entity theories and the aggregate theory all describe the essence of a corporation, they inevitably make claims about attributes or rights of corporations.\textsuperscript{60} Attributes such as limited shareholder liability, free transferability of shares, perpetual existence, and centralized governance may seem to clearly define the corporation. Corporations, however, like people, can have an indefinite number of characteristics, some of which seem inconsistent.\textsuperscript{61} For example, a corporation can sign a document

---

\textsuperscript{56} See Citizens United v. FEC, 130 S. Ct. 876, 913 (2010) ("[T]he Government may not suppress political speech on the basis of the speaker's corporate identity."); Avi-Yonah, supra note 8, at 41 ("The entire Citizens United opinion, both the majority and the dissent, are thus reflective of the real entity view. Corporations stand on their own, independent of both the state that created them and the shareholders that own them.").

\textsuperscript{57} See Millon, supra note 4, at 204 (stating that normative implications follow from positive assertions about corporations).

\textsuperscript{58} See Avi-Yonah, supra note 8, at 3 ("[T]hroughout all of [the changes in the legal conception of the corporation], spanning two millennia, the same three theories of the corporation can be discerned.").

\textsuperscript{59} See infra notes 66–72 and accompanying text (pointing out major flaws in the theories).

\textsuperscript{60} See H.L.A. Hart, Essays in Jurisprudence and Philosophy 42 (Oxford Univ. Press 1983) ("Theories of the traditional form . . . make the common assumption that [expressions for corporate bodies] must stand for or describe something, and then give separate and incompatible accounts of its peculiarity as a complex or recondite or a fictitious entity . . . ").

\textsuperscript{61} See id. ([T]he peculiarity lies . . . in the distinctive characteristics of expressions used in the enunciation and application of rules.").
(and we picture this as requiring a human agent)\textsuperscript{62} but cannot swear an oath (though we could imagine a human proxy on the stand).\textsuperscript{63} In fact, as corporations accumulate characteristics, inconsistencies suggest confusion in the line of questioning between asking about the corporation, on the one hand, and asking about the people who comprise it, on the other hand.\textsuperscript{64} Although some scholars offer explanations for the “oscillation between the three views” over time,\textsuperscript{65} the reality is that no single theory can consistently define a corporation.\textsuperscript{66}

Two inherent flaws prevent any entity or aggregate theory from fully and accurately defining a corporation. First, each of the two major categories distorts the corporate form in its own direction: entity theories have trouble explaining rights arising from the aggregate, and aggregate theories struggle with traits arising from a holistic entity.\textsuperscript{67} Under an entity theory, deriving personal rights on the corporation’s behalf may be difficult to

\begin{flushright}
\textsuperscript{62} See Del. Code Ann. tit. 8, § 142(a) (2012) (“Every corporation organized under this chapter shall have such officers with such titles and duties . . . necessary to enable it to sign instruments . . .”).
\end{flushright}

\begin{flushright}
\textsuperscript{63} See Bruner, supra note 7, at 1388 (providing the seventeenth century view that corporations do not possess bodies or souls and thus can neither “be in person, nor swear” an oath (quoting 5 Edward Coke, The Reports of Sir Edward Coke, Knt.: In Thirteen Parts pt. X, at *32b (photo. reprint 2002) (1826))).
\end{flushright}

\begin{flushright}
\textsuperscript{64} See Hart, supra note 60, at 43 (discussing group personalities, such as “[t]he crowd was angry”).
\end{flushright}

\begin{flushright}
\textsuperscript{65} Avi-Yonah, supra note 8, at 17. “The Court is trying to strike a balance between the rights of the corporations, which can best be protected under either the aggregate or the real entity views, and the regulatory power of the state, which is best reflected in the artificial entity view.” \textit{Id.} at 17–18.
\end{flushright}

\begin{flushright}
\textsuperscript{66} Hart provides examples of inconsistencies under the entity theories: [J]ust as a Realist theory appears to tell us that a company “cannot” be bound by an agreement empowering another company to direct its business and appoint its personnel because this would be “to degrade to the position of a tool” a person with a real will, so a Fiction theory appears to say that company “cannot” be guilty of certain crimes because it has no mind.
\end{flushright}

\begin{flushright}
\textsuperscript{67} See Hart, supra note 60, at 43 (“Under what conditions do we refer to numbers . . . of men as aggregates of individuals and under what conditions do we adopt instead unifying phrases extended by analogy from individuals?”).
\end{flushright}
explain; the only sense in which a corporation can act on its personal rights is via individual people, and if the corporation is in fact a separate entity it arguably should be regulated as such. 68 Under the aggregate theory, on the other hand, replacing a corporation’s entire membership would seem to form a new corporation. 69 As corporations have developed over time, each category of theories (aggregate theories and entity theories) faces the same obstacle: it must “simultaneously recognize[] entity and aggregate characteristics” 70 and be somewhat “schizophrenic about what ultimately the firm is.” 71

The second inherent flaw in corporate-rights theories is that defining a corporation by listing its characteristics presumes corporations have inherent attributes that resemble those of persons. 72 Recognizing the circularity in this reasoning, H.L.A. Hart concluded that statements about inherent characteristics do not define a corporation; that would beg the question by defining what a corporation is. 73 Applying Hart’s theory, no corporate essence exists from which to derive rights; rather, the legislatures and courts must say what rights a corporation has. 74 Therefore, we must abandon the notion that a corporation can be defined by its characteristics and instead ask the question: “Under what

---

68. See Cohen, supra note 47, at 435 (discussing the aggregate theory as responding to the problem of government regulation of the corporate entity).

69. See Arthur W. Machen, Jr., Corporate Personality, 24 Harv. L. Rev. 253, 259 (1911) (“Any group of men . . . whose membership is changing, is necessarily an entity separate and distinct from the constituent members.” (citation omitted)).


71. Cohen, supra note 47, at 436.

72. See HART, supra note 60, at 45 (“These statements [about corporations’ entity or ontological capacity] confuse the issue because they look like eternal truths about the nature of corporations given us by definitions . . . .”); Bruner, supra note 7, at 1388–89 (considering that there may not be an inherent soul-like ontology of a corporate entity).

73. See HART, supra note 60, at 42–47 (discussing the definition of a corporation).

74. See id. at 29–30 (analogizing the corporation to a game in which a rule “attach[es] a single consequence to the successive actions of a set of different men—as when a team is said to have won a game”).
types of conditions does the law ascribe liabilities”—and rights—
“to corporations?”

The circularity problem arises when answering the question
What is a corporation? by listing corporate rights as if they
naturally belong in the definition. The problem does not arise in
every discussion about corporate qualities. If the question is How
does a corporation function in the real world?, listing corporate
rights such as “corporations can sign contracts” is informative
and avoids the circularity problem. The claim is now helpful
because it identifies a corporate right. For example, identifying
which crimes a corporation is capable of committing reveals its
capacity as a legal device. Consider perjury: a corporate official
may lie under oath in his role as an agent of the corporation, yet
courts have held that the corporation is not a person capable of
taking an oath and cannot be criminally liable for the crime of
perjury. Therefore, legislatures and courts determine which
crimes a corporation could potentially commit.

Corporations seem suspect insofar as they retain, through
statutes or court decisions, certain humanlike capacities while
escaping moral and legal accountability for humanlike actions. If Congress and state legislatures (and courts, via judicial review)
make the rules, they cannot hide behind an ontological argument
to defend unpopular corporate traits. Legislatures and courts
seem unlikely to use these ontological theories as a cop-out,

75. Id. at 43. (internal quotation marks omitted).

76. See id. (stating that the discussion of the law ascribing rights to
corporations will “bring out the precise issues at stake when judges, who are
supposed not to legislate, make some new extension to corporate bodies of rules
worked out for individuals”).

77. See, e.g., State v. Saint Paul Fire & Marine Ins. Co., 835 So. 2d 230,
perjury); Note, Criminal Liability of Corporations for Acts of Their Agents, 60
HARV. L. REV. 283, 284 (1946) (discussing corporations’ capacity to commit
(2d Cir. 1983) (concluding that a corporation could be convicted of subscribing to
a false federal income tax return even though it has been held that corporations
are incapable of taking an oath).

78. “The structure of the public corporation ‘insulates shareholders from
social and moral sanctions and processes,’ both by rendering them ‘largely
anonymous’ to the public, as well as by virtue of their ‘relative lack of
information about how corporate operations may impact the public interest.’”
Bruner, supra note 7, at 1393 (quoting Einer Elhauge, Sacrificing Corporate
Profits in the Public Interest, 80 N.Y.U. L. REV. 733, 798 (2005)).
however, but rather as a legal defense of the corporate rights they determine. In a 1906 case, the U.S. Supreme Court stated:

[W]e do not wish to be understood as holding that a corporation is not entitled to immunity, under the Fourth Amendment, against unreasonable searches and seizures. A corporation is, after all, but an association of individuals under an assumed name and with a distinct legal entity. In organizing itself as a collective body it waives no constitutional immunities appropriate to such body. Its property cannot be taken without compensation. It can only be proceeded against by due process of law, and is protected, under the Fourteenth Amendment, against unlawful discrimination. Corporations are a necessary feature of modern business activity, and their aggregated capital has become the source of nearly all great enterprises.79

The Court uses the aggregate view of the corporation as a means to justify its application of the personal due process protection.80 Using the aggregate theory as a scapegoat would look more like: We wish we could hold that a corporation is not entitled to immunity against unreasonable searches and seizures, as humans are, but we simply cannot because the corporation is an association of individuals in all contexts. Legislatures’ and courts’ hiding behind an ontological theory to explain corporate rights, then, is most likely a figment of the public’s imagination.

C. The Stance of the Courts

A common belief among corporate-law theorists is that courts, before upholding a corporation’s constitutional rights, examine the values and policies underlying those rights with respect to individuals and determine whether those justifications apply to corporations.81 Courts have performed this task in two ways. The first reflects the circularity trap, as described above.82

80. See id. (indicating that the corporation derives personal rights from its being an association of individuals).
81. See, e.g., Krannich, supra note 36, at 64, 104–08 (positing that the U.S. Supreme Court should analyze constitutional corporate rights “in light of the values and policies that are thought to underlie it.” (citation omitted)).
82. See supra Part II.B (explaining the circularity of presuming inherent corporate characteristics).
This occurs when courts use language that presumes the existence of a truth-value for corporate characteristics, as if corporations have inherent rights to be protected and we only need to discover what they are. For example: Is it true that a corporation has the capacity for speech? If so, perhaps corporate speech merits constitutional protection. At times the U.S. Supreme Court has adopted this view, most famously in *Trustees of Dartmouth College v. Woodward*: “A corporation is an artificial being, invisible, intangible, and existing only in contemplation of law. Being the mere creature of law, it possesses only those properties which the charter of its creation confers upon it, either expressly, or as incidental to its very existence.”

At other times, courts have not deduced corporate rights from the essence of a corporation. In these cases, traits or rights often attach to corporations simply based on courts’ interpretations of statutory law. If not directly citing to statute, courts have justified the attachment of a trait based on precedent or have simply refused to elaborate on how or when a corporate trait attaches. In *Citizens United*, for example, the Court supported the premise that “First Amendment protection extends to corporations” with a paragraph solely citing to case law, but with no reference to when or how the right attached. The right most

83. See Hart, supra note 60, at 45 (arguing that statements such as “[a] company has no mind and therefore cannot intend” . . . confuse the issue because they look like eternal truths about the nature of corporations”).

84. The opposing view would be that the claim has no truth-value: one can neither say that a corporation has the inherent ability to speak or that it does not. This might be attributable to the fact that corporations can speak in some contexts but not others. See supra Part II.B (discussing inconsistencies in deducing inherent corporate characteristics).


88. See, e.g., Horwitz, supra note 33, at 66–67 (discussing how *Santa Clara* was the first U.S. Supreme Court assertion that “a corporation was a person under the Fourteenth Amendment” and how the Court failed to elaborate or provide a rationale).

89. Citizens United v. FEC, 130 S. Ct. 876, 899–900 (2010). The Court cited twenty-one cases to support this contentious point. Id.
likely attached in the predecessor cases that the Court relies upon now for support.\textsuperscript{90} When courts do not cite to statute or precedent, one would hope that the courts' language and reasoning indicate the source of the corporate right; a sheer lack of elaboration can be “puzzling and controversial.”\textsuperscript{91} The corporate right to constitutionally protected speech, traced back through precedent, ultimately rests on the premise that corporations are persons within the meaning of the Fourteenth Amendment\textsuperscript{92}—a premise that is, itself, unexplained.\textsuperscript{93}

Part of the reason for the U.S. Supreme Court's silence might be an uncertain balance between state and federal powers. Generally, for matters dealing with the internal affairs of corporations, the Court looks to the state of incorporation to interpret the law.\textsuperscript{94} Ultimately, however, Congress retains power over the entire corporate function via the Commerce Clause.\textsuperscript{95} This leaves the U.S. Supreme Court and state courts in a strange

\begin{itemize}
\item \textsuperscript{90} See, e.g., FEC v. Mass. Citizens for Life, Inc., 479 U.S. 238, 252, 263 (1986) (stating that a restriction on direct expenditure of corporate funds for electioneering violated the corporation's First Amendment rights); HORWITZ, supra note 33, at 66–67 (discussing the impact of an 1886 U.S. Supreme Court decision finding—without explanation—that "a corporation was a person under the Fourteenth Amendment") (citing Cnty. of Santa Clara v. S. Pac. R.R. Co., 118 U.S. 394 (1886)); Krannich, supra note 36, at 62 ("[T]he Court has granted corporations constitutional rights without engaging in the preliminary inquiry of whether a corporation is entitled to them under the Constitution.").
\item \textsuperscript{91} HORWITZ, supra note 33, at 66.
\item \textsuperscript{92} See First Nat'l Bank of Bos. v. Bellotti, 435 U.S. 765, 780 & n.15 (1978) ("Freedom of speech . . . always ha[s] been viewed as . . . safeguarded by the Due Process Clause, and the Court has not identified a separate source for the right when it has been asserted by corporations." (citations omitted)).
\item \textsuperscript{93} See HORWITZ, supra note 33, at 66–67 (discussing how the Court simply stated, without explanation, that "a corporation was a person under the Fourteenth Amendment"); see also HORWITZ, supra note 88 (providing the case that asserts the corporate right without explanation).
\item \textsuperscript{94} See, e.g., Rogers v. Guar. Trust Co. of N.Y., 288 U.S. 123, 130 (1933) ("[A] court—state or federal—sitting in one State will . . . decline to interfere with . . . the management of the internal affairs of a corporation organized under the laws of another state but will leave controversies as to such matters to the courts of the state of the domicile.").
\item \textsuperscript{95} See U.S. CONST. art. I, § 8, cl. 3 (granting Congress the power to regulate commerce); Norman R. Williams, Why Congress Cannot “Overrule” the Dormant Commerce Clause, 53 UCLA L. Rev. 153, 216 (2005) (discussing Laurence Tribe's position that the Commerce Clause empowers Congress to set rules for corporations).
\end{itemize}
conundrum: because state courts do not answer questions of federal constitutional law and the Supreme Court avoids questions of corporate law, judicial interpretation of the corporate entity remains a mystery.\textsuperscript{96} When the Court then issues decisions asserting corporations’ personal rights without explanation, the public, not surprisingly, views the decision with a degree of cynicism, or at least skepticism.\textsuperscript{97}

In state courts, of course, the Fourteenth Amendment is not the source, but rather the vehicle, of corporations’ personal rights.\textsuperscript{98} In fact, the amendment exacerbates the confusion by virtue of its own far-reaching applications. Although corporations are persons under the Fourteenth Amendment,\textsuperscript{99} the Court cannot mean that they share the exact same set of Fourteenth Amendment rights as individual persons. Even if corporations’ rights were a subset of persons’ rights, corporations could not “vote on equal terms with natural persons,”\textsuperscript{100} much less attend the same elementary school.\textsuperscript{101} Thus the question remains: What subset of person-type rights do corporations possess?

If legislatures refuse to say that corporations are a subset of persons or share the same set of rights as its members, and likewise do not state that corporations’ only rights are those

\textsuperscript{96} This Note credits Professor Christopher Bruner with pointing out the federalism problem here. Other authors have discussed, for example, “whether the internal affairs doctrine is only a choice-of-law rule or whether it is also a rule of constitutional law.” Jack B. Jacobs, \textit{The Reach of State Corporate Law Beyond State Borders: Reflections Upon Federalism}, 84 N.Y.U. L. REV. 1149, 1164–67 (2009).


\textsuperscript{98} See, e.g., Gitlow v. New York, 268 U.S. 652, 664, 666 (1925) (incorporating the First Amendment to the states via the Fourteenth Amendment).


explicitly provided by charter, then the courts cannot wholly
define a corporation by ascribing some of its rights in a few
different contexts. Finding a holistic framework for
determining, in theory, whether and when particular personal
rights ought to apply to corporations thus becomes an arduous
task—one the U.S. Supreme Court has not fully or consistently
undertaken. \[103\] \textit{Citizens United} embodies this struggle with
respect to corporate political speech.\[104\]

\section*{III. Citizens United}

\subsection*{A. Background}

Aside from legal limitations, corporations have the financial
ability to make large donations, in the form of direct
“contributions” to candidates or “independent expenditures”\[105\]
toward electioneering, using funds amassed from the efforts of
individual persons. Contributions are direct donations to
candidates and have historically been prohibited as a means for
corporations to spend funds on federal candidates.\[106\] Independent
expenditures—the type of spending at issue in \textit{Citizens United}—
are “money[s] spent for a communication that expressly
advocates the election or defeat of a clearly identified Federal
candidate.”\[107\] The first congressional prohibitions on corporate

\begin{footnotesize}
\begin{enumerate}
\item[102.] See Hart, supra note 60, at 41 (“It is easy to see that a statement about
the rights of a limited company is not equivalent to the statement that its
members have those same rights.”).
\item[103.] See Krannich, supra note 36, at 62 (observing that the Court has
frequently used the entity or aggregate theories as metaphors to interpret the
Constitution in the corporate context, but that the use has been ad hoc and
without regard to any simultaneous mutual exclusiveness among the theories).
\item[104.] Compare Citizens United v. FEC, 130 S. Ct. 876, 900 (2010) (stating
without explanation that the U.S. Supreme Court has treated corporations as
people under the First Amendment), \textit{with id.} at 928 (Scalia, J., concurring)
(“[T]he individual person's right to speak includes the right to speak in
association with other individual persons.” (emphasis removed)).
\item[105.] 11 C.F.R. § 100.16(a) (2000) (defining \textit{independent expenditure} as “an
expenditure by a person for a communication expressly advocating the election
or defeat of a clearly identified candidate . . . .”).
\item[106.] See \textit{id.} at 887 (stating that the corporate contribution ban persists).
\item[107.] Federal Election Commission, \textit{Citizens’ Guide Brochure} (Feb. 2004),
\end{enumerate}
\end{footnotesize}
contributions and independent expenditures were the Tillman Act of 1907\textsuperscript{108} and the Labor Management Relations Act of 1947 (the Taft–Hartley Act).\textsuperscript{109} The complete ban on corporate general fund political spending endured until the Federal Election Campaign Act of 1971 (FECA)\textsuperscript{110} provided an exception.\textsuperscript{111} FECA allowed corporations to establish separate funds through so-called Political Action Committees (PACs) to solicit voluntary donations that could be contributed to federal campaigns.\textsuperscript{112}

In 1976, the Court in \textit{Buckley v. Valeo}\textsuperscript{113} upheld limits on individual direct contributions to candidates, even though they suppressed political speech, because the government had a prevailing interest in preventing quid pro quo corruption (or the appearance thereof) by candidates accepting large contributions.\textsuperscript{114} The U.S. Supreme Court overturned individual
independent expenditure limits, however, because they were a severe restriction on political-speech rights and did not “serve any substantial governmental interest in stemming the reality or appearance of corruption in the electoral process.” (This rule—that the government may restrict individual contributions but not individual independent expenditures—is essentially what Citizens United later applied to corporations.) Two years after Buckley, in First National Bank of Boston v. Bellotti, the Court held that corporations may spend unlimited funds on issues and initiative campaigns so that the public may hear the corporate perspective.

In 1990, Austin v. Michigan Chamber of Commerce stated that a prohibition on a corporation’s use of treasury funds for direct independent campaign expenditures amounted to suppression of political speech. Yet the Court upheld a

---

115. See id. at 25 (“The expenditure limitations contained in the [Federal Election Campaign Act of 1971] represent substantial rather than merely theoretical restraints on the quantity and diversity of political speech.”).

116. Id. at 47–48.


118. First Nat’l Bank of Bos. v. Bellotti, 435 U.S. 765 (1978). In Bellotti, the Court held that a Massachusetts statute prohibiting certain corporate political expenditures violates the First Amendment. Id. at 776.

119. See id. at 784 (“[I]t amounts to an impermissible legislative prohibition of speech based on the identity of the interests that spokesmen may represent in public debate over controversial issues and a requirement that the speaker have a sufficiently great interest in the subject to justify communication.”); see also Nathaniel Persily, Contested Concepts in Campaign Finance, 6 U. PA. J. CONST. L. 118, 121 (2003) (discussing Bellotti’s holding).

120. See Austin v. Mich. Chamber of Commerce, 494 U.S. 652, 654–55 (1990) (holding that a statute prohibiting corporations from using corporate treasury funds for independent expenditures was constitutional because the provision was narrowly tailored to a compelling governmental purpose). In Austin, the Court found that preventing corruption (or the appearance of corruption) in the political arena to be a compelling governmental interest because mass corporate treasuries could unfairly influence election outcomes. See id. at 660 (“Corporate wealth can unfairly influence elections when it is deployed in the form of independent expenditures . . . .

121. See id. at 657 (stating that requiring corporations to make political contributions only through PACs “burdens corporate freedom of expression”
THE REAL ERROR IN CITIZENS UNITED

restriction on corporate expenditures because it found a compelling governmental interest in preventing the “distorting effects of immense aggregations of wealth that are accumulated with the help of the corporate form.”122 The Bipartisan Campaign Reform Act of 2002 (McCain–Feingold Act or BCRA)123 and McConnell v. FEC124 reemphasized that PACs provide a sufficient outlet for corporate speech and asserted the constitutionality of prohibiting corporate independent expenditures not from these designated funds.125 Specifically, BCRA amended 2 U.S.C. § 441b to prohibit corporations from using general treasury funds to expressly advocate the election or defeat of a political

---

122. Austin, 494 U.S. at 660.


125. See McConnell, 540 U.S. at 104–05 (“Because those entities may still organize and administer segregated funds, or PACs, for such communications, the provision is a regulation of, not a ban on, expression.”); see also Case Comment, supra note 121, at 76 n.11 (pointing out that the four Justices concurred with the portion of the opinion regarding corporate expenditures).
candidate. In 2010, the U.S. Supreme Court considered the constitutionality of this provision in *Citizens United*.127

**B. The Opinion**

*Citizens United*, a nonprofit corporation with funds primarily donated by individuals, wanted to air a film that clearly advocated the defeat of Senator Hillary Clinton in the Democratic Party’s 2008 Presidential primary elections.128 This violated the terms of the § 441b prohibition on corporate independent expenditures.129

Justice Kennedy, writing for the five-Justice majority, placed the weight of the decision on whether corporations have the same rights as individuals with respect to political speech.130 Two predecessor cases, taken together, suggested they do: *Bellotti* upheld First Amendment protection of corporate speech rights131 and *Buckley* maintained that contribution limits curb political speech.132 Members of the public have sometimes generalized this as: corporations are people, and money is speech, so corporate expenditures cannot be restricted.133


128. See id. at 887 (stating the facts).

129. Id.

130. See id. at 903 (“The Court is thus confronted with conflicting lines of precedent: a pre-*Austin* line that forbids restrictions on political speech based on the speaker’s corporate identity and a post-*Austin* line that permits them.”).


Justice Kennedy overruled *Austin* as inconsistent with “the principle established in *Buckley* and *Bellotti* that the Government may not suppress political speech on the basis of the speaker’s corporate identity.” To protect this speech, the Court applied the most protective standard: “Laws that burden political speech are subject to strict scrutiny, which requires the Government to prove that the restriction furthers a compelling interest and is narrowly tailored to achieve that interest.”

A statute restricting political speech based on the antidistortion rationale in *Austin* fails to meet this standard because the government has no interest “in equalizing the relative ability of individuals and groups to influence the outcome of elections.” Rejecting the antidistortion rationale in *Austin*, the Court stated that “[n]o sufficient governmental interest justifies limits on the political speech of nonprofit or for-profit corporations.” By finding the conflicting precedent in *Austin* to be a speech restriction based on the speaker’s corporate identity, contrary to the First Amendment, Justice Kennedy overruled the case as poorly reasoned. The Court also overruled *McConnell* to the extent that it upheld the statute restricting corporate independent expenditures, finding the antidistortion interest “unconvincing and insufficient.”

After dismissing the antidistortion rationale as being inconsistent with precedent and with the Constitution, the Court rejected the Government’s other main justifications for the § 441b restrictions, that limits are necessary to prevent corruption of political officials and that shareholders would be compelled to fund corporate speech. The Court followed *Buckley*, concluding that “[t]he anticorruption interest is not

---

136. *Id.* at 904 (quoting *Buckley*, 424 U.S. at 48) (citation omitted).
137. *Id.* at 913.
138. *Id.*
139. See *id.* at 904–08 (providing reasoning to overrule *Austin*).
140. See *id.* at 908–11 (rendering the anticorruption and shareholder-interest rationales invalid).
141. See *id.* at 908 (stating that “[t]he governmental interest in preventing corruption and the appearance of corruption [is] inadequate to justify [the ban]
sufficient to displace the speech here in question” and that “independent expenditures, including those made by corporations, do not give rise to corruption or the appearance of corruption.”142 The shareholder-interest rationale, the Court stated, “would allow the Government to ban the political speech even of media corporations” when the abuse could instead be “corrected by shareholders through the procedures of corporate democracy.”143

In terms of novelty, therefore, Citizens United’s real contribution was its rejection of the government’s policy rationales for permitting a ban on corporate political speech, rather than the already-established proposition that corporate speech is protected to the same extent as individual speech.144 Regardless of the legal reality, however, the media and the public have sometimes interpreted the case differently.

C. Public Impact

Citizens United has prompted a wide range of discussion among the general public, primarily criticism.145 But the fear that large public corporations would use amassed funds to make sweeping expenditures has greatly subsided as the public has realized how relatively little corporations have spent on campaigns.146 Corporations’ hesitancy is most likely because

142. Id. at 908–09.
143. Id. at 911 (quoting First Nat’l Bank of Bos. v. Bellotti, 435 U.S. 765, 794 (1978)).
144. See Allison R. Hayward, Citizens United: Correct, Modest, and Overdue, 2 AKRON J. CONST. L. & POL’Y 1, 2–3 (2010) (discussing that Citizens United does not overturn long-standing precedent).
publicly traded companies have to answer to shareholders and customers about any questionable use of funds. The election statistics from the first presidential general election since Citizens United support this result. For example, companies “contributed roughly $75 million to super PACs in the 2012 election cycle,” compared with the approximately $661 million super PACs raised overall in that cycle.

Some fears about the influence of super PACs and Citizens United, however, have materialized. For example, although super PACs may only advocate for—not directly contribute to—political candidates, those supporting presidential candidates in the 2012 general election collected and spent more funds than the candidates themselves: Priorities USA Action spent $67.5 million supporting President Barack Obama and liberal agendas,

147. See id. ("[Public companies] know those contributions might become public at some point, and no company that sells a product wants to risk [a substantial negative consumer reaction]."); Ctr. for Responsive Politics, Mystery Firm is Election’s Top Corporate Donor at $5.3 Million, OPENSECRETS.ORG (Nov. 5, 2012), http://www.opensecrets.org/news/2012/11/mystery-firm-is-elections-top-corpo.html (last visited Nov. 7, 2012) ("Fortune 500 companies are the least likely to be the ones who will be out in front giving publicly,” said Rick Hasen, a law professor at the University of California-Irvine. “They want to have influence over elections and elected officials, but they don’t want to alienate customers.”) (on file with the Washington and Lee Law Review).


150. See Peter Overby, As ‘Citizens United’ Turns 2, SuperPACs Draw Protests, NAT’L PUB. RADIO (Jan. 20, 2012), http://www.npr.org/2012/01/20/145500168/superpacs-celebrate-anniversary-of-citizens-united-case (last visited Nov. 7, 2012) (quoting lawyer Ken Gross as stating that super PACs “are metastasizing” and are “almost bigger than the party committees”) (on file with the Washington and Lee Law Review); Alan Greenblatt, Big Money: Stuffing the Ballot Box?, NAT’L PUB. RADIO (June 4, 2012), http://www.npr.org/2012/05/29/153914560/big-money-and-the-ballot-box (last visited Nov. 7, 2012) (quoting Republican consultant Ed Goeas as stating, “Money is now at the end that's furthest away from the candidates and furthest away from the parties... The money is with these other groups that are having more impact on the campaign than the campaign itself.”) (on file with the Washington and Lee Law Review).
and Restore Our Future spent $142.7 million supporting Republican nominee Mitt Romney and conservative agendas.\textsuperscript{151} Possibly even more influential are the 501(c)(4) organizations, such as Crossroads GPS, co-founded by Karl Rove,\textsuperscript{152} and Americans for Prosperity, affiliated with the Koch brothers; these organizations can fund advertisements without disclosing how much they have spent until after the election and without ever disclosing their donors.\textsuperscript{153}

Although the laws regulating these groups’ expenditures and disclosures are not the direct result of \textit{Citizens United},\textsuperscript{154} some people view the case as converting the political race to a money standoff and “unleash[ing] a torrent of poorly disclosed, if disclosed at all, spending by the superwealthy”\textsuperscript{155} and “of money from businesses and the multimillionaires who run them, and as a result we are now seeing the corporate takeover of American politics”\textsuperscript{156}—however inaccurate that perspective may be.\textsuperscript{157} Still,

\begin{flushright}

\textsuperscript{152} Crossroads GPS also owns a super PAC, American Crossroads, which spent $104.7 million for conservative agendas during the 2012 election cycle. \textit{Id.}


\textsuperscript{154} \textit{Id.} (stating that “[social welfare groups’] secret spending . . . is not due to the 2010 \textit{Citizens United} decision”).


\textsuperscript{156} Bai, \textit{supra} note 146.

the 2012 election cycle saw more outside\textsuperscript{158} spending than every other election cycle since 1990—combined;\textsuperscript{159} the $1,032,901,165-total is almost 3.5 times the 2008 election cycle’s spending.\textsuperscript{160} When adding in official spending, the people of the United States spent a total of approximately $6 billion on federal election campaigns for 2012.\textsuperscript{161}

Regardless of the actual factors driving the substantial increase in expenditures for this presidential election, many Americans are legitimately concerned that \textit{Citizens United} overturned precedent that restricted corporate political speech rights, stating that such corporate speech is protected by the U.S. Constitution.\textsuperscript{162} One extreme generalization takes the form: “the Supreme Court held that corporations are ‘people,’” but they do not deserve the same rights as human people.\textsuperscript{163} This argument indicates a belief that the Court based its decision on an entity theory.\textsuperscript{164} Some commentators, and perhaps even Justice Stevens,\textsuperscript{165} might agree that the Court assumes corporations are

\begin{flushleft}
\textsuperscript{159} Id.
\textsuperscript{160} Id.
\textsuperscript{162} Citizens United v. FEC, 130 S. Ct. 876, 912–13 (2010) (overruling \textit{Austin} because it “abandoned First Amendment principles” by restricting corporate political speech).
\textsuperscript{164} See supra Part II.A (discussing entity theories).
\textsuperscript{165} See \textit{Citizens United}, 130 S. Ct. at 972 (Stevens, J., dissenting) (‘[Corporations’] ‘personhood’ often serves as a useful legal fiction. But they are not themselves members of ‘We the People’ by whom and for whom our Constitution was established.’).
\end{flushleft}
entities—either real or artificial. In general, those viewing the Court as adopting an entity theory are suspicious of this theory (as they see the Court taking it); the critics see the Court’s conclusion in *Citizens United* as having adopting an entity theory without any reason for doing so. Some theorists have argued that the opinion supports either an entity or an aggregate theory, and they criticize the normative conclusions that they draw from the theory’s application in *Citizens United*. Professor David Millon, however, recognized the social cost of this seemingly endless debate:

> [E]fforts to derive ‘ought’ from ‘is’ have not succeeded. Indeed, such intellectual exercises may have stood in the way of careful examination of the truly urgent questions raised by corporate activity. Analysis of difficult questions of social policy have probably been hindered by assumptions about the distinctiveness of activity in the corporate form, whether the corporation is thought to be an entity or instead is an aggregation of people distinct from the rest of society.

*Citizens United* supports Professor Millon’s point because of the significant (and negative) public reaction. Perhaps the policy questions should be at the forefront of individuals’ minds because those are the real terms that affect corporate rights; even if the

---

166. See, e.g., Avi-Yonah, supra note 8, at 41 (“The entire *Citizens United* opinion, both the majority and the dissent, are thus reflective of the real entity view. Corporations stand on their own, independent of both the state that created them and the shareholders that own them.”).


168. See id. (arguing that the Court had no reason to rule broadly, had mistaken assumptions, and wrongly dismissed the likelihood of corruption).

169. See, e.g., Avi-Yonah, supra note 8, at 41 (arguing that the Court used a real-entity theory); Padfield, supra note 8, at 225 (arguing that the Court used an aggregate theory); Kerr, supra note 145, at 314 n.28 (asserting the artificial-entity theory in a *Citizens United* analysis).

170. See Avi-Yonah, supra note 8, at 3 (“[T]hroughout all of [the changes in the legal conception of the corporation], spanning two millennia, the same three theories of the corporation can be discerned.”).


172. See supra Part III.C (discussing *Citizens United*’s public impact).
Supreme Court adopted an entity theory, no one ought to care about the metaphor when policy implications are in dispute. In fact, *Citizens United* may implicate some new policy considerations, particularly in the campaign finance context.

**D. Legal Impact**

*Citizens United* might constrain courts or affect the public in unanticipated ways, which could bear on corporate rights. First, the Court upheld constitutional protection of corporate speech to the same extent that individual speech is protected under the First Amendment. By proclaiming this right to be facially constitutional instead of addressing the issue as-applied, the Court has limited its discretion. For example, corporations currently cannot contribute directly to candidates, but individuals may currently contribute up to $2,500 per federal candidate, per election cycle. Based on the reasoning in

---

173. See infra Part IV.C (discussing the difference between making metaphysical versus function claims).

174. See infra Part V (discussing policy arguments).


176. See id. at 894 (“[It] is necessary . . . to consider the [statute's] facial validity.”).

177. See Tillman Act of 1907, Pub. L. No. 59-36, 34 Stat. 864 (codified as amended at 2 U.S.C. § 441b(a) (2006)) (prohibiting corporate monetary contributions to political campaigns); see also *Buckley v. Valeo*, 424 U.S. 1, 58 (1976) (stating that individual contribution limits are constitutional because they “serve the basic governmental interest in safeguarding the integrity of the electoral process without directly impinging upon the rights of individual citizens and candidates to engage in political debate and discussion”).

Citizens United that monetary restrictions burdening political speech are subject to strict scrutiny, and based on the Court’s rejection of the antidistortion rationale, the U.S. Supreme Court might strike down the ban on corporations contributing directly to candidates as impermissibly restricting political speech. Though a decision extending the protection to corporate contributions might affect legal discourse about public policy, this seems to be the extent of such a ruling’s collateral impact on corporate theory because further protecting the corporate right to speech through political contributions would most likely rely on the same reasoning as the Court used to permit corporate independent expenditures. Thus, a possible future decision to permit corporate contributions might necessarily follow from Citizens United.

Second, although the media has sometimes exaggerated or misinterpreted Citizens United, some legal implications of the corporate monetary contributions to political campaigns).

179. See Citizens United, 130 S. Ct. at 898 (applying strict scrutiny to laws burdening political speech).

180. See id. at 913 (finding the antidistortion interest “unconvincing and insufficient”).


182. Id. But see, e.g., United States v. Danielczyk, 683 F.3d 611, 616 (4th Cir. 2012) (reversing a federal district court decision that applied Citizens United to ban corporate contributions); Minnesota Citizens Concerned for Life, Inc. v. Swanson, 692 F.3d 864, 880 (8th Cir. 2012) (affirming a federal district court decision upholding a state’s ban on direct corporate contributions to candidates); Eugene Volokh, Constitutional to Ban Corporate Contributions to Candidates (as Opposed to Independent Expenditures), THE VOLOKH CONSPIRACY (May 16, 2011, 12:20 PM), http://volokh.com/2011/05/16/constitutional-to-ban-corporate-contributions-to-candidates-as-opposed-to-independent-expenditures/ (last visited Nov. 7, 2012) (arguing it is constitutional to ban corporate contributions to candidates) (on file with the Washington and Lee Law Review).

opinion might indeed have far-reaching effects on the public and on campaign finance reform. Groups forming under § 527 of the Internal Revenue Code, including PACs, may gather uncapped corporate and individual donations and use the funds for independent campaign expenditures. Since Citizens United, courts have had no discretion to limit these groups. Citizens United also laid the groundwork for a federal circuit court decision two months later, SpeechNow.org v. FEC, which led to the clear ability for individuals to form “super PACs” to make only independent expenditures. Super PACs have proliferated as a means to “raise unlimited sums of money from corporations, unions, associations and individuals, then spend unlimited sums to overtly advocate for or against political candidates.” Rather than focusing on the potential for public corporations’ overwhelming influence in the political sphere, perhaps the public ought to be more concerned with individual associations, unions, closely held companies, and other coordinating campaign spenders that can raise and spend unlimited funds.


185. Id. at 3–5 (discussing the effect of Citizens United).

186. See id. at 4 (stating that the Court enables 527 organizations to solicit unlimited donations).

187. See SpeechNow.org v. FEC, 599 F.3d 686, 693 (D.C. Cir. 2010) (“The independence of independent expenditures was a central consideration in [Citizens United]. By definition, independent expenditures are not made in concert or cooperation with or at the request or suggestion of such candidate [or his committees or agents]” (internal quotation marks omitted)).

188. See id. at 698 (concluding that contribution limits cannot constitutionally be applied to individuals pooling money only to make unlimited independent expenditures).


190. Id.

191. See Interview with Jason Torchinsky, Partner, HoltzmanVogelJosefiak PLLC, in Lexington, Va. (Feb. 29, 2012) (suggesting that publicly held corporations are not the primary concern after Citizens United and SpeechNow.org). These groups “cannot coordinate with campaigns” and must disclose all expenditures frequently, allowing competing groups to track their
Third, *Citizens United* may have committed courts and legislatures to a broader conception of constitutionally protected speech than anticipated. For example, although the decision upheld disclosure requirements on independent expenditures,\(^{192}\) social welfare organizations formed under § 501(c)(4) of the Internal Revenue Code can still accept unlimited anonymous donations and use the funds for electioneering expenditures,\(^{193}\) and since 2010, corporations have been able to donate their funds to those organizations.\(^{194}\) By allowing unlimited independent expenditures and not protecting against loopholes such as the use of 501(c)(4) organizations, *Citizens United* opened the door for unlimited undisclosed corporate expenditures,\(^{195}\) and the courts will have no discretion within the Constitution to ban these expenditures if they become substantial.\(^{196}\) In addition, *Citizens United* seems to have little or no effect on restrictions on charity lobbying efforts.\(^{197}\)

*Citizens United*’s legal impact is thus significant in the context of electioneering, and the developments in this area of spending. Id.


193. 501(c)(4)s may remain tax-exempt if the communications are “issue-based,” but advertisements can still reflect particular candidates in a bad light. See Peter Overby, *A Fine Line: Distinguishing Issue Ads From Advocacy*, NAT’L PUB. RADIO (June 19, 2012), http://www.npr.org/2012/06/19/155325685/a-fine-line-distinguishing-issue-ads-from-advocacy (last visited Aug. 23, 2012) (showing how an issue advertisement can walk the line of advocating the defeat of a candidate) (on file with the Washington and Lee Law Review).

194. See *Blasio*, supra note 184, at 4 (explaining 501(c)(4)s after *Citizens United*).

195. See id. at 3 (“While the Court acknowledged the permissibility of disclosure requirements for political spending, the decision enabled many corporations to spend money on independent political broadcasts without disclosing the donors that fund their activities.”).

196. See *Citizens United*, 130 S. Ct. at 892 (“[T]he Court cannot resolve this case on a narrower ground without chilling political speech, speech that is central to the meaning and purpose of the First Amendment.”).

law are deeply interrelated with the policy arguments regarding corporate rights. 198

IV. Corporate Theory After Citizens United

As some legal scholars were ready to move on from the question of corporate personhood, 199 the question became vastly more interesting to the public. 200 Commentators have interpreted Citizens United as adopting the aggregate theory, the real-entity theory, the artificial-entity theory, or no theory of corporate rights. 201 Courts do not consistently adopt one of these views over the others. 202 Even various U.S. Supreme Court Justices have adopted, explicitly or implicitly, each of these three theories at some point—and in some cases, more than one theory in the same opinion. 203 Given the inconsistent and seemingly incompatible application of theories propounded by legal scholars and courts alike, the public’s misunderstanding of Citizens

198. See infra Part IV.D (concluding that the real source of the corporate right in Citizens United is policy).

199. See, e.g., Millon, supra note 171, at 58 (arguing that we may be better off with a more appropriately focused debate). But see, e.g., Padfield, supra note 8, at 225 & n.74 (noting Millon’s and Dewey’s respective arguments against the corporate theory debate and concluding that the debate is relevant to the question of the corporate role in society).


201. See, e.g., Avi-Yonah, supra note 8, at 41 (arguing that the Court used a real-entity theory); Kerr, supra note 145, at 314 n.28 (asserting the artificial-entity theory in a Citizens United analysis); Millon, supra note 171, at 56–57 (discussing the modern relevance of entity or aggregate theories of corporate personhood); Padfield, supra note 8, at 225 (arguing that the Court used an aggregate theory). This Note does not discuss the many other theories that answer questions other than those concerning the nature or existence of the corporate form. See, e.g., Padfield, supra note 8, at 215 (discussing other prominent corporate theories).

202. See supra Part II.C (discussing the various positions courts have taken on the corporate form and corporate personhood).

203. See Krannich, supra note 36, at 62 (observing that the Court has frequently used the entity or aggregate theories as metaphors to interpret the Constitution in the corporate context, but that the use has been ad hoc and without regard to any simultaneous mutual exclusiveness among the theories).
United is not surprising. Moreover, the public is misguided about the source of the corporate speech right because the public relies on the media’s rendition of an already inadequate opinion.204

A. The Public’s Misuse of “Person”

Public attention to corporate rights, after Citizens United in particular, amplifies a main problem with theories of corporate personhood: the fallacy that metaphysical discussions about corporations can imply functional, legal conclusions.205

The use of the phrase “corporations are people” to deduce personal corporate rights is a misclassification.206 Consider dividing descriptive language about corporations, taken in context, into three distinct sets: (1) metaphysical language, such as “corporations are entities” or “corporations have rights derived from their constituents,” which describes an ontological conception of what a corporation is;207 (2) functional language, such as “corporations provide limited shareholder liability” or “corporations are established by charter,” which describes what corporations are in real, legal, or practical terms;208 and (3) language that is neither metaphysical nor functional, such as normative or qualitative language. Focusing on the first two of these three sets, the phrase “corporate person” can have two different meanings. Metaphysically, it refers to a corporate ontology in the same way that corporate-essence theories attempt to represent corporations by using one of two metaphors—either as entities separate from their shareholders or as an aggregate of

204. See, e.g., PBS, supra note 183 (headlining an error that corporate contributions are constitutional, as opposed to independent expenditures—to which the anchors correctly refer).

205. See supra Part II.B (discussing the circularity problem with presuming an inherent truth about some corporate essence and then defining the essence by deriving traits from that essence).

206. See John Dewey, The Historic Background of Corporate Legal Personality, 35 Yale L.J. 655, 660 (1926) (distinguishing the metaphysical from the pragmatic use of “person”).

207. See id. (providing “two radically different types of definitions,” one being “a definite metaphysical conception regarding the nature of things”).

208. See id. at 660–61 (providing a second type of definition “which proceeds in terms of consequences, . . . stated in terms of specific effects extrinsically wrought in other things”).
participants. Importantly, this language does not ascribe real traits to corporations, but rather allows us to speak about what a corporation should be, based on its ontology or essence. Functionally, “corporate person” defines a set of human rights that corporations are permitted to adopt if the issue arises, as if a court were saying, “Treat the corporation like a human in this context.” This language is used to establish that corporations actually have explicit, real rights when legislatures, courts, or charters assign such rights.

Sometimes people confuse these two types of language in order to answer questions about what a corporation or corporate person is. For example, consider the two main branches of the corporate-rights dichotomy: artificial and natural. Both attempt to define what rights a corporation has to function within the law. But when the law does not state whether a corporation has a particular right, it may be tempting to ask the question, “What sort of thing is a corporation—one that ought to have this right, or one that ought not?” This crosses over into the metaphysical category because we are speaking about the essence of a corporation as a metaphor for something we can otherwise only explain to the extent the law provides details.

Using language from one category in an attempt to draw conclusions in another would be a “category error,” as Gilbert Ryle defined the term in his philosophy of the human mind. In Ryle’s classic example of a category error, a foreigner visits a university, tours all the buildings and fields, sees the students and faculty, and then asks: “But where is the University?” Ryle explains that the foreigner asking the question commits a category error by inferring that the university exists as another “member of the class of which [the colleges, libraries, playing fields, etc.] are members.” Classifying the language in Ryle’s example into the three categories listed above makes the precise flaw obvious. The term “where” indicates that the speaker wishes to know a physical (functional) location. If he instead asked the

---

209. See id. at 661–62 (stating that calling corporations “persons” is a verbal matter and does not impute new behaviors).
211. Id.
212. Id.
implicit question ("What is a university?"), the expected response would remain in the metaphysical category and commit no error.213

Before applying the category-error notion to the corporate-theory context, it might help to illustrate why committing a category error is problematic. Consider a mythical figure such as the stork that delivers babies. People construct this idea for when children ask the inevitable conception question and for the purpose of answering without being accountable for the implications of the true answer, as a child would perceive it. Parents can ascribe fictional traits to the stork: it can fly; it carries the baby in swaddling clothes to the proper doorstep; maybe even, it likes pickles. The stork continues to serve its real purpose, and parents are satisfied.

But the stork is a fabrication by humans for a purpose. Realities persist: just as limited shareholder liability does not mean nobody pays a loss, babies are actually being conceived regardless of children’s naivety. Talking about the stork in the same construct as real beings or objects produces inconsistencies. A husband who tells his wife, "Honey, ask the stork if he wants more pickles," will leave the wife bewildered; it puts the stork in the "existing" category when it should be in the "non-existing" category. Likewise, discourse about corporations on a metaphysical level using functional language (where the legal world seems to operate) commits a category error.214 The only way we can consistently speak about what corporations ought to be is on the metaphysical level, but we can speak about what corporations can actually, legally do on the functional level.215

Once one spouse commits the error by talking about the stork in functional and metaphysical terms simultaneously, the other spouse rightly questions whether the stork is overstepping its purpose; no partner wants to be replaced by a mythical stork. Returning to the logical flaw in reasoning, the phrase “ask the stork if he wants more pickles” presumes there exists a truth-

213. See id. (dissecting the university example).
214. See id. at 16 (explaining category mistakes).
215. See Dewey, supra note 206, at 660 (describing two different types of definitions: metaphysical regarding the nature of things and pragmatic regarding consequences).
value about whether he does. But the answer to the question is derived from the parents, not the stork; there is no answer until the parents make it up. Likewise, after Citizens United, the public's fear that corporations are people in the metaphysical sense, deriving rights from the person mold it attempts to fill, illuminates the real problem at issue: the corporate right to free speech via independent expenditures does not derive from corporate personhood.

When criticizing Citizens United, the public interprets the Court as speaking about the essence of a corporation, which belongs in the metaphysical category. In fact, Justice Kennedy uses terms that describe corporate constituents or qualities in themselves, which belong in the functional category. For a simple example, “Citizens United wanted to make [the film] available . . . . It feared, however, that both the film and the ads would be [banned].” Most people would recognize that these personal verbs are only being used in the functional sense; it is hard to imagine this language as an indication that corporations literally and naturally are capable of having desires or fears.

The problem with using theories to derive corporate traits arises with the idea that corporations have an essence from which we can deduce corporate rights. In 1926, John Dewey distinguished the corporate body as a mere right-and-duty-bearing unit from the corporate body as having a nature such that rights can be ascribed to it. He rejects the notion that even

216. See supra Part II.B (discussing the circularity of deducing characteristics that only exist if fabricated).

217. See supra Part II.B (concluding that deducing rights from corporate personhood begs the question).


220. Id. (emphasis added).

221. See Dewey, supra note 206, at 660 (arguing that ordinary connotations of “person” do not apply in the sense of the corporate person as a unit bearing rights and duties).

222. See id. at 658–59 (2010) (discussing whether we assume corporations have a nature).
“right-and-duty-bearing unit[s] should have a character of [their] own”223 by asserting that implicit theories in the minds of jurists, such as the aggregate and entity theories, do not vest in jurisprudence without “confusion, conflict and uncertainty.”224 Dewey exposes the error of packing the term “person” with extraneous uses of the word, such as psychological or philosophical uses; “person” is meant for the limited use with which courts or legislators explicitly dictate.225

Although the public might misunderstand Citizens United’s use of person, the opinion’s use is precisely to give constitutional protection to corporations to the same extent as it applies to individuals.226 Indeed, if the courts and legislatures could physically list every manifestation of personal speech rights and then map that list onto corporations, the debate would not concern corporate personhood but rather purely normative claims about whether policy supports a corporation’s having those rights. “Person” is a mere convenience to answer questions about corporate rights and duties, and deducing corporate rights from some corporate essence is a semantic error—a category error from conflating the metaphysical with the functional.227

B. Looking for Theoretical Language in Citizens United

1. Justice Kennedy’s Majority Opinion

In Citizens United, Justice Kennedy, writing for the majority, avoids committing a category error because he does not deduce the corporate right from the essence of a corporate person. Instead, he begins with general language about a speaker’s rights without establishing first whether a corporation qualifies as a speaker: “the Court cannot resolve this case on a narrower

223. Id. at 660.
224. Id.
225. See id. at 655–56 (arguing that the term “person” in the legal sense is affected by extraneous influences, such as what “person” signifies in popular speech, psychology, philosophy, or morals).
227. See Hart, supra note 60, at 42–43 (explaining the circularity in deducing corporate characteristics from a corporate form).
ground without chilling political speech, speech that is central to
the meaning and purpose of the First Amendment.” Further,
“[t]he right of citizens to inquire, to hear, to speak, and to use
information to reach consensus is a precondition to enlightened
self-government and a necessary means to protect it.” Next,
Justice Kennedy finds corporations’ right of free speech: “The
Court has recognized that First Amendment protection extends to
corporations.” He cites this sentence to twenty-one cases that
support similar claims based on a chain of precedent tracing back
to the single 1886 case, which declares the corporate right
without justification. Justice Kennedy follows suit by providing
only this one sentence to explain the extension of that right to
corporations. From that right, Justice Kennedy immediately
extends the protection to political speech, and then to the primary
issue of independent expenditures, never returning to the
application of personal rights to corporations.

In the following sections of the opinion, Justice Kennedy
addresses and rejects the arguments that the government has
sufficient interest in preventing corruption (or its appearance),
in preventing a distortion based on the speaker’s wealth, and in
protecting shareholders from being compelled to fund corporate
speech, concluding that none of those arguments adequately
protect a valid constitutional right of speech. He specifically
applies each of these conclusions to corporations, stating that
Buckley and Bellotti “could not have been clearer” that “the
Government cannot restrict political speech based on the
speaker’s corporate identity” and rejecting the “post-Austin line

228. Citizens United, 130 S. Ct. at 892.
229. Id. at 898.
230. Id. at 899.
231. See Horwitz, supra note 33, at 66–67 (discussing how Santa Clara was the first U.S. Supreme Court assertion that “a corporation was a person under the Fourteenth Amendment” and how the Court failed to elaborate or provide a rationale).
233. Id. at 900–01.
234. Id. at 908.
235. Id. at 904.
236. Id. at 911.
237. Id. at 902.
[of cases] that permits [corporate-based restrictions].” Indeed, in Bellotti, the U.S. Supreme Court recognized the need to stick to constitutional interpretation, as opposed to corporate theory: “The proper question therefore is not whether corporations ‘have’ First Amendment rights and, if so, whether they are coextensive with those of natural persons. Instead, the question must be whether [the statute] abridges expression that the First Amendment was meant to protect.”

Justice Kennedy could only be construed to make statements about the corporate essence or corporate rights in two places: First, he states that “[c]orporations, like individuals, do not have monolithic views. On certain topics corporations may possess valuable expertise . . . .” This commits no category error; while the former sentence seems to be metaphysical, it does not infer any legal conclusions based on that claim. The latter sentence does not definitively assert any right or trait. Second, Justice Kennedy states: “[W]ealthy individuals and unincorporated associations can spend unlimited amounts on independent expenditures. Yet certain disfavored associations of citizens—those that have taken on the corporate form—are penalized for engaging in the same political speech.” In context, Justice Kennedy is making the point that wealthy individuals are sometimes as influential as corporations and that corporations should be treated as other associations of citizens. This too seems purely functional because Justice Kennedy already asserted that corporations are treated as persons for speech rights, so this only asserts rights already derived from legal precedent—it does not imagine new rights that ought to attach to the corporate citizen due to its form or essence. Therefore, Justice Kennedy’s language appears clear of subscription to any theory for purposes of improperly deriving corporate rights.

Some commentators, however, interpret Justice Kennedy’s language as suggesting the aggregate theory, referring to the majority’s statements about banning “the political speech of

238. Id. at 903.
241. Id. at 908 (citation omitted).
millions of associations of citizens” and penalizing “certain disfavored associations of citizens—those that have taken on the corporate form.”

One scholar asserts, “[t]he majority viewed the corporation as fundamentally little more than an association of citizens.” Some of the Court’s claims, however, are trivially true by virtue of the corporate function in society, such as saying a corporation is an association of individuals. This does not reflect any metaphysical claim about the essence of a corporation from which we can deduce rights. To read the opinion as asserting that the corporation is an aggregate retaining the rights of its constituents is problematic: on one hand, if it inherits all of its constituents’ rights, contradictions will arise for those rights a corporation has been said not to possess, such as the right to swear an oath; on the other hand, if the aggregate only gleans some of the personhood rights, the court will still have to determine which rights it retains, which begs the question—it retains precisely those rights that the Court stipulates. The only sense in which the opinion could properly be advocating the aggregate theory, then, is in the metaphorical sense, merely to discuss the corporation, not to deduce functional rights.

Professor Avi-Yonah argues that the majority’s language can only indicate a real-entity view. He arrives at this conclusion by process of elimination among the aggregate and entity theories, and he uses Justice Kennedy’s statement that “the ban on corporate speech was not alleviated by the fact that a PAC organized and controlled by the same corporation could speak freely because ‘[a] PAC is a separate association from the

243. Id. (quoting Citizens United, 130 S. Ct. at 908).
244. Id. at 224.
245. For example, if several individuals incorporate their business and hang a sign, it would seem to be speech of a corporation—an association of individuals—under any theory.
246. See HART, supra note 60, at 42–43 (“Under what conditions do we refer to numbers and sequences of men as aggregates of individuals and under what conditions do we adopt instead unifying phrases extended by analogy from individuals?” (internal quotation marks omitted)).
247. See id. (explaining the circularity in deducing corporate characteristics from a corporate form).
First, Professor Avi-Yonah asserts that Justice Kennedy does not advocate the aggregate theory because “under the aggregate view both the corporation and the PAC are owned by the same ultimate shareholders,” so they could not both be equivalent with their parts if they are “‘separate association[s].’” Second, Professor Avi-Yonah tackles the artificial–real entity distinction (the corporate-rights dichotomy) to narrow down his interpretation that Justice Kennedy takes an entity position. He rejects that Justice Kennedy assumes an artificial-entity view, which posits that the corporation is an entity with rights to the extent the state grants them. This theory, Professor Avi-Yonah claims, is inconsistent with Justice Kennedy’s statement about a PAC being separate from the corporation because “both the PAC and the corporation are created by the same state” and controlled by the same corporation, so would presumably be the same association of people.

Professor Avi-Yonah presumes he is left with only the real-entity theory, which states that corporations are entities consisting of shareholders with separate management and corporate rights beyond what the legislature prescribes. He mentions John Dewey’s argument that the aggregate and entity theories are circular and “could be deployed to suit any purpose.” Professor Avi-Yonah then briefly considers how Dewey’s argument “dismisses as irrelevant the debate among the aggregate, artificial entity, and real entity views of the corporation.” He acknowledges the credibility of this view but then promptly moves on, stating that, “As a practical matter, . . . the real entity view predominated for

249. Id. (alteration in original) (quoting Citizens United v. FEC, 130 S. Ct. 876, 897 (2010)).
250. Id. (quoting Citizens United, 130 S. Ct. at 897).
251. See id. at 33 (defining the real- and artificial-entity theories and the aggregate theory).
252. Id. (stating that the artificial entity theory “views the corporation as a creature of the State”)
253. See id. at 39 (discussing how the majority adopts the real entity view because it implicitly rejects the aggregate view and “does not even mention the artificial entity view”).
254. See id. at 33 (defining the real- and artificial-entity theories and the aggregate theory).
255. Id. at 23 (citing Dewey, supra note 206, at 673).
large, publicly traded corporations.\textsuperscript{256} Rather than rejecting Dewey's view, therefore, he merely observes that its popularity faded.\textsuperscript{257} In sum, Professor Avi-Yonah rejects that Justice Kennedy aligns with the aggregate or artificial-entity theories but provides no direct language in support of his claim that Justice Kennedy espouses a real-entity theory.\textsuperscript{258} This lack of affirmative foundation for the real-entity theory in \textit{Citizens United} does not pair well with Professor Avi-Yonah's willingness to overlook Dewey's argument that none of the three theories are on point. To agree with Professor Avi-Yonah, either we are left to our own devices to reject Dewey's argument or we must defer to general legal scholars' silence in reaction to Dewey's theory, which seems to have left them scratching their heads until enough time passed to sweep his point under the rug as too devastating to centuries of corporate-law theories.\textsuperscript{259}

In fact, Dewey's point is not so devastating. Professor Avi-Yonah seems to recognize that his argument may have contributed to a more practical (functional) use of the theories, rather than theoretical (metaphorical).\textsuperscript{260} Still, as long as scholars attempt to use metaphor-based theories (such as the corporate-essence theories) to answer questions about corporations' functional, real rights, Dewey's point remains cogent: “'person' signifies what law makes it signify.”\textsuperscript{261} A court stating that “corporations are people” under one law only means to qualify \textit{people} to include \textit{corporations}—it does not add a trait to the corporate form labeling it as a person.

In other cases, the Court has not been so careful,\textsuperscript{262} and the members of the public frequently commit this error.\textsuperscript{263} Justice

\begin{itemize}
  \item \textsuperscript{256} \textit{Id.}
  \item \textsuperscript{257} \textit{See id.} (“Dewey was influential in that the theoretical debate on corporate personality largely disappeared until the 1970s.”).
  \item \textsuperscript{258} \textit{See id.} at 39–43 (applying \textit{Citizens United} language to align the justices with corporate-rights theories).
  \item \textsuperscript{259} \textit{See id.} at 23 (glossing over the failure of Dewey's theory to catch on).
  \item \textsuperscript{260} \textit{See id.} (“Dewey was influential in that the theoretical debate on corporate personality largely disappeared until the 1970s. As a practical matter, however, the real entity view predominated for large, publicly traded corporations.”).
  \item \textsuperscript{261} Dewey, \textit{supra} note 206, at 655.
  \item \textsuperscript{262} \textit{See, e.g.}, Hale v. Henkel, 201 U.S. 43, 76 (1906) (“A corporation is, after all, but an association of individuals under an assumed name and with a
Kennedy, it seems, either has inadvertently avoided a category error by limiting his justification of the corporate person to essentially one sentence (“The Court has recognized that First Amendment protection extends to corporations.”) or, more likely, has properly stayed within the narrow lines of federal jurisdiction by carefully choosing his language to avoid expanding the corporate essence based on a mere image of what a corporation ought to be. Despite his tiptoeing around corporate law, however, he is not off the hook for the confusion surrounding Citizens United.

2. Justice Scalia’s Concurring Opinion

Justice Scalia offers more debatable language in his concurring opinion. He finds the majority opinion consistent with the Framers’ intent to constitutionalize free speech for individual Americans:

All the provisions of the Bill of Rights set forth the rights of individual men and women—not, for example, of trees or polar bears. But the individual person’s right to speak includes the right to speak in association with other individual persons . . . . It is the speech of many individual Americans, who have associated in a common cause, giving the leadership of the [group] the right to speak on their behalf . . . . [A corporation] cannot be denied the right to speak on the simplistic ground that it is not “an individual American.”

To some this may seem like language characteristic of the aggregate view, but Professor Avi-Yonah correctly presumes distinct legal entity.”).
that Justice Scalia means “corporate management working together as an association of persons” for a common cause. In fact, Justice Scalia’s language, such as referring to the corporation as a “legal entity,” is consistent with Professor Avi-Yonah’s interpretation that Justice Scalia might adopt a real-entity theory, which would be proper so long as Justice Scalia does not transgress into metaphysical notions about corporate nature.

Most telling, however, of Justice Scalia’s use of the entity conception of a corporation is his apparent assumption that the First Amendment applies to corporations insofar as it does not exclude them outright. Because Justice Scalia essentially presumes corporations have the same First Amendment speech rights that people have, he provides no justification for this premise and we are left to wonder if he grounded it in a right derived from the corporate essence or in actual law. If the right comes from his notion of the corporate entity, this is an impermissible category error: obviously an association of people cannot have all of the same rights as the individuals that make it up, such as the right to vote in an election, but the question “Which rights does it derive from its individuals?” cannot be answered with “The same rights that its individuals are entitled to.” Conversely, if Justice Scalia implicitly justified the right using law, he would have had to rely on precedent that interprets the First Amendment to apply to corporations. This presents its own problems, as Part II.C explains: all of the precedent supporting corporate personal constitutional rights traces back to

269. See, e.g., Avi-Yonah, supra note 8, at 40 (arguing that Scalia might appear to adopt the aggregate view, but “what Scalia meant was presumably corporate management working together as an association of persons”).

270. Justice Scalia actually uses the term “artificial legal entity,” but his use of the adjective artificial must not be confused with the term used in this Note, which refers to the corporation as a mere creature of the law, whose only rights are those conferred by the state. See supra note 38 and accompanying text (defining the artificial-entity theory). Justice Scalia seems to use artificial to mean that the legal entity is not a tangible, visible thing but merely one constructed for thinking about the concept of a corporation.

271. See Avi-Yonah, supra note 8, at 39–40 (arguing that Justice Scalia adopts the real-entity view).

272. See, e.g., Citizens United v. FEC, 130 S. Ct. 876, 926 (2010) (Scalia, J., concurring) (finding no support for “the proposition that the First Amendment excludes conduct of artificial legal entities from the scope of its protection”).
one sentence in one 1886 case, which simply declares the right without any discussion.273

3. Justice Stevens's Dissenting Opinion

Justice Stevens provides ample language analyzing corporate nature. For example, he argues that “speech can be regulated differentially on account of the speaker's identity”274 and that this applies especially to corporations because they raise special concerns, such as the potential for corruptive influence.275 By concluding this criticism with a list of other avenues for expressing corporate speech, Justice Stevens seems to adopt an entity viewpoint and assume rights based on that notion.276 He presumes that the corporations are just a “different class[] of speakers,”277 who are “not natural persons, much less members of our political community.”278 Justice Stevens also asserts that the Framers “held very different views about the nature of the First Amendment right and the role of corporations in society”279—that “it was the free speech of individual Americans that they had in mind”280 and that “[a] corporation [was thought of as] an artificial being, invisible, intangible, and existing only in contemplation of the law[,] . . . possess[ing] only those properties which the charter of its creation confers upon it.”281 In particular, he asserts that:

---

273. See supra notes 89–93 and accompanying text (tracing the precedent back to an unsubstantiated opinion declaring that corporations are people within the meaning of the Fourteenth Amendment).
274. Citizens United, 130 S. Ct. at 945 (Stevens, J., dissenting).
275. See id. at 947 (“[L]egislatures are entitled to decide that the special characteristics of the corporate structure require particularly careful regulation in an electoral context.” (internal quotation marks omitted)).
276. See id. at 943 (“The laws upheld in Austin and McConnell leave open many additional avenues for corporations' political speech.”).
277. Id. at 946.
278. Id. at 947.
279. Id. at 949.
280. Id. at 950.
281. Id. (quoting Trs. of Dartmouth Coll. v. Woodward, 17 U.S. (1 Wheat.) 518, 636 (1819)) (providing, within a citation, language of an 1819 case to represent views of the Framers).
[W]hereas we have no evidence to support the notion that the Framers would have wanted corporations to have the same rights as natural persons in the electoral context, we have ample evidence to suggest that they would have been appalled by the evidence of corruption that Congress unearthed in developing BCRA and that the Court today discounts to irrelevance.282

Justice Stevens explains that “corporations are different from human beings” for many reasons, including their properties of limited liability for owners and management, perpetual life, separation of ownership and control, potential for foreign control, an available treasury that reflects economically motivated decisions of investors and customers, and a lack of beliefs, feelings, thoughts, and desires.283 “[T]heir ‘personhood’ often serves as a useful legal fiction. But they are not themselves members of ‘We the People’ by whom and for whom our Constitution was established.”284 He continues to evaluate “who is even speaking when a business corporation places an advertisement,”—not the customers, employees, or shareholders, leaving possibly a few managers who must sometimes overlook even their own personal convictions regarding the message.285 When a four-Justice dissent continuously criticizes “[t]he majority’s unwillingness to distinguish between corporations and humans,”286 the public seems more justified in believing that Citizens United stands for the claim that corporations are people.

In addition to explicit language, Justice Stevens introduces arguments that rely on an understanding of the majority as adopting an entity viewpoint. For example, he criticizes that the majority considers the statute a categorical ban on speech, when actually “every shareholder of every corporation remains entirely free... to do however much electioneering she pleases outside of the corporate form.”287 This criticism indicates that Justice Stevens views the majority opinion as adopting an entity theory and discerning rights from that notion, but he stops short of

282. Id. at 963.
283. Id. at 971.
284. Id. at 972.
285. Id.
286. Id. at 975.
287. Id. at 943.
identifying the problem with deducing rights from this metaphorical corporate entity. Justice Stevens argues that the majority “beg[s] the question what types of corporate spending are constitutionally protected and to what extent.”

Although Justice Stevens criticizes this as circular, the inference (and thus the circularity) cuts the other way: the Court is charged with saying what types of corporate spending are constitutional when a statute poses the question.

In sum, no *Citizens United* Justices attempt to deny that corporations have some personal constitutional rights under the First Amendment. Justices Kennedy (for the majority), Scalia (concurring), and Stevens (for the dissent) each provide justification for the corporate right to political speech in one of two ways: they either improperly derive the right from a mirage of what they believe a corporation is or ought to be, or else they base the right on unexplained precedent. Which of the two justifications each Justice chooses is unclear.

### C. What *Citizens United* Makes Clear

The U.S. Supreme Court has not been careful to provide the public with a consistent image of the corporate form. In fact, the Court’s previous manifestations of the error, sometimes deriving functional rights from metaphysical notions about inherent corporate form, have tainted the public’s impression of the source of corporate rights. Therefore, the reaction to *Citizens United* makes clear that, due to the Court’s cursory

---

288. *Id.* at 934.
289. *Id.* at 930 (“The concept that corporations must be treated identically to natural persons in the political sphere is not only inaccurate but also inadequate to justify the Court’s disposition of this case.”).
290. *See* Bruner, *supra* note 7, at 1388 (explaining, under Hart’s position, that “we make the rules, and it is only by reference to those rules that the corporation can be said to ‘be’ anything at all” (citations omitted)).
291. *See* Krannich, *supra* note 36, at 62 (observing that the Court has frequently used the entity or aggregate theories as metaphors to interpret the Constitution in the corporate context, but that the use has been ad hoc and without regard to any simultaneous mutual exclusiveness among the theories).
analysis, the public has a misguided notion of the source of the political speech right.

The inference the public often draws has two interrelated flaws, the first logical and the second semantic. First, the reasoning is backwards: the public interprets *Citizens United* as using the premise “corporations are people” to deduce traits, but that should be the conclusion—and only in the particular instances when the law treats corporations as persons. Hart identified this error in his claim that defining a corporation by inherent rights is circular. The corporation only has characteristics to the extent that courts and legislatures stipulate them. The public would be marginally better off by saying, “Corporations have the same speech rights as people, so corporations are people,” but the overly broad conclusion does not follow. In a constitutional analysis of personal rights, then, the only proper way to claim that “corporations are people” is in response to the question “What is the extent of corporate rights pursuant to this particular constitutional provision?” Saying “corporations are people” merely means corporations have the same set of rights as people under the law in question; it is a language tool to understand what law applies. Theorists and even courts join the general public in often overlooking Hart’s point when speaking about corporate personhood, but his point follows clearly from the fact that corporations exist by virtue of statutory law, judicial interpretation, and human participation. The *Citizens United* Justices fail to eradicate this misconception, and they possibly even commit the mistake themselves by assuming what they seek to prove: that corporations are entitled to First Amendment protection to the same extent as individuals.


294. See *id.* (explaining that corporations only have traits explicitly granted by law).

295. Similarly, Hart uses the example of a trick in a card game to show how attempting to define it in one context leads to inconsistencies when that definition is “substituted for [the word trick] whenever it is used.” *Id.* at 33. (internal quotation marks omitted).

296. See *supra* Part II.B (discussing the circularity of using theories to deduce functional rights).

297. See *supra* Part IV.B (providing theoretical hints from each Justice’s
Second, the inference commits a category error by using functional language about a particular corporate right to infer a metaphysical claim about the essence of corporations. John Dewey explained that person is merely a term used for convenience to say that, with respect to this particular issue, all the rights of a person apply. Dewey would argue that person is merely meant for the functional sense, and using it in the metaphysical sense would be improper and would pack it full of the real implications that come along with the word. The particular category error evident in the public’s response to Citizens United makes the same move by using functional language to draw a metaphysical conclusion about corporations. Again, the Citizens United Justices—more likely in the concurring and dissenting opinions—possibly commit this error themselves by failing to justify the corporate right to speech in functional terms and, instead, deducing certain rights from a vague conception of corporations.

Each flaw—the logical and the semantic—produces the same consequence: corporate personhood cannot be the source of corporate rights and duties.

D. The Real Source of the Corporate Right in Citizens United

Part of the reason the public misunderstands Citizens United as deriving the corporate speech right from corporate personhood is that the opinion itself does not adequately explain how corporations are endowed with the right. Whereas the public itself typically has brought about many corporate rights by virtue of its influence in legislatures, the Supreme Court confuses the public by establishing a corporate right without explaining the source. Further, the public needs to know the source in order to scrutinize the Court; after all, Citizens United overturned a

language).

298. See Ryle, supra note 210, at 16 (explaining category mistakes).
299. See Dewey, supra note 206, at 655 (arguing that “person” signifies only what the law makes it signify).
300. See id. at 660 (differentiating the two uses).
301. See supra Part IV.B (providing theoretical hints from each Justice’s language).
statute on constitutional grounds, rendering the decision largely immune from future legislative, and thus public, interference. If the corporate personal speech right did not arise from the corporate essence, what is the source?

Obvious sources include statutes or public policy. Or perhaps courts in the past have merely fabricated the right and now courts blindly perpetuate it. The question is, at the moment a right attaches, what is its justification? This question avoids a category error because it is functional, not metaphysical: the justification (if one exists) must be specific and with a legal or social purpose—not abstract or with merely a theoretical purpose. Recall the example of the stork. The stork is just a metaphor to ease discussion; it cannot actually fly because it does not exist. The parents can argue about whether the stork should theoretically be able to fly, but they know it is their discretion—they do not begrudgingly accept some truth that all imaginary storks must fly. In contrast, the reality is that babies are conceived and born; the physical world restricts what they can do, and parents try to keep them in line.

Hiding behind a metaphor allows us to imagine that corporations are exactly what we want them to be in any given circumstance, regardless of contradictions this might pose in reality. We created corporations to be just like people in some ways, but not like people in other ways—often aiming to obtain the best of both worlds, but we forget the inconsistencies a piecemeal nature inevitably brings. When it comes to addressing these inconsistencies or questions about corporate rights, only two boundaries apply: corporations’ physical potential to do certain humanlike activities and the rules that we set.

At issue in *Citizens United* is whether corporations should be permitted to speak, in the form of political independent

---


304. See supra Part IV.A (discussing the “stork” example).

305. For example, with limited liability, we stand to earn huge rewards if the business prevails, but the potential for devastating loss is limited to our investment; if that is exhausted, other members of society foot the bill.
expenditures, to the extent that persons are permitted to do so. While the public sees this decision as the Court bestowing a major liberty on corporations, they do not see why or how; the Court is virtually silent. The Court’s task is to justify the decision and explain—even to the public—why it opines as it does. This is the most important sense in which the Court errs in *Citizens United*—not by its improperly deducing the right from a metaphysical corporate essence but in failing to explain the actual source of the right. The public is left to analyze the decision.

Looking at the moment the Court attaches the personal right of speech to corporations, what must the source be? First, the Court infers the corporate speech right by stating that “First Amendment protection extends to corporations” and citing plenty of cases. Next the Court considers public policy arguments, such as whether the corporate influence in the political sphere might corrupt officials or distort the picture of public opinion. Looking to its predecessor cases, too, indicates public policy justifications. As it appears from the decision, which lacks any explanation for why this personal right extends to corporations, *Citizens United* seems to ground the right in public policy. This presents an irony: the public frames its dissatisfaction around the corporate-rights debate, but as Professor Millon warned, the unnecessary discourse of the corporate-rights debate pushes important policy considerations aside. What makes *Citizens United* important is that the source of the right actually is the party who is most confused about the source of the right—the public. Because the public’s traditional avenue to implement preferred policy into law (through the legislatures) has been removed by this Constitution-based decision, criticizing the Court

---

308. *See, e.g.*, id. at 904–05 (considering the antidistortion rationale).
309. *See Avi-Yonah, supra* note 8, at 33–41 (looking at cases preceding *Citizens United*).
310. *See Milon, supra* note 171, at 58 (stating that theoretical discussions may detract from the focus on important policy considerations).
for finding a corporate personal right is not the most useful discourse.

V. Proper Grounds for Criticizing Citizens United

Per the foregoing discussion, one could criticize the opinion by saying that the U.S. Supreme Court Justices rely on metaphysical language to make a functional claim.\textsuperscript{311} The majority opinion in \textit{Citizens United}, however, seems to cleanly remain in the functional category, and even the concurring opinions make no substantial logical or semantic errors.\textsuperscript{312} If the public views the opinion as not deducing rights from a corporate essence, as this Note suggests it should, our society might still benefit from a critical look at the opinion.

A. Attacking the Court’s View of Policy

One proper form of criticism would be that the Court inadequately defended the policy rationale. The public’s negative reaction to the opinion indicates that the Court, in relying on public policy as the source of the corporate right, might have ruled improperly by finding no sufficient government interest in policy concerns.\textsuperscript{313} Apparently from the language of the majority and concurring opinions and from the public’s reaction, the only policy issues with any steam are the antidistortion rationale and the anticorruption rationale.\textsuperscript{314}

\textsuperscript{311} See, e.g., \textit{Citizens United v. FEC}, 130 S. Ct. 876, 930 (2010) (Stevens, J., dissenting) (arguing that the majority uses a theory as a basis for the right, and that the theory is improper).

\textsuperscript{312} See supra Part IV.A (looking at the Justices’ language).


\textsuperscript{314} See \textit{Citizens United}, 130 S. Ct. at 911 (dismissing the shareholder interest and foreign national arguments with little discussion and on several grounds).
With respect to the antidistortion rationale, the decision did not turn on whether there was a real risk of “an unfair advantage in the political marketplace by using resources amassed in the economic marketplace,” but rather on whether the government had a compelling interest in this protection. The grounds for reviving this issue would be the Court’s lack of due consideration for the effect of distortion. The Court relies only on Buckley to support a rigid claim that the government never has an interest in this regard. But if corporate law has its basis in utilitarian concerns, the public’s overwhelming dissatisfaction might be a sign that the government does have an interest in permitting checks on corporate power, even if it means overturning precedent or allowing restrictions on political speech.

In considering the anticorruption rationale, the Court has rejected that corporate independent expenditures give rise to corruption or the appearance of corruption. In early 2012, the U.S. Supreme Court, in a one-paragraph per curiam decision, reversed a Montana Supreme Court ruling that upheld a statute restricting corporate political expenditures despite Citizens United and despite the fact that it restricted corporate speech.

315. Id. at 904 (citation omitted).
316. See id. ("Buckley rejected the premise that the Government has an interest in equalizing the relative ability of individuals and groups to influence the outcome of elections." (citation and internal quotation marks omitted)).
317. Id. at 904–05 (stating that the government does not have an interest in equalizing abilities to influence elections and that wealth is not sufficient to limit political speech).
318. See Bruner, supra note 7, at 1387 (stating that utilitarianism is “corporate law’s implicit moral theory”).
319. See generally Hasen, supra note 124 (discussing how the antidistortion rationale was improperly orphaned).
323. See id. (reversing the judgment that upheld a law contrary to the Citizens United holding).
The Montana court found a compelling state interest in “preserving the integrity of its electoral process.”\textsuperscript{324} The U.S. Supreme Court initially stayed the decision pending a timely filing and disposition of a petition for a writ of certiorari.\textsuperscript{325} In an order written February 17, 2012, Justice Ginsburg stated:

Montana’s experience, and experience elsewhere since this Court’s decision in \textit{Citizens United v. Federal Election Comm'n}, make it exceedingly difficult to maintain that independent expenditures by corporations “do not give rise to corruption or the appearance of corruption.” A petition for certiorari will give the Court an opportunity to consider whether, in light of the huge sums currently deployed to buy candidates’ allegiance, \textit{Citizens United} should continue to hold sway. Because lower courts are bound to follow this Court’s decisions until they are withdrawn or modified, however, I vote to grant the stay.\textsuperscript{326}

An anticorruption argument would need to refute the \textit{Citizens United} finding that “independent expenditures, including those made by corporations, do not give rise to corruption or the appearance of corruption.”\textsuperscript{327} Then, the government would have to prove its interest in preventing quid pro quo corruption or the appearance of it, but this may be a slighter burden than for the antidistortion rationale.\textsuperscript{328} Indeed, the Court draws the measure of interest from \textit{Buckley}, which upheld limits on direct contributions on anticorruption grounds.\textsuperscript{329}

The Montana Supreme Court proceeded on other grounds, holding that a statute regulating independent corporate expenditures did not violate the corporation’s speech rights and,

\textsuperscript{326} See Supreme Court of the United States, SUPREMECOURT.GOV, http://www.supremecourt.gov/orders/courtdockets/021712zr1.pdf (providing the U.S. Supreme Court order pending the case (citations omitted)).
\textsuperscript{327} Citizens United v. FEC, 130 S. Ct. 876, 909 (2010).
\textsuperscript{328} See id. at 908 (explaining the government’s interest in preventing quid pro quo corruption).
\textsuperscript{329} Buckley v. Valeo, 424 U.S. 1, 63 (1976) (upholding contribution limits).
even if it did, satisfied a compelling state interest in preserving the integrity of the electoral process.\textsuperscript{330} Regardless of its reversal before the U.S. Supreme Court, \textit{Western Tradition} and other cases\textsuperscript{331} blatantly refuting \textit{Citizens United} might indicate a need to limit corporate speech via independent expenditures on public policy grounds. From here, perhaps a future Court of a different composition or a constitutional amendment, however unlikely, would serve to protect the public’s concerns.

\subsection*{B. Attacking the Court’s Application of Precedent}

Another proper form of criticism would be that the Court departed from, or misinterpreted, precedent.\textsuperscript{332} It seems virtually impossible for the Court to properly interpret precedent when the Court has never consistently subscribed to one corporate-rights theory.\textsuperscript{333} One can criticize the court’s application of different metaphors in various corporate-rights cases.\textsuperscript{334} With respect to \textit{Citizens United}, one can use this inconsistency to show that the Court has not properly applied precedent because the precedent itself lacked steady footing.\textsuperscript{335} Alternatively, one can attack the precedent underlying the decision for its use of metaphysical language. This analysis would likely involve a review of the

\textsuperscript{330} W. Tradition P’ship, Inc. v. Attorney Gen., 2011 MT 328, ¶ 48, 363 Mont. 220, 240, 271 P.3d 1, 13 (2011) (stating that when “applying the principles enunciated in \textit{Citizens United}, it is clear that Montana has a compelling interest to impose the challenged rationally-tailored statutory restrictions”).

\textsuperscript{331} See, e.g., Ognibene v. Parkes, No. 09-0994-CV, 2012 WL 89358, at *1 (2d Cir. Jan. 12, 2012) (finding that laws reducing the campaign contribution limits restricting lobbyists’ contributions “are closely drawn to address the significant governmental interest in reducing corruption or the appearance thereof”).


\textsuperscript{333} See Kranich, \textit{supra} note 36, at 62 (observing that the Court has frequently used the entity or aggregate theories as metaphors to interpret the Constitution in the corporate context, but that the use has been ad hoc and without regard to any simultaneous mutual exclusiveness among the theories).

\textsuperscript{334} See id. (criticizing the use of metaphors without regard for consistency).

\textsuperscript{335} See Avi-Yonah, \textit{supra} note 8, at 33–41 (providing support for the real-entity theory in the cases preceding \textit{Citizens United}).
Court’s use of statutory construction to interpret precedent or statutory law. 336

C. Attacking Public Policy as a Legal Basis in Itself

The strongest criticism, if not the lack of a strong policy foundation, might be that the Court’s idea of policy is simply not a good reason to rule in this way. Perhaps other options can bypass the reliance on public policy arguments. Some people (humans, no doubt) have advocated for “a constitutional amendment that removes for-profit corporations from the speech protections of the First Amendment.” 337 Congress could potentially effect a similar result pursuant to its Commerce Clause authority by requiring a Federal Charter stipulating that corporations do not have constitutional protection of personal rights such as speech. Despite the overwhelming strength of corporations in the lobbying sphere and in electioneering, 338 individuals at least carry the right to vote if the majority opposing Citizens United is truly overwhelming.

Although a basis in policy arguments can be weaker than one in case or statutory law, policy can be very important in situations involving the mass public, such as election law. 339 Further, a primary check the public has on the U.S. Supreme Court is to lobby the legislature after an unpopular ruling based on policy or statutory grounds, 340 a concept tainted by the fact

that lobbyists are often corporations or influenced by them. *Citizens United*, however, based its ruling on the Constitution.\(^{341}\) This ups the ante for legislatures, and hence the public, to challenge the ruling: it essentially requires another Court’s decision overturning *Citizens United* or a constitutional amendment.\(^{342}\) Therefore, if the *Citizens United* Court failed to adequately justify its decision, and if the public’s fears materialize, then policy might ultimately be sufficient grounds to challenge the ruling.\(^{343}\)

**VI. Conclusion**

The U.S. Supreme Court has equivocated and even, over time, contradicted itself in its use of corporate-rights metaphors.\(^{344}\) The Court has at times even succumbed to the error of deducing functional corporate rights from some metaphysical essence of a corporation.\(^{345}\) This foundation has perpetuated the error because the public resounds the ruling (with the compounded effect of a media spin),\(^{346}\) and it has alerted the public to rightly question the Court’s conception of corporate form. The public’s overwhelming criticism of *Citizens United* reveals an important point in the discussion of corporate form: the U.S. Supreme Court has provided inadequate guidance for its

\(^{341}\) See *Citizens United v. FEC*, 130 S. Ct. 876, 913 (2010) (stating that the First Amendment does not give the government the authority to restrict corporate political speech).


\(^{343}\) See Weissman, *supra* note 337, at 980–81 (arguing for a constitutional amendment after *Citizens United*).

\(^{344}\) See Krannich, *supra* note 36, at 62 (observing that the Court has frequently used the entity or aggregate theories as metaphors to interpret the Constitution in the corporate context, but that the use has been ad hoc and without regard to any simultaneous mutual exclusiveness among the theories).

\(^{345}\) See, e.g., Hale v. Henkel, 201 U.S. 43, 76 (1906) (using the aggregate view to justify personal due process protection).

\(^{346}\) See, e.g., PBS.ORG, *supra* note 183 (headlining an error that corporate contributions are constitutional, as opposed to independent expenditures—to which the anchors correctly refer).
justification of the corporate right to speech. By virtue of the Court’s failure to explain the source of the corporate right to political speech, *Citizens United* has the unintended consequence of revealing that corporate-rights theories are altogether ineffective as inherent truths about corporations. (Still, with precise language, commentators can use the theories in the metaphysical sense to articulate justifications for various laws, so long as the theories are not used as the source of those laws.) With respect to independent expenditures, the *Citizens United* Court finds the source of the corporate right not in some metaphysical corporate essence, but rather in public policy. Many voices ought to be heard in the political sphere, and the true protection of the public’s voice lies in proper discussion about corporate theory and what we realistically want corporations to be.

---

347. See supra Part III (discussing *Citizens United* and its public impact).
348. See supra Part IV (revealing the error in *Citizens United*).
349. See supra Part IV.D (discussing policy as the source of the right).
350. See supra Part V (providing proper grounds of criticism).
A Case Against the Entire Market Value Rule

Anthony D. Raucci*

Table of Contents

I. Introduction ................................................................... 2233
II. Present State of Patent Damages Law .............................. 2234
III. Enter the Entire Market Value Rule ............................... 2243
IV. Recent Cases Applying the Rule .................................... 2248
V. District Courts and the EMVR ...................................... 2254
VI. Remedies in Copyright Law .......................................... 2257
VII. An Attempt to Impart Clarity ....................................... 2259
VIII. Support from the Federal Rules of Evidence .............. 2264
IX. Harmonizing Patent Damages with the Goals of Patent Law ................................................................. 2267
X. Conclusion ...................................................................... 2269

I. Introduction

Throughout the history of patent law, courts and juries have struggled to determine the value of a single component in a multifaceted device or process when that single component alone is found to have infringed a valid patent.¹ Given the intricacies

* Candidate for J.D., Washington and Lee University School of Law, May 2013; B.S.N.E., University of Florida, 2010. Special thanks to Professors Josh Fairfield, Doug Rendleman, and J.D. King for helpful comments and criticisms.

¹ See, e.g., Westinghouse Elec. & Mfg. Co. v. Wagner Elec. & Mfg. Co., 225 U.S. 604, 617 (1912) (“[Judge Lindley] said in Siddell v. Vickers, 9 Rep. Pat. Cas. 162, that there ‘was no form of account more difficult to work out than an account of profits.’”). As a prelude, note the explicit reference to profits as opposed to the difficulty involved in deriving a value attributed to a certain
inherent in electronic devices today, a large number of patents often cover a single product. The difficulties in placing a value on a single infringed patent incorporated in such a device have become apparent in several recent cases before the United States Court of Appeals for the Federal Circuit. These cases have muddled the rules applied in determining patent-infringement damages. This Note intends to challenge the logic of applying what has become a court-adopted evidentiary exclusion rule, known as the entire market value rule (EMVR), to reasonable royalty determinations. This Note concludes that courts have transformed what was originally a sword for plaintiffs into a shield for defendants without clearly justifying the change under remedies law principles.

II. Present State of Patent Damages Law

Pursuant to constitutional authority, Congress has authorized limited-term monopolies, in the form of patents, to be granted to inventors for their inventions. The limited monopoly seeks to give adequate incentive for undertaking financial risks associated with invention, while still providing competition in the market over the long term. The United States Patent and Trademark Office decides whether to grant or deny a patent after application for one, but when a patent holder alleges

---

component more abstractly. Juries are referenced because the amount of damages in patent infringement cases is a question of fact. See Smithkline Diagnostics, Inc. v. Helena Labs. Corp., 926 F.2d 1161, 1164 (Fed. Cir. 1991) (“[T]he amount of a prevailing party’s damages is a finding of fact.”).


3. See infra Part IV.

4. Thanks to Professor Doug Rendleman for this useful metaphor.

5. U.S. CONST. art. I, § 8, cl. 8; see 35 U.S.C. § 1 (2006) (“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, subject to the conditions and requirements of this title.”).

infringement, federal courts typically decide on the patent’s ultimate validity. A valid patent is infringed if another party makes, uses, or sells the patented invention without the patentee’s consent. The United States Court of Appeals for the Federal Circuit has exclusive jurisdiction over appeals from final decisions of a federal district court when the underlying civil action arises under the patent statute.

Section 284 of Title 35 of the United States Code provides that “[u]pon finding for the claimant [a] court shall award the claimant damages adequate to compensate for the infringement, but in no event less than a reasonable royalty for the use made of the invention by the infringer.” The statute, based on the Patent Act of 1946, has been interpreted since 1964 as precluding the remedy of disgorgement of the infringer’s profits related to a utility patent. In the event of a conscious and willful infringement, the court may decide to “increase the [total damages award] up to three times the amount [of compensation damages] assessed.” Thus, punitive damages are available under the statute. Furthermore, courts may grant injunctions prohibiting an infringer from further infringement subject to the “traditional four-factor framework that governs the award of injunctive relief.”


8. Id.


11. Id.


13. Union Carbide Corp. v. Graver Tank & Mfg. Co., 282 F.2d 653, 675 (7th Cir. 1960). Other situations may also lead to imposition of punitive damages. See Laskowitz v. Marie Designer, Inc., 119 F. Supp. 541, 554 (S.D. Cal. 1954) (stating that a court may award punitive damages when there is oppressive or fraudulent conduct on the part of the defendant).

14. eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388, 394 (2006). The four factors that an injunction-seeking plaintiff must demonstrate are:

   (1) that it has suffered an irreparable injury; (2) that remedies
In court, a patentee has two options when seeking compensation from an infringer. First, the patentee may seek profits that it lost due to the existence of the infringer’s product in the marketplace. In seeking lost profits, a patentee must show that but for infringement, it would have made additional sales or could have charged a higher price for its own product. This can be accomplished, for example, by establishing the four factors expressed in Panduit Corp. v. Stahlin Bros. Fibre Works:

1. The patented feature was in demand;
2. The patentee had the capability to manufacture and market the patented feature to fulfill that demand;
3. There were no noninfringing substitutes available to the infringer; and
4. A profit would have been made on the additional sales.

available at law, such as monetary damages, are inadequate to compensate for that injury; (3) that, considering the balance of hardships between the plaintiff and defendant, a remedy in equity is warranted; and (4) that the public interest would not be disserved by a permanent injunction.

Id. at 391; see also John M. Golden, Injunctions as More (Or Less) than “Off Switches”: Patent-infringement Injunctions’ Scope, 90 Tex. L. Rev. 1399 (2012).

15. See Rite-Hite Corp. v. Kelley Co., 56 F.3d 1538, 1546 (Fed. Cir. 1995) (describing the lost-profits determination as a “but for” analysis limited by “objective foreseeability”).

16. See id. (establishing a but-for test).

17. Panduit Corp. v. Stahlin Bros. Fibre Works, Inc., 575 F.2d 1152, 1156 (6th Cir. 1978). In Panduit, Panduit Corporation successfully sued Stahlin Brothers for infringing a patent “covering duct for wiring of electrical control systems.” Id. at 1155. As part of the proceedings below, a master was appointed to determine Panduit’s deserved damages. Id. The master’s report found that the plaintiff failed to establish sufficient evidence of lost profits, and the report recommended a reasonable royalty. Id. at 1155–56. This report was adopted by the district court below. Id. at 1155. Panduit appealed the decision, arguing that it was entitled to lost profits on lost sales and price competition. Id. In the alternative, it sought a higher reasonable royalty rate. Id. The court of appeals denied the lost profits argument, ruling that there was insufficient evidence to establish the amount of profit it would have made. Id. at 1156. But the court of appeals agreed with Panduit on the insufficiency of evidence to support the master’s reasonable royalty (which was lowered due to his conclusion that there were acceptable noninfringing substitutes on the market). Id. at 1162. The case was remanded to reexamine the royalty based on (1) the lack of acceptable noninfringing substitutes; (2) Panduit’s policy of not licensing the patent at issue; (3) the future business and profit Panduit would expect to lose by licensing at the time of initial infringement; and (4) the fact that the infringed patent gave the entire marketable value to the product sold. Id. at 1164.

18. Id. at 1156. These factors are not the exclusive means of proving lost profits. See Rite-Hite, 56 F.3d at 1545.
If a patentee cannot prove lost profits, the default measure of damages is no less than a reasonable royalty for the infringed patent. A reasonable royalty has been defined as the amount that "a person, desiring to manufacture and sell a patented article, as a business proposition, would be willing to pay as a royalty and yet be able to make and sell the patented article, in the market, at a reasonable profit." Georgia-Pacific Corp. v. United States Plywood Corp. is the seminal case setting forth the factors courts weigh when determining a reasonable royalty. Most often, this involves reconstructing a hypothetical licensing negotiation between a willing licensor and a willing licensee on the date of initial infringement. Typically, an established royalty rate for the same patented component will be the controlling factor. In the absence of an established royalty rate, the other Georgia-Pacific factors are considered in determining

19. See 35 U.S.C. § 284 (2006) (stating that damages adequate to compensate for infringement should "in no event [be] less than a reasonable royalty for the use made of the invention by the infringer").

20. Panduit, 575 F.2d at 1157–58 (citing Goodyear Tire & Rubber Co. v. Overman Cushion Tire Co., 95 F.2d 978, 984 (6th Cir. 1937)).

21. Georgia-Pacific Corp. v. U.S. Plywood Corp., 318 F. Supp. 1116, 1116 (S.D.N.Y. 1970), modified sub nom. Georgia-Pacific Corp. v. U.S. Plywood-Champion Papers, Inc., 446 F.2d 295 (2d Cir. 1971). In Georgia-Pacific, U.S. Plywood (USP) successfully proved infringement of one of its patents for striated fir plywood in the proceedings prior. Id. at 1117. A special master was appointed to determine USP's compensatory damages. Id. The award of $685,837 was based on Georgia-Pacific's (GP) profit on the sale of the infringing article. Id. The judge presiding over the trial took exception to the master's report, and found that GP's profits were not the proper measure of recovery, but rather, the proper measure was a reasonable royalty. Id. The court then analyzed several factors in coming to a reasonable royalty of $50 per thousand square feet of the patented product, amounting to $800,000. Id. at 1120–43. It should be noted that apportionment was not applied in the court's analysis because the "patent was not one for an improvement on an article nor was GP's infringement of a patented feature sold together with unpatented parts. [The patent covered] a marketable article—a panel of striated fir plywood—as an entirety." Id. at 1132.

22. See Unisplay, S.A. v. Am. Elec. Sign Co., 69 F.3d 512, 518 (Fed. Cir. 1995) (stating that the royalty should not be based on an "after-the-fact assessment"). This is an oft-reiterated point of analysis that courts should consider banishing. The benefit of hindsight allows a court to craft an equitable result for both parties.

what a reasonable royalty would have been at the time of initial infringement.24

To fully grasp this Note’s central argument, it will be useful to have an understanding of what these factors are. As stated in Georgia-Pacific, the factors relevant to making a reasonable royalty determination are:

1. The royalties received by the patentee for the licensing of the patent in suit, proving or tending to prove an established royalty.

2. The rates paid by the licensee for the use of other patents comparable to the patent in suit.

3. The nature and scope of the license, as exclusive or non-exclusive; or as restricted or non-restricted in terms of territory or with respect to whom the manufactured product may be sold.

4. The licensor’s established policy and marketing program to maintain his patent monopoly by not licensing others to use the invention or by granting licenses under special conditions designed to preserve that monopoly.

5. The commercial relationship between the licensor and licensee, such as, whether they are competitors in the same territory in the same line of business; or whether they are inventor and promoter.

6. The effect of selling the patented specialty in promoting sales of other products of the licensee; the existing value of the invention to the licensor as a generator of sales of his non-patented items; and the extent of such derivative or convoyed sales.

7. The duration of the patent and the term of the license.

8. The established profitability of the product made under the patent; its commercial success; and its current popularity.

9. The utility and advantages of the patent property over the old modes or devices, if any, that had been used for working out similar results.

24. See Georgia-Pacific, 318 F. Supp. at 1121 (stating that there was an absence of an established royalty, and thus, it was “necessary to resort to a broad spectrum of other evidentiary facts probative of a ‘reasonable’ royalty”).
10. The nature of the patented invention, the character of the commercial embodiment of it as owned and produced by the licensor; and the benefits to those who have used the invention.

11. The extent to which the infringer has made use of the invention; and any evidence probative of the value of that use.

12. The portion of the profit or of the selling price that may be customary in the particular business or in comparable businesses to allow for the use of the invention or analogous inventions.

13. The portion of the realizable profit that should be credited to the invention as distinguished from non-patented elements, the manufacturing process, business risks, or significant features or improvements added by the infringer.

14. The opinion testimony of qualified experts.

15. The amount that a licensor (such as the patentee) and a licensee (such as the infringer) would have agreed upon (at the time the infringement began) if both had been reasonably and voluntarily trying to reach an agreement; that is, the amount which a prudent licensee—who desired, as a business proposition, to obtain a license to manufacture and sell a particular article embodying the patented invention—would have been willing to pay as a royalty and yet be able to make a reasonable profit and which amount would have been acceptable by a prudent patentee who was willing to grant a license.25

If a running royalty is presumed, the reasonable royalty calculation requires the parties to ascertain both the royalty base and the royalty rate.26 The base is the total amount to which the rate is applied in determining the damages owed by the infringer.27 If a lump-sum royalty is presumed, there are additional factors to consider, including the risk discount involved in the lump-sum valuation.28

25. Id. at 1120.
27. Id.
Taking only the above explanation of the current state of the law, it appears that courts have crafted a logical process for evaluating what a patent embedded in a device covered by a sea of patents would be worth in calculating a patentee's damages. Even if a hypothetical negotiation at the time of the infringement seems nearly impossible to ascertain, its ultimate purpose is at least objectively fair: to give the patentee what it would have received had the infringer obtained a license in the first instance. This goal aligns with the patent-statute's goal of providing the patentee with damages adequate to compensate it for the infringement.29

Special attention should be directed to Georgia-Pacific factor thirteen, which offers as relevant to the reasonable royalty calculation “[t]he portion of the realizable profit that should be credited to the invention as distinguished from non-patented elements, the manufacturing process, business risks, or significant features or improvements added by the infringer.”30 This evaluation, historically called apportionment, is arguably the fairest starting point for retroactively forging a reasonable royalty.31 The “realizable profit . . . credited to the invention” impliedly requires introducing into evidence the market value of the product that includes the patented component; otherwise, there is no base profit value to apportion.32 Similarly, factor eight refers to the “established profitability of the product made under the patent;”33 the language implies a consideration of the profits made after a hypothetical negotiation, but before the instant

31. See Eric E. Benson & Danielle M. White, Using Apportionment To Rein In the Georgia-Pacific Factors, 9 COLUM. SCI. & TECH. L. REV. 1, 46 (2008) (arguing that “apportionment should be a required part of every reasonable royalty analysis, not just a factor that may or may not be given any weight”). The Article continues with the argument that the reasonable royalty calculation should not only be bounded by the profits attributable to the infringed patent but also by some portion of those profits since a prudent licensee would not seek a license if it did not yield some profit for itself. Id. at 47.
33. Id.
litigation. Thus, it is important to note that apportionment has been recognized to be one element of the reasonable-royalty analysis.

Patent damages are meant to be compensatory, meaning the measurement should be based on the plaintiff's loss, rather than the defendant's gain. The damages determination starts with the patentee trying to prove its lost profits, and if that fails, then it tries to prove a reasonable royalty based on a hypothetical negotiation informed by other factors. Although this Note makes the argument that a reasonable royalty should almost always be informed by the profits on the infringing product, presently, the reasonable royalty analysis is being perverted to exclude such profits in instances involving a product covered by numerous patented and nonpatented components.

Eric Benson and Danielle White propound that a hypothetical negotiation seeking to compensate a patentee via a reasonable royalty should be informed by the actual profits made on a product incorporating the infringed patent—stating that “apportionment should be a _threshold_ determination in every reasonable royalty analysis.” There is much to be said for this suggestion. Less convincing, however, is their necessarily bounding a patentee’s potential royalty by the infringer’s profits. Under such a limitation, a patentee’s compensation for the lost opportunity to bargain for a license could be circumvented by an inefficient infringer. Nevertheless, Benson and White’s argument implicitly recognizes the importance of using the income attributable to the entire product even if the patented component at issue is not the basis for the product’s demand in the market.

The reason for introducing an infringing product’s profits, even if the infringed patent does not contribute the basis for demand for that product, is that it gives the jury concrete

---

35. See infra Part III (discussing the EMVR).
37. See _id._ at 2 (“By treating apportionment as a threshold question, courts can ensure that the resulting reasonable royalty award is properly confined to a _portion_ of the profit attributable to the patent.”).
evidence of what the patent might be worth to the infringer, and this evidence would be relevant to determining what amount the infringer would have agreed to pay as a royalty. With such evidence before a jury, it is unlikely that any infringement of a minor component, taken alone, would threaten the infringing entity with major economic disruption. It therefore helps to avoid uninformed awards that may wander into the punitive category. At the same time, the patentee’s compensation is necessarily in correlation to the patent’s contribution to the product incorporating it rather than some abstractly determined value. An infringer would likely have made the licensing decision with respect to the perceived value of the overall product. Actual profits provide an objective, reliable value for the jury in cases in which the opposing parties’ experts often provide largely divergent royalty estimates.

Even if the compensatory nature of patent damages could rebut bounding reasonable royalties by an infringer’s profits, the profits on the infringing product should still be considered an essential factor in a hypothetical negotiation. But the importation of the EMVR into reasonable royalty determinations has, in effect, needlessly made this factor nearly impossible to get into evidence.

38. Note that the present state of the law does not provide a right to an accounting of profits. See Georgia-Pacific Corp. v. U.S. Plywood Corp., 318 F. Supp. 1116, 1120 (S.D.N.Y. 1970), modified sub nom. Georgia-Pacific Corp. v. U.S. Plywood-Champion Papers, Inc., 446 F.2d 295 (2d Cir. 1971) (referring to profits as a factor to be considered). The proposed method essentially seeks to place far greater weight on actual profits than present law effectually allows in the reasonable royalty analysis.

39. See, e.g., Uniloc USA, Inc. v. Microsoft Corp., 632 F.3d 1292, 1318–20 (Fed. Cir. 2011) (stating that Microsoft’s expert suggested a royalty of $7 million, and Uniloc’s expert suggested a royalty of nearly $565 million).

40. The language of the patent damages statute does not clearly preclude classifying reasonable royalties as restitutionary. Compare 35 U.S.C. § 284 (2006) (providing that in no event should a patentee of an infringed patent be awarded less than a “reasonable royalty”), with RESTATEMENT (THIRD) OF RESTITUTION AND UNJUST ENRICHMENT § 50 cmt. e (2011) (providing that the remedy in restitution for an innocent recipient of some unjust enrichment is measured by a “reasonable value” standard).
Before addressing a present point of confusion in patent damage calculations, it is useful to examine the definition and origin of the EMVR. In seeking to elicit the present justification behind the rule, this Note will compare the rule’s original logic to how it is being applied by courts today.

The EMVR, most near its original understanding, allows a patentee to recover damages that encompass unpatented features sold with the patented feature if: (1) the patented feature is the basis for customer demand for the entire product or products sold; (2) the patentee reasonably anticipates the sale of the unpatented parts along with the patented component; and (3) there is a functional relatedness between the patented and unpatented components. The idea is that whatever value was contributed by the unpatented components would not have been realized without the patented component—it was the reason the entire product had market value at all. Note that the entire market value rule is not clearly an evidentiary exclusion rule; rather, it is an affirmative grant of a degree of recovery provided the requisite elements are met. Plainly interpreted, the rule allows the patentee to recover the full value of an infringing product if the infringing component, on its own, drove demand.

Presumably, the patentee would have captured the value of the unpatented components itself, but for infringement. Thus, the rule seems to make sense when viewed in conjunction with the lost-profits analysis. But, in actuality, the EMVR does not fully comport with the language of the patent statute even in lost-profits scenarios absent proof of the patentee’s plans and capability to manufacture the entire device.

One of the earliest cases in which the rule appears is Garretson v. Clark. The patent at issue in Garretson was an

---

42. See id.
43. See id. at 1549 (“We have held that the entire market value rule permits recovery of damages based on the value of a patentee’s entire apparatus containing several features when the patent-related feature is the ‘basis for customer demand.’”).
44. Garretson v. Clark, 111 U.S. 120 (1884).
improvement for a method of fastening mop-heads. The patentee proved infringement and sought all of the infringer's profits in damages. The U.S. Supreme Court affirmed the lower court's denial of all of the infringer's profits, invoking the EMVR:

The patentee . . . must in every case give evidence tending to separate or apportion the defendant's profits and the patentee's damages between the patented feature and the unpatented features, and such evidence must be reliable and tangible, and not conjectural or speculative; or he must show, by equally reliable and satisfactory evidence, that the profits and damages are to be calculated on the whole machine, for the reason that the entire value of the whole machine, as a marketable article, is properly and legally attributable to the patented feature.

Thus, Garretson provided two clear alternatives: (1) apportion the infringer's profits and the patentee's lost profits according to the value of the patent, or (2) prove that the entire value of the product is derived from the patented component to receive the full value of the profits associated with the product, sold by the patentee, which incorporated the patented component. Note how these two choices are restitutionary in nature, with the damages being measured by the defendant's gain attributable to the patented component. There was no requirement that the patentee prove manufacturing capability or any other factor supporting a claim of actual loss.

If Garretson had produced the requisite evidence to prove that his improvement was worth 25% of the value of a mop, the rule would apportion to Garretson that amount of the infringer’s net profits (this restitutionary measure is no longer available). Note that an improvement that contributed 25% of the value of a mop would mathematically not constitute the basis of consumer demand, yet the original rule (in conjunction with the apportionment alternative) would necessarily allow into evidence the entire amount of profits made on the infringing product. This

45. Id. at 121.
46. Id.
47. Id.
48. Id.
49. See supra note 12 and accompanying text.
understanding of the rule continued into the early twentieth century.

In *Westinghouse Electric & Manufacturing Co. v. Wagner Electric & Manufacturing Co.*, the U.S. Supreme Court found Wagner to have infringed a patent for an electrical transformer incorporating a nonconducting cooling oil. Prior to reaching the Supreme Court, it was determined that Wagner had made a profit of $132,433 on infringing transformers, but because Westinghouse failed to prove what portion of the profits was attributable to the infringement, it was only awarded nominal damages. Note that this case was decided before the present patent-damages statute was enacted. In this case, the Court correctly delineated between apportionment and recovery of all of an infringer's profits. The Court, in describing the law, stated that when the patent “is only part of the machine and creates only part of the profits[,] ... [the plaintiff] is only entitled to recover that part of the net gains.” It then quoted Garretson's delineation between apportioning defendant's profits and awarding the entirety of the profits (as would be the case if the EMVR were satisfied).

Although the Court sensibly explained the EMVR in *Garretson*, later, lower court decisions expanded it from a rule requiring a threshold to be met in recovering all of the infringer's profits to a rule that excluded evidence of the infringer's profits in the reasonable royalty context. The confusion appears to have originated in *Marconi Wireless Telegraph Co. v. United States*, when the United States Court of Claims found that the patent at

51. *Id.* at 607.
52. *Id.* at 614.
53. *Id.* at 614–15.
54. *Id.* at 615.
55. Marconi Wireless Tel. Co. v. United States, 99 Ct. Cl. 1 (1942), aff’d *in part and vacated in part on other grounds*, 320 U.S. 1 (1943). The Court of Claims was abolished in 1982. Note that while the Court of Claims was meddling with the EMVR's original application, the Court of Appeals for the Seventh Circuit was remaining true to it. See *Cover v. Chi. Eye Shield Co.*, 130 F.2d 25, 28–30 (7th Cir. 1942) (affirming the master's finding that, because the entire market value of infringing device was not attributable to the patent at issue, the profits from the device had to be apportioned).
issue “substantially created the value of the component parts utilized in radio transmitters and receivers purchased . . . and that it therefore falls within the [EMVR]. The complete cost of the transmitting and receiving sets [was] used as the base in the application of a reasonable royalty.” 56 Marconi dealt with a reasonable royalty, a compensatory measure, and thus it can be seen as expanding the applicability of the EMVR. The Court of Claims did not claim that Marconi’s patent was the basis for consumer demand; rather, it created a new standard, claiming that it “substantially created the value of the component parts.” 57

The court in Marconi applied the EMVR and lowered the standard—from “the basis of consumer demand” to “substantially created value”—but did not totally abrogate the rule. The court did not explicitly provide its reasons for doing so, but based on a reading of the facts, some possible justifications present themselves. First, the court may have recognized that in the valuation of a component in a device as complex as the radio was at the time, it would be difficult to pinpoint one component that was the basis for demand. Implicit in this reasoning is that the market value of the product is an essential tool in placing a value on a single component of that product. Second, creating a tension with the first point, the court likely recognized that, as devices were beginning to interact with large numbers of other devices, it would become necessary to limit the royalty base in some manner to protect against applying inherently uncertain royalty rates to exorbitant royalty bases. A good example of this today is a patentee of a computer program (other than something as integral as an operating system) seeking damages based on the market value of a desktop computer sold by the infringer.

Although these presumed concerns are valid, the Marconi court’s standard was far too restrictive when applied, essentially, as an evidentiary rule in reasonable royalty scenarios—that is, by excluding the profit made on the entire product from evidence when the standard is not met. A major problem with the court’s reasoning is its belief that it had to invoke the EMVR, a progeny of restitution, when in fact it was calculating a compensatory reasonable royalty. Thus, in Marconi, two remedies theories—

56. Id. at 49.
57. Id.
those of damages and restitution—were obscured and a legal
device was forged to limit what is often times the only reliable
evidence available to ensure that a royalty is indeed reasonable.
It is important to note that in 1952, Congress passed the present
patent statute, including 35 U.S.C. § 284, which provided for
damages “adequate to compensate for the infringement, but in no
event less than a reasonable royalty for the use made of the
invention by the infringer.” It is difficult to rectify the continued
existence of the restitutionary EMVR with this codification.

Interestingly, the “reasonable royalty” measure of damages
arose out of evidentiary proof difficulties when courts sought to
apportion profits in disgorgement scenarios. Courts have
interpreted the patent statute’s granting of compensatory
damages or a reasonable royalty as precluding the remedy of
disgorgement. While this interpretation is arguably
questionable, it need not be erroneous to logically align with
this Note’s modest proposal. Determining a reasonable royalty
and apportioning an infringer’s profits both initially aim to find
what value the patent contributes to the overall product. The
patent statute’s reference to a “reasonable royalty” may preclude
disgorgement of profits related to an infringed patent, but it does
not necessarily preclude calculating a royalty in reference to the
infringer’s profits. Thus, a balance can be struck between full
disgorgement under a restitutior•nary theory and the exclusion of
profits entirely from consideration in a reasonable royalty
calculation. A more nuanced evidentiary standard makes more
sense than a categorical bar.

59. See id. at 661 (“[C]ourt opinions interpreted the alteration as eliminating the patent owner’s traditional equitable remedy of an accounting of the infringer’s profits.” (citations and internal quotation marks omitted)). The legislative history for the patent statute does elicit an attitude against using apportionment in every infringement suit. See S. REP. No. 79-1503, at 2 (1946) (stating that it is impossible to apportion profits due to an improvement patent in a complex machine and referring to the associated long, expensive road to dubious justice in that pursuit).
60. See Roberts, supra note 58, at 661–64 (discussing the ambiguous reasoning behind excising an accounting of profits from the patent statute).
61. See supra Part II (discussing the patent remedies law).
62. See Roberts, supra note 58, at 683–84 (arguing that under existing law, reference to an infringer’s profits should be allowed when there is a relevant
IV. Recent Cases Applying the Rule

To illustrate the issue, it is necessary to analyze how recent cases have applied the EMVR. In *Lucent Technologies, Inc. v. Gateway, Inc.*,63 Lucent initially brought a patent infringement claim against Gateway for a method patent involving entering information into fields on a computer screen without the use of a keyboard.64 Microsoft Corporation subsequently intervened, and Lucent alleged indirect infringement of the patent by Microsoft’s sale of Microsoft Money, Microsoft Outlook, and Windows Mobile.65 The jury awarded $357,693,056.18 to Lucent after finding Microsoft liable on claims covering all three products.66 On appeal, Microsoft argued that “the jury should not have applied the entire market value rule to the value of its three software products.”67 At trial, the parties agreed that the measure of damages was to be determined using a hypothetical negotiation in accord with *Georgia-Pacific*.68 Under this approach, Lucent sought a running royalty, but Microsoft argued it would have paid a lump sum of $6.5 million for the rights to incorporate the patented method in its products.69 Because the jury indicated that its award was a lump-sum payment, the U.S. Court of Appeals for the Federal Circuit sought to determine whether there was substantial evidence to find that Microsoft would have paid nearly $358 million up front for the use of the patent.70

In support of its finding that there was not substantial evidence to support the award, the court cited precedent stating that “[t]he issue of the infringer’s profit is to be determined not on the basis of a hindsight evaluation of what actually happened, but on the basis of what the parties to the hypothetical license negotiations would have considered at the time of the

64. *Id.* at 1308; see also U.S. Patent No. 4,763,356 (filed Dec. 11, 1986) (describing the patented technology).
65. *Lucent*, 580 F.3d at 1308.
66. *Id.* at 1309.
67. *Id.* at 1323.
68. *Id.* at 1325.
69. *Id.*
70. *Id.*
A CASE AGAINST THE ENTIRE MARKET VALUE RULE

negotiations.” The court seemingly disregarded this logic, however, in claiming that evidence regarding the frequency of use by the infringing users of the software programs was not barred, even though that particular evidence originated after the hypothetical negotiation. It is unclear why evidence of actual use of the patented method would be helpful in determining the reasonableness of a royalty while evidence of actual profit from the infringing product would not be. If “[e]xperience is . . . available to correct uncertain prophecy” on evidence of actual use, why not on evidence of actual profits? Both would necessarily be matters of speculation during a hypothetical negotiation.

Microsoft claimed that the jury must have applied some percentage rate to the entire market value of its infringing products relevant to the award determination. Microsoft claimed that this was an improper application of the EMVR. The court recognized that “it is difficult to understand how the jury could have chosen its lump-sum figure down to the penny unless it used a running royalty calculation.” Note that from a negotiation standpoint, while there is a difference between a lump sum and a running royalty, the lump sum will be mathematically connected to a running royalty—as determination of a lump sum will be based on a speculated running royalty, discounted to present value, with further discounts for risk.

Here, the court agreed that, if the jury applied the EMVR, it would have been error. This statement is correct as applied to

71. Id. at 1333 (quoting Hanson v. Alpine Valley Ski Area, Inc., 718 F.2d 1075, 1081 (Fed. Cir. 1983)).
72. See id. at 1333–34 (“[Evidence of frequency of usage], assuming it meets admissibility requirements, ought to be given its proper weight, as determined by the circumstances of each case.”).
73. Id. at 1333 (quoting Sinclair Ref. Co. v. Jenkins Petroleum Process Co., 289 U.S. 689, 698 (1933)).
75. Id. at 1323.
76. Id.
77. Id. at 1336.
78. See id. at 1327 (analyzing an expert’s determination of a lump sum).
79. Id.
the historical meaning of the EMVR. Recall that the EMVR was originally applied to allow the patentee to recover all of the infringer’s profits when a product was not entirely covered by an infringed patent, but the patent was nevertheless the basis for consumer demand. The high threshold of proving that a component was the basis of demand was presumably necessary because depriving the infringer of the entirety of its profits is a harsh result. But the court of appeals did not follow this line of reasoning.

The court found two problems with applying the EMVR, which it interpreted as allowing an award to be associated with the entire market value of a product even if a low royalty rate was applied. First, there was no evidence that the patent provided the basis of consumer demand. Second, the court criticized the methodology employed by Lucent’s expert. The expert initially applied a royalty rate to the price of an entire computer, but when the district court excluded this from evidence, the expert increased his royalty rate as applied to the software price and came to the same number as when he used the aggregate computer sales as the royalty base. Although the court should be concerned with deceptive tactics, it appears that this criticism of the expert’s royalty-rate inflation weighs more toward the quality of the evidence and does not justify excluding the price of the software from the royalty base if the patentee could prove what value of that price was profit attributable to the infringed patent.

The court itself acknowledged this point in dictum, which was rebuked in subsequent cases, when it stated, “Simply put, the base used in a running royalty calculation can always be the value of the entire commercial embodiment, as long as the

80. See supra notes 41–57 and accompanying text (explaining the EMVR).
81. See Lucent Techs., Inc. v. Gateway, Inc., 580 F.3d 1301, 1337 (Fed. Cir. 2009) (“The first flaw with any application of the entire market value rule in the present case is the lack of evidence demonstrating the patented method of the Day patent as the basis—or even a substantial basis—of the consumer demand for Outlook.”).
82. See id. at 1338 (“The second flaw with any application of the entire market value rule in this case lies in the approach adopted by Lucent’s licensing expert.”).
83. Id.
magnitude of the rate is within an acceptable range (as determined by the evidence).” It was not necessary for the court to rule on the issue of whether the jury erroneously applied the EMVR because it found that there was not substantial evidence to support the damages award regardless, but it is more than apparent that the award would have been reversed on EMVR grounds had it been necessary to decide the issue. The damages award did not give the entire market value or all of the profits to the patentee. Thus, when the court claimed that a lack of evidence proving the patented component was the basis for consumer demand of the software programs was fatal to the application of the EMVR, it is apparent that the court believed the EMVR was applied even though the award amounted to approximately 5.5% of the total sales revenue of the infringing product. The court implied that the EMVR is applied when the market value of the product is used as the royalty base. This implication parallels the reasoning in Marconi but runs counter to the court’s dicta quoted above. The tension in the court of appeals’s logic would be resolved—incorrectly, this Note asserts—two years later in Uniloc USA, Inc. v. Microsoft Corp.

In Uniloc, the patent at issue was directed at preventing unauthorized installation of copied software. Uniloc alleged that Microsoft’s Product Activation feature on Microsoft Word XP, Word 2003, and Windows XP infringed the patent. The jury

84. Id. at 1338–39.
85. See id. at 1338 (arguing the flaws with applying EMVR in this case). Again, because the patentee was not seeking the entirety of Microsoft’s profits, the EMVR, as originally understood, was not applied. Note how applying the “substantial evidence” standard allowed the court to overturn the large damages award without any need for the EMVR. This provides further support for abrogating the EMVR entirely.
86. Id. at 1336–38. It is not clear why the experts did not focus on the net profits, which are what this Note encourages courts to look to in solving the problem of unjust damage awards. Of course, this evidence should be subject to other admissibility considerations.
87. See infra Part III (discussing Marconi).
88. Uniloc USA, Inc. v. Microsoft Corp., 632 F.3d 1292 (Fed. Cir. 2011).
89. Id. at 1296; see also U.S. Patent No. 5,490,216 (filed Sept. 21, 1993) (describing the patented product). An understanding of the functionality of the patented process is not necessary for the purposes of this Note.
90. Uniloc, 632 F.3d at 1296.
found infringement and awarded Uniloc $388 million.\footnote{Id. at 1311.} Uniloc's expert explained to the jury, based on a document internal to Microsoft, that product activation keys could be worth anywhere from $10 to $10,000 each, depending on usage—instead of installing a copy of the software, users will presumably have to buy the software to get a valid key.\footnote{Id. (citing Uniloc USA, Inc. v. Microsoft Corp., 632 F. Supp. 2d 147, 150 n.2 (D.R.I. 2009)).} The expert then took the lowest estimated value, $10, and assumed this was the isolated value of the patent per product license sold.\footnote{Id.} He then utilized the 25% rule-of-thumb, which presumes that 25% of the value of the patent would go to the patent owner and the rest would go to the party selling the product.\footnote{Id.} Thus, $2.50 per license was attributable to the patent.\footnote{Id.} After finding that the Georgia-Pacific factors balanced out, the expert applied the $2.50 rate to the total number of Office and Windows product licenses sold and determined that $565 million was a reasonable royalty.\footnote{Id.} As a check to this calculation, the expert then estimated gross revenues for the products by multiplying the total number of licenses by the average sales price per license, yielding a gross revenue value of $19.28 billion.\footnote{Id.} Thus, the royalty rate turned out to be 2.9% of the estimated gross revenues.\footnote{Id.} He then opined that this was reasonable because, in his experience, royalty rates for software are on average generally above 10%.\footnote{Id. at 1318.}

This methodology of creating the value of the infringed patent from scratch, instead of first apportioning profits, should only be resorted to in the rare case in which a defendant cannot produce evidence of the profits for the larger device of which the infringed patent is a component. The concern that the conclusion in the expert's methodology might be seen as fair, because there is a low royalty percentage, is deceiving partially because he is comparing the royalty to revenues instead of profits; the
manufacturer is bearing all the risk, and this is reflected in the market price. This illustrates why the reasonable royalty calculation should, in most instances, be tied to net profits. Of course, it would be more helpful if the parties actually presented evidence on net profits instead of just gross revenues.

Microsoft contested the admission of the expert’s testimony, alleging that it used the EMVR, even if it was only part of the check on the reasonable royalty calculation. Microsoft’s contention was based on Uniloc’s lack of evidence proving that the product activation feature was the basis of consumer demand for the product. The district court granted a new trial on damages based on this alone, stating that the “$19 billion cat was never put back into the bag.” The U.S. Court of Appeals for the Federal Circuit affirmed. First, the court found that the application of a 25% rule-of-thumb was arbitrary and inadmissible under Daubert v. Merrell Dow Pharmaceuticals, Inc. and the Federal Rules of Evidence. Second, the court addressed Microsoft’s allegation that the expert’s check was an inadmissible application of the EMVR. After aptly laying out the law, citing Lucent, Rite-Hite, and Garretson, the court of appeals agreed with the district court’s finding that the “entire market value of Office and Windows in the form of the $19 billion figure was ‘irrelevant’ and ‘taint[ed]’ the jury’s damages award.” The court of appeals then resolved the tension in logic described in Lucent, stating that “[t]he Supreme Court and this court’s precedents do not allow consideration of the entire market value of accused products for minor patent improvements simply by asserting a low enough royalty rate.”

100. Id. at 1312.
101. Id.
102. Id.
103. Id. at 1321.
104. Id. at 1315 (citing Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579 (1993)).
105. Id. at 1318.
106. Id. at 1319 (alteration in original).
107. Id. at 1320 (citations omitted). It seems that there is a misinterpretation of the phrase based on. As originally intended, to award damages based on the entire value of the product means to give the entirety of the profits. See Garretson v. Clark, 111 U.S. 120, 121 (1884) (focusing on the profits to determine damages).
Following this logic, evidence of the entire market value of a product will not even be admitted unless the patented component is proven to provide the basis of demand. What started as a standard for finding that a patentee was entitled to the entire value of a product—a sword for plaintiffs—has essentially been broadened into an evidentiary admissibility standard—a shield for defendants. Considering the complexity inherent in technology today, there is little chance that a single component will ever be the basis of consumer demand. The market value of a product in which a patented component is incorporated informs rather than taints a jury in its quest of finding a reasonable royalty. The Federal Circuit’s decision with respect to the EMVR is blind to the patent damages statute’s construction and to how the EMVR has been perverted since it was promulgated by the Supreme Court.

V. District Courts and the EMVR

The error in the application of the EMVR is spreading. It can be found in district courts that are now relying on Uniloc. In Mirror Worlds, L.L.C. v. Apple, Inc., the patents at issue involved a document stream operating system and method that stored documents in a chronological order that was transparent to the user. The user could scroll through the documents in a stack and see a miniature version of the document before accessing it. The patentee claimed that Apple’s operating systems and mobile devices (the iPhone, iPod, and iPad) infringed the patents. To determine the revenue pool, Mirror Worlds “used both software and hardware incorporating the accused [operating system] to calculate a $72 billion royalty base.” Mirror Worlds argued that the EMVR did not apply. This argument was correct as applied to the originally promulgated

109. Id. at 708.
110. Id.
111. Id. at 709.
112. Id. at 726.
113. Id.
EMVR. Mirror Worlds was not seeking the entire market value in damages. In fact, Mirror Worlds was attempting to apportion the value of the infringed patent based on the total value of the product in which it was incorporated as a starting point for finding a reasonable royalty. In such a scenario, it is illogical to exclude from evidence the total value of the product. If Mirror Worlds were able to provide substantial evidence that its patent constituted precisely 1% of the market value of the product (including the hardware), wouldn’t the starting point for a just damages award be 1% of the profits relating to the infringing product? And from there, the 1% profit figure could be tweaked based on a hypothetical negotiation, presuming a licensee would not negotiate to give up all profits related to the license. Including the hardware in the royalty base is arguably reasonable because an operating system is necessary to make the hardware useful for a consumer; thus, a sufficient functional relatedness exists between the patented component and the product incorporating it.

Based on the following excerpt from the opinion, it seems the court would exclude the profits from the evidence unless the patentee could prove that the infringed features created “the basis for customer demand” or “substantially create[d] the value of the component parts.”

Despite Mirror Worlds’ protestations that the entire market value rule does not apply, it undisputedly used the entire market value of the accused commercial products in calculating its royalty base—and the accused products contain several features, both accused and non-accused. Therefore, Mirror Worlds must show the connection between the accused commercial products, which form its royalty base, and the patented features. Accordingly, at trial Mirror Worlds was obligated to show that the Spotlight, Cover Flow, and Time Machine features create the “basis for customer demand” or “substantially create[d] the value of the component parts” in the accused software and hardware products that contain Mac OS X 10.4–6.

114. Id.
115. Id. (internal quotation marks omitted) (quoting Uniloc USA, Inc. v. Microsoft Corp., 632 F.3d 1292, 1318 (Fed. Cir. 2011)).
116. Id. (quoting Uniloc USA, Inc. v. Microsoft Corp., 632 F.3d 1292, 1318 (Fed. Cir. 2011)).
From this excerpt, it is apparent that the district court has now imported two concepts present in the history of patent damage calculations that were originally applied in both lost profits and apportionment scenarios. In apportionment, the theory of damages sought to deprive the infringer of its unjust enrichment, and thus, the patentee was entitled to deprive the infringer of the entire value attributed to the patent.117 This is harsher than the present statute because a reasonable royalty would not deprive a hypothetical licensee from the entirety of the profits associated with the licensed patent because a licensee would expect some profit from the deal.118 Even so, apportionment is a necessary element and should be treated as such in the reasonable royalty determination, as discussed in Part II.

In lost-profits scenarios, the recovery of all of an infringer's profits may sometimes comport with the adequate compensation standard; here the patentee may recover more than just the precise value of the patent as incorporated in the infringer's product; it may also recover the value from reasonably anticipated market transactions associated with the patent and lost sales. Thus, if the infringed patent was the basis of demand, the patentee could have arguably made all of the infringer's profits. But Mirror Worlds did not seek all of Apple's profits. Rather, it sought a reasonable royalty based on the precise value of its patent with respect to Apple's infringing products. The cases discussing the EMVR following Lucent v. Gateway unreasonably apply the same high standard—the "basis of consumer demand" test—to two distinguishable subsets of compensatory damages.119

117. See Dowagiac Mfg. Co. v. Minn. Moline Plow Co., 235 U.S. 641, 646 (1915) (stating that the burden of proof was on the patentee because it was entitled to recover the portion of the profits attributable to the patent). Dowagiac ended with a remand to determine apportionment, in part because of the Westinghouse decision. Id. at 651.

118. See Panduit Corp. v. Stahlin Bros. Fibre Works, Inc., 575 F.2d 1152, 1157–58 (6th Cir. 1978) ("A reasonable royalty is an amount 'which a person, desiring to manufacture and sell a patented article, as a business proposition, would be willing to pay as a royalty and yet be able to make and sell the patented article, in the market, at a reasonable profit.'" (quoting Goodyear Tire & Rubber Co. v. Overman Cushion Tire Co., 95 F.2d 978, 984 (6th Cir. 1937)).

119. See Pall Corp. v. Micron Separations, Inc., 66 F.3d 1211, 1223 (Fed. Cir. 1995) ("[T]he purpose of compensatory damages is not to punish the infringer, but to make the patentee whole.").
VI. Remedies in Copyright Law

It seems absurd to declare evidence of the actual revenues as “irrelevant” to a reasonable royalty calculation when the projected revenues would have most certainly been a part of that determination in a hypothetical negotiation. To address this absurdity, this Note suggests that courts draw experience from copyright law. In particular, it may be of interest to look at how courts determine remedies in a copyright infringement suit in which the infringing work has substantial noninfringing components. Under the copyright statute:

The copyright owner is entitled to recover the actual damages suffered by him or her as a result of the infringement, and any profits of the infringer that are attributable to the infringement and are not taken into account in computing the actual damages. In establishing the infringer’s profits, the copyright owner is required to present proof only of the infringer’s gross revenue, and the infringer is required to prove his or her deductible expenses and the elements of profit attributable to factors other than the copyrighted work.

Thus, copyright law provides a copyright holder with a right to choose between compensation and restitution. Any reasonable infringer would do everything possible to present evidence of expenses and attributes of the work not covered by the plaintiff’s copyright in attempting to reduce a restitutionary award. Drawing some parallels between the copyright and the patent statutes, one initially finds that the patent statute grants damages “adequate to compensate for the infringement,” while the copyright statute grants “actual damages” suffered as a result

120. See infra Part VII (discussing the elements of licensing valuation).

121. Cf. Sony Corp. of Am. v. Univ. City Studios, Inc., 464 U.S. 417, 439 (1984) (finding it appropriate to refer to patent law as an analogy in a copyright case “because of the historic kinship between patent law and copyright law”).


123. See Dobbs, supra note 6, at 54 (“In effect, the owner may recover either damages specified by the statute or the infringer’s profits, whichever sum is higher.”); see also RESTATEMENT (THIRD) OF RESTITUTION & UNJUST ENRICHMENT § 42, cmt. d, illus.1 & 2 (2011) (illustrating the same result under the copyright statute).

Both forms of recovery are compensatory. The patent statute then provides that in no event should an award be less than a reasonable royalty. This language has no requirement that the plaintiff prove its actual loss, and thus is not clearly compensatory in every case. In a case where a patentee was not seeking to license or produce under the patent, if a reasonable royalty is awarded, how is the award classified? It appears such an award could be labeled quasi-restitutionary. In such a case, looking to the defendant’s profits seems even more justified. Instead of a reasonable royalty floor to recovery, the copyright statute grants the right to completely disgorge profits attributable to the infringement from the infringer.

Whereas the copyright holder may choose between compensation and full restitution, the patent holder may only choose compensation. The lack of a restitutionary option in patent law arguably results in insufficient deterrence. When disgorgement is an option, as in copyright law, the infringer gives up all gains from the infringement—leaving the infringer indifferent to the act of infringement. When recovery is based solely on compensation, a liability rule has been adopted, and a rational infringer will choose to infringe when its expected gain exceeds the patentee’s expected loss. This gives an infringer every incentive to prevent a patentee from establishing a loss. The patent statute does provide punitive deterrence in allowing

127. See Dobbs, supra note 6, at 34 n.2 (stating that “a plaintiff who is not exploiting a patent at all will still be entitled to a reasonable royalty”).
128. See Restatement (Third) of Restitution and Unjust Enrichment § 50 cmt. e (2011) (providing that the remedy in restitution for an innocent recipient of some unjust enrichment is measured by a “reasonable value” standard).
129. See Robert D. Cooter, Punitive Damages, Social Norms, and Economic Analysis, 60 L. & CONTEMP. PROBS. 73, 77 (1997) (“[P]erfect disgorgement restores the injurer to the same indifference curve as if no injury had occurred.”).
130. See id. at 88 (“[A]ssuming certain ideal conditions including perfectly compensatory damages, a liability rule causes efficient self-monitoring by a rational actor.”).
131. See infra Part IX (discussing the perverse incentives created by the present patent damages regime).
for treble damages in the event of willful infringement;\textsuperscript{132} however, willfulness is difficult to prove in practice.\textsuperscript{133}

Applying the EMVR in lost-profits scenarios to award a patentee the entire market value of a product when the patent is the basis (but not the sole source) of demand for the product is restitutionary, and thus contrary to the present patent statute’s grant of a compensatory remedy.\textsuperscript{134} Applying the EMVR in reasonable-royalty scenarios to block a patentee from basing a royalty on the infringing product’s profits further favors an infringer in a system that already encourages, to some extent, efficient infringement. The EMVR has no place in patent law.

Most importantly, the copyright statute explicitly allows a copyright holder to introduce an infringer’s profits on a larger work into evidence without requiring proof that the infringed component constitutes the basis of demand.

\textbf{VII. An Attempt to Impart Clarity}

The EMVR has evolved to its present state because the decisionmaker in patent infringement determinations is often a jury of laypeople. As such, the rule is not entirely without purpose, for when a patented component contributes less than 1\% of the value of the device being sold, it becomes increasingly difficult for an infringer to prove, for instance, that the patent is worth 0.01\%, rather than only 0.001\%.\textsuperscript{135} But this problem is only

\begin{footnotesize}
\begin{itemize}
\item 134. There is no requirement under the EMVR (as opposed to the requirements expressed in the \textit{Panduit} factors) that there be proof that the patentee had the manufacturing capability to exploit the demand for the overall product, or even that the patentee contemplated the idea of marketing the overall product. See \textit{Panduit Corp. v. Stahlin Bros. Fibre Works, Inc.}, 575 F.2d 1152, 1156 (6th Cir. 1978).
\item 135. See LaserDynamics, Inc. v. Quanta Computer, Inc., 694 F.3d 51, 68 (Fed. Cir. 2012) (“Admission of such overall revenues, which have no demonstrated correlation to the value of the patented feature alone, only serve to make a patentee’s proffered damages amount appear modest by comparison, and to artificially inflate the jury’s damages calculation beyond that which is adequate...”)
\end{itemize}
\end{footnotesize}
a minor issue. Assume an infringer of a minor patent made $100,000,000 in profits on a larger device containing other unpatented components. The difficulty in proving that the infringed patent was worth either 0.01% or 0.001% means the difference between a reasonable royalty of $10,000 and one of $1,000. Considering the cost of litigation, such a difficulty in proof is of little consequence. Only when the profits are in the billions of dollars does the difficulty yield consequential differences, and in such a case, the trial judge must be more demanding in scrutinizing a damages expert’s methodology. Although it may be easier to pinpoint a value of a patent when it contributes a substantial portion of the value to the broader device, adoption of the EMVR does more harm than good because it excludes material evidence in most cases without exception.

Let us take Uniloc to its logical conclusion when omitting the EMVR and placing heavier weight on the evidence of profits. The gross revenues for the products at issue were approximately $19.28 billion. Microsoft’s latest profit margin, based on its 2011 annual report, was approximately 33%. Thus, we could assume that the profits on the products at issue in Uniloc were approximately $6.36 billion. Uniloc’s expert multiplied $2.50 by the number of licenses sold to get roughly $565 million; thus there were approximately 226 million licenses sold by Microsoft. The expert’s calculation would value a reasonable royalty (after adjusting for a hypothetical negotiation) at approximately 8% of the profit for Microsoft Office and Windows. To evaluate the reasonableness of this rate, it is necessary to provide some context.

to compensate for the infringement.” (internal quotation marks omitted) (citing Uniloc USA, Inc. v. Microsoft Corp., 632 F.3d 1292, 1320 (Fed. Cir. 2011)).

136. See Mark A. Lemley & Carl Shapiro, Patent Holdup and Royalty Stacking, 85 Tex. L. Rev. 1991, 2021 (2007) (providing an example of how a low percentage royalty based on the product price might seem reasonable but may easily exceed the royalty the parties would have negotiated, or even exceed the entire value of the infringing component).

137. Uniloc USA, Inc. v. Microsoft Corp., 632 F.3d 1292, 1312 (Fed. Cir. 2011).


139. Uniloc, 632 F.3d at 1311.
Digital rights management (DRM) refers to methods of preventing unauthorized use of a company’s software; the term encompasses product activation codes. Microsoft’s PlayReady, as one example to show the amount charged for DRM, is a content access and protection technology for digital entertainment content, products, services, and devices. The fee structure for PlayReady is based on a client’s quarterly unit volume, with the royalties per unit decreasing as unit volume increases. The most expensive royalty per unit is $0.20. Taking this conservative rate, and assuming it is in line with industry norms, Microsoft would have incurred a cost of about $45,200,000 for the 226 million licenses sold (assuming it was paying for a product comparable to its own). Yet the jury awarded $388 million in damages. Based on this rough estimate, the jury’s award was more than eight times the value that Microsoft charges for its own DRM software. Considering this large difference, it would not be a stretch for a judge to rule that no reasonable jury could find that the evidence supported a $388 million award. Notice how the reasonableness of an award can be attacked at the trial level without any reference, and therefore without any need, for the EMVR.

140. Pamela Samuelson & Jason Schultz, Should Copyright Owners Have to Give Notice of Their Use of Technical Protection Measures?, 6 J. TELECOMM. & HIGH TECH. L. 41, 42 (2007) (“Some copyright owners have adopted technical protection measures (or ‘TPMs’, sometimes also referred to as ‘digital rights management’ or ‘DRM’ technologies) to control unauthorized access to and uses of digital content in mass-market products and services.”).


142. Id. It appears that Microsoft has taken the fee structure cited herein off of its website. See www.diigo.com/item/image/8c52/5wt6?size=0 for a clipped version.

143. Fee Structure for Microsoft PlayReady Final Products, http://www.microsoft.com/playready/PPPLFees.mspx (last visited Feb. 4, 2012). Check the website mentioned in note 142 if this site is no longer available.

144. This figure does not apply the tiered royalties per unit that the fee structure would in actuality. Id. This is conservative because it applies the highest rate to all licenses.

145. Uniloc USA, Inc. v. Microsoft Corp., 632 F.3d 1292, 1311 (Fed. Cir. 2011).

146. See FED. R. CIV. P. 50 (stating the standard for granting a judgment as a matter of law).
Admittedly, the above analysis is not entirely accurate considering that the patents covering PlayReady (assuming they exist) were not at issue in *Uniloc*. But the analysis can be supported as relevant because the marginal production cost of software is typically near zero. This low marginal cost would apply to both the product activation code at issue in *Uniloc* as well as the code for Microsoft PlayReady. Thus, it is reasonable to assume the costs per unit and the sale price of the two DRM programs, when dealing with economies of scale, would be comparable.

Perhaps contributing to excessive, misinformed jury awards are the sometimes confusing jury instructions. Looking at a standard jury instruction, the contradictions in logic would confuse a lawyer, let alone a layperson: “A reasonable royalty is the royalty that would be reasonable for the infringer to pay and for the patent owner to accept for use of a patent that they both know is valid and that the infringer wants to use.” The instruction then provides: “You are to decide what a reasonable royalty would be based on circumstances as of the time just before [defendant] began [selling or using] the infringing [product or process].” Taking this sentence literally, the actual profits of the product, including the infringed patent, would seemingly be precluded from consideration. But the jury instruction then provides a list of factors the jury should consider, including “the portion of the profit that is due to the patented invention, as compared to the portion of the profit due to other factors, such as unpatented elements or unpatented manufacturing processes, or features or improvements developed by [defendant].” The instruction, like the case law, contradicts itself. It asks the jury to base its royalty as of the time of initial infringement but, at the same time, to consider the actual profits attributed to the patent that necessarily incur after this point in time. It would be wise to eliminate the contradiction and tell the jury that when an infringer cannot prove lost profits, a reasonable royalty informed

149. *Id.*
150. *Id.* at 198.
by the infringer’s profits (amongst other factors) should be determined.\footnote{This assumes that the patentee has provided sufficient evidence to prove the value that the patent contributed to the product.}

The primary forces driving royalty rates for intellectual property are the amount of profits, the duration of profits, and the risk associated with the expected profits.\footnote{Russell Parr, Royalty Rates for Licensing Intellectual Property 124 (2007).} In licensing, “[t]he basis of all value is cash.”\footnote{Id. at 143.} And to determine net cash flow, one must take into account gross revenues.\footnote{See id. at 144 (providing a basic net cash flow calculation).}

In light of the above analysis, this Note proposes a change to the test that courts apply when determining whether to exclude total product revenue (en route to finding profits) from evidence: A patentee may recover damages associated with the value of an entire apparatus containing several features when the feature patented constitutes some basis for customer demand and there is a sufficient functional relatedness between the patented feature and the product.\footnote{This language (other than the emphasized change) was taken from TWM Mfg. Co. v. Dura Corp., 789 F.2d 895, 901 (Fed. Cir. 1986) (citations omitted). Courts have impliedly rejected this change. For example, in Schindler Elevator Corp. v. Otis Elevator Co., 561 F. Supp. 2d 352 (S.D.N.Y. 2008), the court granted a motion in limine to exclude the testimony of plaintiff’s damages expert because it relied on proof that the infringed patent was a substantial basis for consumer demand rather than the higher standard requiring that it was the basis for consumer demand. See Schindler Elevator Corp. v. Otis Elevator Co., 1-06-cv-05377 (S.D.N.Y. June 23, 2011) (granting motion in limine).} This change is suggested in the spirit of eventually abrogating a rigid test entirely in moving toward judges’ increased scrutiny of experts under \textit{Daubert} and their increased willingness to grant judgments notwithstanding a jury verdict on damages. The United States District Court for the District of Delaware has, at least in one case, interpreted the rule in a way that most aligns with the spirit behind this change. In \textit{Lucent Technologies, Inc. v. Newbridge Networks Corp.},\footnote{Lucent Techs., Inc. v. Newbridge Networks Corp., 168 F. Supp. 2d 181 (D. Del. 2001).} the district court stated: “[T]he test for the application of the entire market value rule is not whether the unpatented products are
VIII. Support from the Federal Rules of Evidence

Presumably because the EMVR originated before the current Federal Rules of Evidence were enacted, there are no cases discussing how the EMVR comports with these rules of evidence. Originally, it did not operate as an evidence rule at all. As applied today, does the EMVR make sense in light of Federal Rules of Evidence 401, 402, and 403? In general, relevant evidence is admissible unless the United States Constitution, a federal statute, the Federal Rules of Evidence, or rules prescribed by the Supreme Court provide otherwise. The EMVR did originate in the U.S. Supreme Court, and thus it is arguably a rule prescribed by it. But when the Supreme Court has invoked the EMVR, it has never been in association with reasonable royalties—it has been in association with restitutionary remedies. Thus, the Supreme Court exception to relevancy should not apply to the EMVR as perverted.

Is the amount of profits made on a complex device incorporating a patented subcomponent relevant in determining a reasonable royalty if the patent is infringed? This question turns on whether the evidence “has any tendency to make a fact [here, what a reasonable royalty would be] more or less probable than it would be without the evidence; and [whether] the fact is of consequence in determining the action.” The argument opposing relevancy would be that the actual profits on a device

157. Id. at 238.
160. See Garretson v. Clark, 111 U.S. 120, 121 (1884) (applying the EMVR).
161. The EMVR was imposed in reasonable royalty determinations by the United States Court of Claims. See Marconi Wireless Tel. Co. v. United States, 99 Ct. Cl. 1 (1942), aff’d in part and vacated in part on other grounds by 320 U.S. 1 (1943).
incorporating the patent are not relevant because the valuation method relies on a hypothetical negotiation before the product was manufactured and sold; thus, the actual value would not have yet come into existence for consideration. Further, the irrelevancy argument may point to the possibility that without proof of some value to the patent, the profits of the device in which it is incorporated mean nothing in ascertaining the value of the patent to any precision.

Taking the first argument, Georgia-Pacific has fleshed out the elements indicative of a reasonable royalty, and among these elements is the “established profitability of the product made under the patent.” The hypothetical negotiation was not meant to totally preclude consideration of events occurring after the initial infringement. Rather, it is to be determined in context. The second irrelevancy argument is stronger, and certainly the patentee should prove some benefit to the overall product before being able to introduce profits of the product incorporating the infringed patent among other patented or unpatented components. But if the infringed patent had no value whatsoever, why would the infringer find it necessary to include it in its device? Once the patent is found infringed, unless there are no profits on the broader device, the existence of some value attributed to the patent might be presumed. Whether that value is ten cents or ten million dollars is for the jury to decide. The evidence of profits is relevant to this determination.

Even so, a “court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.” Thus, if the market value is relevant under Federal Rules of Evidence 401 and 402, which it likely is, then do any of the above considerations substantially outweigh the probative value of the actual profit of the broader device on the proper amount of a reasonable royalty? If there is a

163. See Panduit Corp. v. Stahlin Bros. Fibre Works, Inc., 575 F.2d 1152, 1164 (6th Cir. 1978) (“The infringer’s profit element, in the post-judgment ‘reasonable royalty’ equation, is not related to the infringer’s actual profit.”).
165. FED. R. EVID. 403.
functional relatedness between the infringed patent and the overall product, and proof of this relatedness to some degree of consumer demand, the answer is “not always.”

Presuming the existence of these elements, there is no persuasive argument that there is unfair prejudice against the infringer. The argument for unfair prejudice might be that if a jury sees a large profit figure it will be more inclined to think the infringer can afford a large award even if the award does not reflect the true value of the infringed patent. Based on that inclination, the jury will unfairly value the patent in the patentee’s favor. Perhaps it would be unfair if the parties only introduced net profits into evidence and sent the jury to deliberate. But the parties employ experts to put that figure into context. The infringer has every opportunity to explain how the device operates, why it is attractive to consumers, and how the other elements contribute to the product’s value. It is not unfair to require an infringer to explain its product.

Nor does the amount of profits tend to mislead the jury. Expert valuations will incorporate expected profits and actual profits. As stated earlier, the basis of value is cash flow. Thus, to determine the value attributable to the infringed patent, the necessary cash value is the profit made. When the entire device is not infringing, the profit attributable to the infringed component must be apportioned from the total profit made en route to determining a reasonable royalty. In summary, the Federal Rules of Evidence do not support a categorical bar to admitting an infringer’s profits into evidence in reasonable royalty determinations (even when the patent isn’t the basis for demand),

---

166. See Rite-Hite Corp. v. Kelley Co., 56 F.3d 1538, 1550 (Fed. Cir. 1995) (requiring a functional relationship between the infringed patent and the product to allow recovery based on unpatented components sold with patented components).

167. See Lemley & Shapiro, supra note 136, at 2024 (stating that in practice, it is rare for an infringer to introduce evidence of other patented components). The Article explains that such evidence is rare because an infringer usually will “not want to admit that it might be infringing other patented inventions.” Id. Further contributing to the dearth of this type of evidence is courts’ reluctance to admit evidence of patents other than that at issue in the trial. See id. (“[C]ourts do not want to admit such evidence because it will require collateral litigations during the damages phase over the existence and value of parts of the product that are not covered by the patent at trial.”).

168. See Parr, supra note 152, at 143.
provided there is a functional relatedness between the infringed patent and the broader product and that some value to the patent can be proven or inferred.\footnote{169} Of course, this determination is inherently fact-specific.

Courts have stated that without the EMVR requirement, “an infringer could be required to pay multiple recoveries on a single product to numerous patentees, each of whom file infringement claims directed to different components of the product without regard to the extent to which its patented component contributed to the overall profitability of the product.”\footnote{170} But this logic does not extend to calculating a reasonable royalty based on profits because, in attempting to formulate a reasonable royalty, the precise goal is to determine the extent to which the patent contributed to overall profitability.

Some decisions have explicitly recognized the relevancy of the infringer’s profits in assessing the reasonableness of a royalty award. The United States District Court for the District of Delaware has stated that “[w]here the commercial value of a product is increased by the use of the patented process and infringement is present, profits are considered a relative element.”\footnote{171}

IX. Harmonizing Patent Damages with the Goals of Patent Law

Congress’s power to grant limited monopolies in the form of patents is to be used “[t]o promote the Progress of Science and the useful Arts.”\footnote{172} The right to exclude others from making, using, or

\footnotetext{169}{See Roberts, \textit{supra} note 58, at 670 ("Measuring the infringer’s profits may sometimes provide a useful proxy for assessing the patent owner’s compensatory loss."). Certain circumstances may justify excluding profits from evidence.}


\footnotetext{171}{Hartford Nat’l Bank & Trust Co. v. E.F. Drew & Co., 188 F. Supp. 353, 358–59 (D. Del. 1960); see also Locklin v. Switzer Bros., Inc., 235 F. Supp. 904, 906 (N.D. Cal. 1964) (stating that “an infringer’s profits may have an evidentiary bearing on the determination of the reasonableness of a royalty”).}

\footnotetext{172}{U.S. \textit{CONST.} art. I, § 8, cl. 8.}
selling a patented invention provides encouragement for taking investment-based risk in developing new technologies.\textsuperscript{173}

If the entire market value is excluded in cases in which the patent does not constitute the basis of demand for the product, proving a reasonable royalty becomes significantly more difficult—especially if a court decides to exclude evidence of other patents incorporated in the device. Intentional infringers know this, so the EMVR as applied produces a moral hazard for an infringer to aggressively market the infringing product in hopes of eliminating any chance for the patentee to establish a record of sales that could support an award of lost profits, thus relegating a patentee to the reasonable royalty remedy. The infringer’s aggressive marketing would also reduce the likelihood that a patentee could successfully negotiate a license for the patent with a potential competitor to the infringer due to the infringer’s head start on manufacturing and advertising and the consequent reluctance of a competitor to enter a saturated market. This would tend to prevent an actual royalty rate from being established. Thus the patentee is put in the position of trying to prove a reasonable royalty without evidence of established rates and, under the EMVR, without reference to the infringer’s profits made on a device incorporating the patent. An infringer has every incentive to put the patentee in this position because it substantially impairs the patentee from rebutting a low-ball royalty suggestion from the infringer’s expert.

The goals of promoting technological advancement are not furthered when inventors cannot get adequate compensation for a violation of their right to exclude. Entities investing in innovation theoretically take the risk of being inadequately compensated for infringement of a potential patent into account—this risk arguably stymies investment to a certain extent. In encouraging investment in the technical arts, it is better to overcompensate patentees in rare instances rather than undercompensate them in most instances of infringement. This idea is supported by the patent statute’s inference that a reasonable royalty is the floor to a potential damages award.\textsuperscript{174}

\textsuperscript{173} King Instruments Corp. v. Perego, 65 F.3d 941, 950 (Fed. Cir. 1995).

\textsuperscript{174} See 35 U.S.C. § 284 (2006) (stating that in no event is the owner of an infringed patent to be awarded less than a reasonable royalty).
X. Conclusion

Some scholars believe that the EMVR has no place in reasonable royalty cases.\textsuperscript{175} Professor Mark Lemley has said that applying the EMVR “necessarily overcompensates the patent owner by giving it value not in fact attributable to the patent.”\textsuperscript{176} This Note agrees with the inapplicability of the rule, but not with Professor Lemley’s reasoning. The rule has no place in a reasonable royalty calculation not because the market value of a product should always be excluded from evidence when the patent at issue covers only a portion of the product, but because such evidence should almost always be included. In most cases, “market value” should more narrowly refer to net profits as long as the infringer has kept adequate records with good-faith accounting for revenues and expenses (and was not unreasonably inefficient). A patentee is not necessarily overcompensated when the royalty base includes the entire product, but it is overcompensated if a judge fails to properly scrutinize damages experts’ methodologies in comparing a royalty rate to net profits or if a judge fails to grant a judgment as a matter of law in the case of an unreasonable award. When an infringer seeks judgment as a matter of law on a reasonable-royalty award that is not reasonably linked to the profits attributed to the patent, a

\textsuperscript{175} See, e.g., Mark A. Lemley, Distinguishing Lost Profits from Reasonable Royalties, 51 WM. & MARY L. REV. 655, 663–64 (2009) (“The logic of the entire market value rule breaks down in reasonable royalty cases . . . .”). “Until courts abandon current doctrine and apply the entire market value rule only when the patented component of the accused devices truly accounts for the entire market demand for the infringed device, patentees will continue to be unjustly rewarded.” Brian J. Love, Patentee Overcompensation and the Entire Market Value Rule, 60 STAN. L. REV. 263, 293 (2007). For a sure-to-be rewarding trek into mathematics, the reader ought to consider that the mathematical proof in Love’s Note might assume its own conclusion. See id. at 273–75 (concluding that when a reasonable royalty is based on the entire value of a device, the benchmark royalty will only be achieved if the infringing component accounts for the entire value of the accused product). That is, $V_{EMVR}$ was assumed to be $V_{product}$ in reaching this conclusion. Id. at 275. Should not $V_{product}$ have been included in the summation of individual patents such that the only conclusion to draw would be that $B^*V_{product} = B^*V_{EMVR} - B^*\sum_{i=2}^{N} v_i$ (assuming that $v_{product} = v_1$ in the summation)? See id.

\textsuperscript{176} Lemley, supra note 175, at 664.
trial judge should find that no reasonable person would find the award supported by sufficient evidence.\textsuperscript{177}

In limiting any unfair prejudice, a court should only admit into evidence the profits made on a product that are functionally tied to the patented component when the patent is shown to have provided some value; the patentee should then prove with sufficient evidence the profit of the entire product that is attributable to the patent.\textsuperscript{178} The limitation on a patentee’s purported royalty base should be the functional nexus plus “some value” standard, not a “basis of consumer demand” or “substantially creates demand” standard. It may be possible to infer value based on positive net profits. Once some value for the patent is established by the patentee, the defendant has the opportunity to rebut by attempting to highlight the value attributed to the other features of the machine. Once these elements are established, the valuation for reasonable royalty purposes should typically be in reference to the profits, not the revenues. The inherent uncertainties in valuation are thereby limited. Furthermore, this reference to profits of the entire device compensates the patentee for the synergistic value of the patent, rather than the value of the patent in a vacuum. It is fair to assume that an infringer/hypothetical licensee would take this into account when negotiating a license even if the patent is not expected to cover a component that is the basis of demand.

In summary, courts have taken a restitutionary sword from plaintiffs and reforged it into a shield for defendants in a

\textsuperscript{177} See Dow Chem. Co. v. Nova Chems. Corp., No. 05-737-JJF, 2010 WL 3056617, at *1 (D. Del. July 30, 2010) (“[A] court may grant judgment as a matter of law if ‘the court finds that a reasonable jury would not have a legally sufficient evidentiary basis’ to find for a party on a given issue after that party has been fully heard.” (citing Fed. R. Civ. P. 50(a))).

\textsuperscript{178} See Rite-Hite Corp. v. Kelley Co., 56 F.3d 1538, 1550 (Fed. Cir. 1995) (“[W]hen recovery is sought on sales of unpatented components sold with patented components . . . the unpatented components must function together with the patented component in some manner so as to produce a desired end product or result.”). Thus, this Note’s proposed revision to reasonable royalties in patent damages law would maintain the functional nexus requirement while dropping the “basis for consumer demand” requirement. Ultimately, upon review, the findings at trial will be judged on a substantial evidence standard. See Lucent Techs., Inc. v. Gateway, Inc., 580 F.3d 1301, 1333 (Fed. Cir. 2009) (stating that on appellate review, findings of fact are reviewed on a substantial evidential support standard).
compensatory system. The result will not be more reasonable
awards, but rather awards that are more likely to be arbitrary
and a system that inadequately deters infringement.
The Physical Consequences of Emotional Distress: The Need for a New Test to Determine What Amounts Are Excluded from Gross Income Under § 104(a)(2)

C. Anthony Wolfe IV*

Table of Contents

I. Introduction ................................................................... 2274

II. Development of the “Personal Injury” Exclusion ........ 2276
    A. Early Development ................................................. 2276
    B. The “Personal Injury” Exclusion After
        Glenshaw Glass ........................................................... 2280
    C. The 1989 Amendments ............................................. 2282
    D. A Two-Part Test Emerges ...................................... 2283
        1. United States v. Burke .......................................... 2284
        2. Commissioner v. Schleier .................................... 2285

III. The Courts Refine § 104(a)(2)’s “On Account Of” Requirement ................................................................. 2289
    A. The “Direct Causal Link” ........................................... 2290

IV. The Small Business Job Protection Act of 1996 ............. 2293

V. Section 104(a)(2)’s Exclusion After the 1996 Act .......... 2296
    A. The Circuits Reach the Right Outcome ................... 2296

VI. The Tax Court Reaches Conflicting Results ................ 2304
    A. Sanford v. Commissioner ....................................... 2304
    B. Domeny v. Commissioner ...................................... 2307

* Candidate for J.D., Washington and Lee University School of Law, May 2013. I would like to thank Professor Brant J. Hellwig for pointing me toward this topic and providing invaluable advice throughout writing this Note. Also, thank you to the members of the Washington and Lee Law Review for all their assistance.
I. Introduction

Congress amended § 104(a)(2) of the Internal Revenue Code in 1996, inserting the word “physical” into § 104(a)(2)’s “personal injury or sickness” exclusion from gross income. Section 104(a)(2)’s gross income exclusion now applies to only damages received “on account of personal physical injuries or physical sickness.” Congress’s amendment further provided that “emotional distress shall not be treated as a physical injury or physical sickness.” Before Congress’s change, the exclusion from gross income applied to all “personal injuries,” regardless of whether they were physical, nonphysical, or emotional.

Since Congress amended § 104(a)(2), courts have struggled to uniformly apply the exclusion and have begun to extend it beyond only physical injuries and physical sicknesses. Part of this struggle is due to the inherent difficulty with defining the word “physical.” Significantly, however, the struggle is mostly to do with the fact that the judicial test used to determine what amounts fall within § 104(a)(2)’s exclusion developed when the

2. See infra Part IV (discussing the Small Business Job Protection Act of 1996).
4. I.R.C. § 104(a) (flush language).
5. See United States v. Burke, 504 U.S. 229, 235 n.6 (1992) (noting that the exclusion’s scope had previously been limited to only physical injuries, but that it was now settled that it extended to “nonphysical injuries to the individual”).
6. See infra Part VI (discussing the conflicting Tax Court cases).
exclusion applied to all “personal injuries.”\textsuperscript{8} Although Congress limited § 104(a)(2)’s exclusion to only physical injuries and physical sicknesses, courts have not adjusted the test used to determine what amounts fall within the exclusion.\textsuperscript{9} This Note proposes that courts modify the test used to exclude damage awards under § 104(a)(2) to accommodate the added “physical” requirement.\textsuperscript{10} The proposed test will allow § 104(a)(2) to be uniformly applied to only physical injuries and physical sicknesses in the way that Congress intended.

Every year taxpayers must calculate their gross income to determine the tax they must pay. To do this, taxpayers must know what amounts are included in gross income under § 61\textsuperscript{11} and what amounts can be excluded from gross income under other Tax Code provisions. In 2011, disputes over what amounts are included in gross income and what amounts can be excluded were among the most litigated taxation issues.\textsuperscript{12} Personal injury claimants—physical and nonphysical—are awarded millions of dollars each year.\textsuperscript{13} Thus, § 104(a)(2)’s gross income exclusion is consistently one of the most litigated exclusionary provisions.\textsuperscript{14}

\textsuperscript{8} See Comm’r v. Schleier, 515 U.S. 323, 336–37 (1995) (determining that for a recovery to be excluded under § 104(a)(2), a taxpayer must establish that the “cause of action giving rise to the recovery [was] ‘based upon tort or tort type rights’” and that the “damages were received ‘on account of personal injuries or sickness’”).

\textsuperscript{9} See Amos v. Comm’r, T.C. Memo. 2003-329, at 4 (2003) (“The 1996 amendment does not otherwise change the requirements of section 104(a)(2) or the analysis set forth in Commissioner v. Schleier; it imposes an additional requirement for an amount to qualify for exclusion from gross income under that section.” (citation omitted)).

\textsuperscript{10} See infra Part VII (proposing a new test).


\textsuperscript{13} See Genny Barret, Note, Did the Sixth Circuit Get It Right in Stadnyk? What To Do About the § 104(a)(2) Personal Injury Damages Exclusion, 2011 BYU L. Rev. 1193, 1193 (2011) (“Each year, millions of dollars are awarded to victims of physical and non-physical personal injury.”).

\textsuperscript{14} See Nat’l Taxpayer Advocate, 2008 Annual Report to Congress 472 (2008), http://www.irs.gov/pub/irs-utl/08_tas_arc_mli.pdf (“Taxation of damage awards spurs litigation every year.”); see also Wright, supra note 7, at 211 (noting that § 104(a)(2)’s exclusion is one of the most litigated issues in federal courts).
Part II of this Note outlines the history of the “personal injury” exclusion before Congress added the word “physical.” Part III discusses judicial opinions limiting the scope of § 104(a)(2)’s “on account of” language. Congress’s amendment to § 104(a)(2) is detailed in Part IV, and the circuit courts’ response to the amendment is discussed in Part V. Part VI of this Note analyzes three Tax Court cases that illustrate the struggle in applying the amended § 104(a)(2) exclusion. To remedy the confusion surrounding § 104(a)(2)’s exclusion since its amendment, this Note proposes a new judicial test in Part VII. Finally, this Note concludes in Part VIII.

II. Development of the “Personal Injury” Exclusion

A. Early Development

Before the first personal injury exclusion appeared in Chapter 18, § 213(b)(6) of the Revenue Act of 1918, the U.S. Department of the Treasury (Treasury) indicated, in Treasury Decision 2135, that amounts received for personal injury were taxable as “gains or profits and income derived from any source whatsoever.” Treasury Regulation 33, which stated, “An amount received as the result of a suit or compromise for personal injury, being similar to the proceeds of accident insurance, is to be

15. See infra Part II (discussing the personal injury exclusion’s history).
16. See infra Part III (discussing judicial opinions interpreting “on account of”).
17. See infra Part IV (discussing the 1996 amendments).
18. See infra Part V (discussing circuit court opinions following the 1996 amendments).
19. See infra Part VI (discussing three conflicting Tax Court opinions).
20. See infra Part VII (proposing a new test).
21. See infra Part VIII (conclusion).
23. T.D. 2135, 17 Treas. Dec. Int. Rev. 39, 42 (1915) (“An amount received as a result of suit or compromise for ‘pain and suffering’ is held to be such income as would be taxable under the provision of law that includes ‘gains or profits and income derived from any source whatsoever.’”); see also Patrick E. Hobbs, The Personal Injury Exclusion: Congress Gets Physical But Leaves the Exclusion Emotionally Distressed, 76 Neb. L. Rev. 51, 56–57 (1997) (noting Treasury’s early treatment of personal injury awards).
accounted for as income,” later confirmed Treasury’s indication.\textsuperscript{24} 

The tax treatment of personal injury awards changed in 1918, however, when the Attorney General urged Treasury to find accident insurance proceeds nontaxable.\textsuperscript{25} 

The Attorney General’s opinion rested on the idea that accident insurance proceeds represented a return of capital.\textsuperscript{26} With tangible assets such as real property, the taxpayer’s invested capital is measured by the amount the taxpayer paid to acquire the property.\textsuperscript{27} This acquisition “cost” can be subsequently adjusted.\textsuperscript{28} This cost is the taxpayer’s basis in such property.\textsuperscript{29} Later, when a taxpayer disposes of the property, she is only taxed on the amount in excess of her basis, or gain.\textsuperscript{30} The amount she receives equaling her basis in the property is her “return of capital” and is not taxed.\textsuperscript{31} 

Applied to personal injury damages, the return of capital approach excludes from income damages equaling the taxpayer’s capital invested in oneself.\textsuperscript{32} This “human capital” theory stands for the principle that accident proceeds and damage awards

\textsuperscript{24} Treas. Reg. § 33, art. 4 (1918); see also Hobbs, supra note 23, at 57 (stating that Regulation 33 confirmed that Treasury regarded personal injury awards as taxable).

\textsuperscript{25} See 31 Op. Att’y Gen. 304 (1918) (urging Treasury to find accident insurance proceeds excludable from income); see also Hobbs, supra note 23, at 57 (noting that the tax treatment of personal injury awards reversed in 1918 when the Attorney General urged Treasury to find accident insurance proceeds excludable from income).

\textsuperscript{26} See 31 Op. Att’y Gen. at 308 (arguing that accident proceeds “merely take the place of capital in human ability which was destroyed by the accident”); see also Hobbs, supra note 23, at 63 (discussing the Attorney General’s extension of “the notion of capital value replacement to the human body”).

\textsuperscript{27} See I.R.C. § 1012 (2006) (defining basis in one’s property as “the cost of such property”).

\textsuperscript{28} See id. § 1016 (adjusting basis “for expenditures, receipts, losses, or other items, properly chargeable to [a] capital account”).

\textsuperscript{29} See id. § 1012 (defining basis as “the cost of such property”).

\textsuperscript{30} See id. § 1001 (providing that a taxpayer’s gain on such property is the difference between the “amount realized” and the basis in the property).


\textsuperscript{32} Id.
should not be taxed because the proceeds are merely a substitute for capital invested in oneself that was diminished in the accident.\textsuperscript{33} Because the proceeds represent a “return of capital” equal to the taxpayer’s invested capital in herself, the proceeds do not increase the taxpayer’s wealth.\textsuperscript{34} Instead, the accident proceeds simply return the taxpayer to her pre-injury condition.\textsuperscript{35} This approach is difficult, however, because invested “human capital” in terms of cost to the owner—the owner’s basis—is often zero.\textsuperscript{36} Moreover, if cost could theoretically be established, it is nearly impossible for a court to allocate accident proceeds to an abstract notion of “human capital.”\textsuperscript{37}

Nonetheless, Treasury accepted the Attorney General’s rationale when it issued Treasury Decision 2747.\textsuperscript{38} Treasury Decision 2747 excluded personal injury awards from income, stating that “an amount received by an individual as a result of a suit or compromise for personal injuries sustained by him through accident is not income.”\textsuperscript{39} Congress agreed with Treasury, and promptly codified the exclusion when it passed the Revenue Act of 1918.\textsuperscript{40} At the time, Congress believed personal

\begin{flushright}
33. See Hobbs, supra note 23, at 63 (“This human capital argument rests on the principle that if an ability or function of the human body, which previously allowed the taxpayer to produce income, is diminished through an accident, then the proceeds are merely a substitute for that ability or function.”).

34. See Cohen-Whelan, supra note 31, at 919–20 (stating that a return of capital does not increase wealth).

35. See id. at 919 (explaining that damages would “return the taxpayer to her pre-injury condition”).

36. See Hobbs, supra note 23, at 64 (explaining that often “[t]he cost to the owner of human capital is zero”); see also Roemer v. Comm’r, 716 F.2d 693, 696 n.2 (9th Cir. 1983) (“[T]here is no tax basis in a person’s health and other personal interests . . . .”).


40. See Revenue Act of 1918, ch. 18, § 213(b)(6), 40 Stat. 1066 (“Amounts received, through accident or health insurance or under workmen’s compensation acts, as compensation for personal injuries or sickness, plus the
injury compensation payments to be a return of human capital, and thus not constitutionally taxable “income” under the Sixteenth Amendment.41

Although the statute used the language “personal injuries or sickness,”42 Treasury rejected extending the exclusion to the alienation of a wife’s affections, interpreting the exclusion to apply to “physical injuries only.”43 Later, however, in the wake of the now-overruled U.S. Supreme Court decision, Eisner v. Macomber,44 in which the Court defined income narrowly to include “gain derived from capital, from labor, or from both combined,”45 Treasury created an administrative exclusion for nonphysical personal injuries.46 Treasury’s administrative exclusion became a judicial exclusion in 1927 when the Board of Tax Appeals—now the Tax Court—relied on Eisner’s narrow definition of income to determine that a settlement award in a defamation suit was not taxable “income.”47 Twenty-eight years later, the Supreme Court drastically broadened its definition of income in Commissioner v. Glenshaw Glass, Co.48 to include any

41. See Dotson v. United States, 87 F.3d 682, 685 (5th Cir. 1996) (“Congress first enacted the personal injury compensation exclusion . . . when such payments were considered the return of human capital, and thus not constitutionally taxable ‘income’ under the 16th Amendment.”).
42. Revenue Act of 1918 § 213(b)(6).
43. S. 1384, 2 C.B. 71 (1920); see Hobbs, supra note 23, at 66 (discussing Senate Bill 1384 and that Treasury “rejected the extension of the ‘human capital’ approach to every personal injury”).
44. Eisner v. Macomber, 252 U.S. 189, 219 (1920) (holding that a “stock dividend” was not income).
45. Id. at 207.
46. See Sol. Op. 132, 1 C.B. 92 (1922) (determining that damages received for alienation of affections, defamation, and for the surrender of child custody rights were excludable from income); see also Hobbs, supra note 23, at 67 (discussing Treasury’s conclusion in Solicitor’s Opinion 132 that the receipts at issue were excludable from gross income based upon the Supreme Court’s decision in Eisner).
47. See Hawkins v. Comm’r, 6 B.T.A. 1023, 1024 (1927) (relying on Eisner to conclude that a taxpayer’s settlement award in a defamation suit was excludable from income); see also Hobbs, supra note 23, at 67–68 (discussing the court’s analysis and conclusion in Hawkins).
48. Comm’r v. Glenshaw Glass, Co., 348 U.S. 426 (1955) (holding that punitive damages fell within I.R.C. § 22(a)’s definition of income because they were “instances of undeniable accession to wealth, clearly realized, and over
“accession to wealth, clearly realized, and over which the taxpayers have complete dominion.” 49 In Glenshaw, the Court observed that Congress intended “to tax all gains except those specifically exempted.” 50 Despite this broad definition, the personal injury exclusion—now § 104(a)(2)—continued to be applied to nonphysical personal injuries. 51

B. The “Personal Injury” Exclusion After Glenshaw Glass

In Glenshaw Glass, the Court reiterated that “personal injury recoveries [were] nontaxable on the theory that they roughly correspond to a return of capital.” 52 The Court determined, however, that the same restoration-of-capital theory could not justify excluding punitive damages from income. 53

Subsequent to Glenshaw Glass, the Internal Revenue Service (the Service) issued a series of revenue rulings reaffirming that nonphysical personal injury receipts were excludable from income. 54 And in 1972, in Seay v. Commissioner, 55 the Tax Court

which the taxpayers have complete dominion”).

49. Id. at 432.
50. Id. at 430.
51. See infra Part II.B (discussing the exclusion after Glenshaw Glass).
52. Glenshaw Glass, 348 U.S. at 432 n.8.
53. Id.
54. See Rev. Rul. 55-132, 1955-1 C.B. 213 (holding that former World War II prisoners-of-war’s receipts for losses of personal rights were not includable in gross income); Rev. Rul. 56-462, 1956-2 C.B. 20 (determining that payments made to former Korean War captives “as compensation for the loss of their personal rights . . . [were] not includible in the gross income”); Rev. Rul. 56-518, 1956-2 C.B. 25 (“[C]ompensation paid by . . . Germany to citizens or residents of the United States . . . on account of . . . damage to life, body, health, liberty, or to professional or economic advancement, are in the nature of reimbursement for the deprivation of civil or personal rights and do not constitute taxable income . . . .”); Rev. Rul. 58-370, 1958-2 C.B. 14 (determining that, for the same reasons as in Revenue Ruling 56-518, compensation paid by the Federal Republic of Austria did not constitute taxable income); see also Hobbs, supra note 23, at 70 (discussing the Service’s rulings and their renewal of the position that nonphysical personal injury receipts were excludable).
55. Seay v. Comm’r, 58 T.C. 32, 37 (1972) (holding that § 104(a)(2) excluded the taxpayer’s settlement award for personal injury claims); see Hobbs, supra note 23, at 71–72 (discussing the Tax Court decision in Seay and noting that the decision was the first time a court addressed the “application of the statutory exclusion to nonphysical injuries”).
similarly reaffirmed that the exclusion extended to nonphysical injuries.\footnote{56}{See \textit{Seay}, 58 T.C. at 38 (finding that § 104(a)(2) extended to the taxpayer’s nonphysical personal injuries).} In \textit{Seay}, the Tax Court concluded that a $45,000 payment as “compensation for . . . personal embarrassment, mental and physical strain and injury to health and personal reputation” was excludable from income under § 104(a)(2).\footnote{57}{Id.} The court cited Revenue Ruling 58-418, and its earlier decision in \textit{Hawkins v. Commissioner},\footnote{58}{Hawkins v. Comm’r, 6 B.T.A. 1023 (1927) (holding that damages received for libel and slander were excludable from income).} in support of its conclusion that the payment at issue fell within § 104(a)(2)’s exclusion.\footnote{59}{See \textit{Seay}, 58 T.C. at 40 (stating that \textit{Hawkins} and Revenue Ruling 54-518 supported the finding that the payment at issue was exempt from taxation under § 104(a)(2)).} \textit{Hawkins} had analogized recovery for injury to personal reputation with compensation by way of life insurance proceeds, and found damages received for libel and slander to be excluded from income.\footnote{60}{See \textit{Hawkins}, 6 B.T.A. at 1024–25 (analogizing a recovery for injury to personal reputation with compensation by way of life insurance proceeds, and finding the damages at issue to be excludable from income).}

Eleven years later, the Ninth Circuit seemingly erased any doubt as to whether § 104(a)(2)’s personal injury exclusion applied to nonphysical personal injuries in \textit{Roemer v. Commissioner}.\footnote{61}{See \textit{Roemer v. Comm’r}, 716 F.2d 693, 697 (9th Cir. 1983) (holding that “the ordinary meaning of a personal injury is not limited to a physical one”).} In \textit{Roemer}, the Ninth Circuit reversed the Tax Court’s decision below, concluding that an entire jury award in a defamation suit was excludable from income under § 104(a)(2).\footnote{62}{See id. at 700–01 (concluding that the award was excludable from gross income under § 104(a)(2)).} But shortly thereafter, the Service issued Revenue Ruling 85-143,\footnote{63}{Rev. Rul. 85-143, 1985-2 C.B. 55.} stating that it would follow the Tax Court’s conclusion including the defamation award in income, rather than the Ninth Circuit’s reversal.\footnote{64}{See id. (stating that the Service would not follow the Ninth Circuit’s decision in \textit{Roemer} and would instead follow the Tax Court’s decision including the defamation award in income).} Finally, however, the Tax Court ended the
debate in *Threlkeld v. Commissioner*\(^{65}\) when it determined that “personal injury” was not limited to a “physical injury.”\(^{66}\) Congress, though, would have its voice heard on the matter before the end of the decade.\(^{57}\)

C. The 1989 Amendments

In 1989, the House of Representatives proposed amending § 104(a)(2) to read, “[G]ross income does not include . . . the amount of any damages received . . . on account of personal injuries or sickness in a case involving *physical* injury or *physical* sickness.”\(^{68}\) According to the accompanying committee report, the House was attempting to override judicial applications of § 104(a)(2)’s exclusion to nonphysical personal injuries.\(^{69}\) Congress rejected the proposed amendment, however.\(^{70}\) Instead, the final amendment read, “[Section 104(a)(2)] shall not apply to any punitive damages in connection with a case not involving physical injury or physical sickness.”\(^{71}\) The final amendment “impliedly extended § 104(a)(2) to compensatory awards for nonphysical injuries,”\(^{72}\) a result contrary to the House’s proposed

\(^{65}\) Threlkeld v. Comm’r, 87 T.C. 1294, 1305 (1986) (stating that § 104(a)(2) extended beyond physical injuries).

\(^{66}\) See id. (“A personal injury has long been understood to include nonphysical as well as physical injuries.”).

\(^{67}\) See infra Part II.C (discussing the 1989 amendments to § 104(a)(2)).


\(^{69}\) See H.R. Rep. No. 101-247, at 1344–55 (1989) (stating that courts had extended § 104(a)(2) to “awards for personal injury that do not relate to a physical injury or sickness” and that the “committee believes that such treatment is inappropriate where no physical injury or sickness is involved”); see also Hobbs, *supra* note 23, at 74 (“The accompanying committee report confirmed the House’s recognition that courts were interpreting § 104(a)(2) too broadly.”).

\(^{70}\) See Hobbs, *supra* note 23, at 74 (“The conference committee rejected the proposed amendment . . . .”)


\(^{72}\) Hobbs, *supra* note 23, at 75.
PHYSICAL CONSEQUENCES

intentions. All courts were now free to apply § 104(a)(2)’s exclusion to nonphysical injuries. This judicial freedom led to confusion in lower courts as to how to apply § 104(a)(2) to nonphysical injuries, especially in the context of recovery for employment discrimination. This confusion ultimately compelled the Supreme Court to intercede.

D. A Two-Part Test Emerges

In the 1990s, the U.S. Supreme Court decided two employment discrimination cases, resulting in a two-part test for damage awards to be excluded under § 104(a)(2). To exclude damage awards under § 104(a)(2), a taxpayer must establish that the “cause of action giving rise to the recovery [was] ‘based upon tort or tort type rights’” and that the “damages were received ‘on account of personal injuries or sickness.’” Subsequent congressional and judicial action refined the Court’s two-part test, but the framework remained applicable.

73. See id. (“This was precisely the opposite of the intent of the House provision . . . .”).
74. See id. (noting that courts were “free to apply § 104(a)(2) to nonphysical injuries”).
75. See id. at 75–78 (noting that “courts struggled in determining the applicability of § 104(a)(2) to employment discrimination recoveries” and discussing the events leading up to United States v. Burke, 504 U.S. 229 (1992)).
76. See id. (noting that “courts struggled in determining the applicability of § 104(a)(2) to employment discrimination recoveries” and discussing the events leading up to Burke).
77. See Comm’r v. Schleier, 515 U.S. 323, 336–37 (1995) (determining that for a recovery to be excluded under § 104(a)(2), a taxpayer must establish that the “cause of action giving rise to the recovery [was] ‘based upon tort or tort type rights’” and that the “damages were received ‘on account of personal injuries or sickness’”); United States v. Burke, 504 U.S. 229, 241 (1992) (determining that damages received must redress tort or tort-like personal injury claims to be excluded under § 104(a)(2)); see also Hobbs, supra note 23, at 81 (stating that, after Schleier, a taxpayer must satisfy two separate requirements to exclude a damage award under § 104(a)(2)).
1. United States v. Burke

In 1992, the U.S Supreme Court decided United States v. Burke. Burke held that the taxpayers’ backpay awards received as settlement of their claims under Title VII of the Civil Rights Act of 1964 (Title VII) were not excludable from income under § 104(a)(2). The Court’s decision did not hinge on whether the claims were physical or nonphysical, however. Instead, the Court specifically addressed whether an award of back wages redressed a “tort-like personal injury within the meaning of § 104(a)(2) and the applicable regulations.” Because Title VII did not allow compensatory damages and limited the available remedies to “backpay, injunctions, and other equitable relief,” the Court concluded that Title VII did not redress “a tort-like personal injury within the meaning of § 104(a)(2) and the applicable regulations.” Before concluding that Title VII backpay awards were not excludable from respondent’s gross income under § 104(a)(2), however, the Court acknowledged a few important precepts of both the Tax Code and § 104(a)(2)’s exclusion.

The Court first noted that the Tax Code’s definition of gross income “sweeps broadly” to include “all income from whatever

---

80. United States v. Burke, 504 U.S. 229, 242 (1992) (holding that “the backpay awards received by respondents in settlement of their Title VII claims [were] not excludable from gross income as ‘damages received . . . on account of personal injuries’ under § 104(a)(2)” (alteration in original)).
82. Burke, 504 U.S. at 242.
83. See id. at 241 (finding that Title VII backpay awards did not redress a “tort-like personal injury” and thus could not be excluded from gross income).
84. See id. (concluding that the award of back wages under a Title VII claim did not redress tort-like personal injuries); see also I.R.C. § 104(a)(2) (2006); 26 CFR § 1.104-1(c) (1991) (“The term ‘damages received (whether by suit or agreement)’ means an amount received (other than workmen’s compensation) through prosecution of a legal suit or action based upon tort or tort type rights.”).
85. Burke, 504 U.S. at 241.
86. Id. at 241–42 (concluding that the award of back wages under a Title VII claim did not redress tort or tort-like personal injuries and were thus not excludable from gross income under § 104(a)(2)).
87. Id. at 233.
source derived." Further, the Court acknowledged Congress’s intention to exert its full taxing power when enacting § 61(a). Also, the Court noted that after its decision in *Glenshaw Glass*, it has interpreted any “accession to wealth” to be within the definition of gross income. Although gross income includes any accession to wealth, exclusions from gross income must be “specifically enumerated elsewhere in the Code.” Additionally, these specifically enumerated exclusions from gross income are “narrowly construed.” Finally, the Court acknowledged that although courts once interpreted the personal injury exclusion to encompass only “physical” injuries, it was now well accepted that the exclusion encompassed nonphysical injuries as well.

2. Commissioner v. Schleier

The U.S. Supreme Court’s next contribution to § 104(a)(2) came three years later when it decided *Commissioner v. Schleier*. Similar to the issue addressed in *Burke*, the Court considered whether § 104(a)(2) authorized the taxpayer to exclude the amount received as settlement for his claims for both backpay and liquidated damages under the Age Discrimination in Employment Act (ADEA) from his gross income. Half of the taxpayer’s settlement was attributed to backpay and half to

---

88. *Id.* (quoting I.R.C. § 61(a) (2006)).
89. *Id.*
90. *Id.* (citations omitted) (quoting Comm’r v. Glenshaw Glass Co., 348 U.S. 426, 431 (1955)).
91. *Id.*
92. See *id.* at 248 (Souter, J., concurring) (noting the “default rule of statutory interpretation that exclusions from income must be narrowly construed”).
93. See *id.* at 235 n.6 (majority opinion) (noting that the exclusion’s scope had previously been limited to only physical injuries, but that it was now settled that it extended to “nonphysical injuries to the individual”).
95. See *id.* at 324–25 (“The question presented is whether § 104(a)(2) . . . authorizes a taxpayer to exclude from his gross income the amount received in settlement of a claim for backpay and liquidated damages under the Age Discrimination in Employment Act of 1967 (ADEA.”).
liquidated damages. The Tax Court determined the entire amount to be excludable, and the Fifth Circuit affirmed. The Supreme Court, however, reversed, holding that § 104(a)(2) did not exclude the taxpayer’s ADEA recovery.

As in Burke, the Court emphasized the broadly sweeping definition of gross income and the corollary that exclusions must be construed narrowly. And also as it did in Burke, the Court looked to the available remedies under the ADEA, noting that the ADEA allowed for liquidated damages “only in cases of willful violations” and that the ADEA did not permit a “separate recovery of compensatory damages for pain and suffering or emotional distress.” The Court, however, did not first analyze the claim under Burke’s “tort or tort-type rights” standard for excludability. Instead, the Court went on to determine whether the settlement award was “on account of personal injuries.”

The Court used injuries resulting from an automobile accident as an example to illustrate its idea of what constituted a “personal injury.” In the Court’s example, a taxpayer’s medical expenses, lost wages, and pain, suffering, and emotional distress would all be excludable under § 104(a)(2) as being on account of personal injuries arising out of the automobile accident. The Court compared backpay in the present case to lost wages resulting from the hypothetical automobile accident. But unlike the lost wages in the automobile accident example, the Court

96. See id. at 326 (“Half of respondent’s award was attributed to ‘backpay’ and half to ‘liquidated damages.’”).
97. See id. at 327 (stating that the Tax Court and the Fifth Circuit allowed the amount to be excluded, but reversing those decisions).
98. See id. at 327–28 (noting the “sweeping scope” of § 61(a) and that exclusions from income are narrowly construed).
99. Id. at 326.
100. See id. at 328–32 (determining whether the backpay portion or the liquidated damages portion of respondent’s settlement was “on account of personal injuries”).
101. See id. at 328–31 (noting potential claims and injuries arising out of a hypothetical car crash to illustrate an example of a personal injury).
102. See id. at 329–30 (determining that all recoveries associated with injuries arising out an automobile accident would be excludable under § 104(a)(2)).
103. See id. at 330–31 (comparing respondent’s recovery of back wages to recovery for lost wages that hypothetically resulted from injuries sustained in a car accident).
determined that the taxpayer’s backpay was not excludable. The Court reasoned that the taxpayer’s back wages were not excludable because any personal injury caused by the age discrimination was not in any way linked to the loss of wages. Thus, age discrimination directly causes two distinct injuries, only one of which is a personal injury. First, the employer’s conduct directly causes lost wages. Second, the tortious conduct often causes a nonphysical personal injury. The lost wages, however, do not flow from the personal injury and therefore do not meet § 104(a)(2)’s on account of requirement.

After the Court concluded that the recovery of back wages under the ADEA was not on account of any personal injury, it then addressed respondent’s argument that § 104(a)(2) excluded the liquidated damages portion of the award. The Court concluded that the liquidated damages also could not be excluded. Justice Stevens, writing for the majority, acknowledged that the liquidated damages might have come within § 104(a)(2)’s exclusion had that been Congress’s intention. Justice Stevens, however, found nothing to indicate that Congress intended the ADEA’s liquidated damages to compensate “personal rather than economic” injuries. Instead,

104. See id. (“[Section] 104(a)(2) does not permit the exclusion of respondent’s back wages because the recovery of back wages was not ‘on account of’ any personal injury and because no personal injury affected the amount of back wages recovered.”).

105. See id. at 330 (“In age discrimination, the discrimination causes both personal injury and loss of wages, but neither is linked to the other.”); see also Hobbs, supra note 23, at 81–82 (stating that Justice Stevens’s language, when writing for the majority in determining that a recovery under the ADEA was not a personal injury, “defie[d] logic”).


108. See id. (concluding that “liquidated damages under the ADEA, like back wages under the ADEA, are not received ‘on account of personal injury or sickness’”).

109. See id. at 331 (determining that the Court’s previous observation—that the liquidated damages authorized by the Fair Labor Standards Act of 1938 (FLSA), 29 U.S.C. §§ 201–219, may compensate for some “obscure” injuries—did not necessarily mean that Congress intended the ADEA’s liquidated damages provision to be “personal rather than economic”).

110. Id.
Justice Stevens found the ADEA’s liquidated damages provision to be punitive in nature. Accordingly, he concluded that the recovery’s liquidated damages portion, along with the portion attributed to backpay, was not received “on account of personal injury of sickness.”

Next, the Court examined whether the taxpayer’s ADEA recovery was “based upon ‘tort or tort type rights’ as that term was construed in Burke.” Again the Court looked to the ADEA’s remedial scheme, declared that the jury trial and liquidated damages provisions were not sufficient to bring an ADEA recovery within the ambit of § 104(a)(2), and concluded that “a recovery under the ADEA [was] not one that is ‘based upon tort or tort type rights.’”

Throughout its opinion, the Court seemed uncomfortable applying § 104(a)(2) to nonphysical injuries. But the Court did not go so far as to provide a rule limiting the exclusion to physical injuries. Instead, Justice Stevens articulated a two-part test for excludability under § 104(a)(2). First, a taxpayer must “demonstrate that the underlying cause of action giving rise to the recovery is ‘based upon tort or tort type rights.’” Second, the taxpayer must “show that the damages were received on ‘account of personal injuries or sickness.’” For a recovery to be excludable, a taxpayer must satisfy Schleier’s two independent requirements.

---

111. Id.
112. Id. at 332.
113. See id. at 334–36 (addressing respondent’s argument that the recovery was excludable under Burke).
114. See id. at 335–36 (determining that neither the ADEA’s jury trial provision nor its liquidated damages provision were sufficient to bring a recovery under the ADEA within the § 104(a)(2) exclusion).
115. See Hobbs, supra note 23, at 82 (“The Supreme Court’s decisions in Burke and Schleier clearly exhibited the Court’s discomfort in applying § 104(a)(2) in a nonphysical context.”).
116. See id. (noting that the Court did not “articulate a rule limiting § 104(a)(2) to physical injuries”).
118. Id.
119. Id.
120. See id. at 336–37 (stating the “two independent requirements that a taxpayer must meet before a recovery may be excluded under § 104(a)(2)”)

Since Schleier, many courts have conceded that a recovery was “based upon tort or tort type rights,” and Treasury has recently finalized regulations eliminating the requirement entirely. Instead, courts have focused on whether the damages were received on account of a personal injury. Schleier, however, provided little detail on when a recovery is on account of a personal injury. After Schleier, courts added more definition to the on account of requirement, but Schleier’s framework remained the standard.

III. The Courts Refine § 104(a)(2)’s “On Account Of” Requirement

In O’Gilvie v. United States, the U.S. Supreme Court addressed whether § 104(a)(2) excluded punitive damages, and provided a narrow interpretation of § 104(a)(2)’s on account of requirement. The Court concluded that the petitioner’s punitive damages were not received on account of personal injury.

---

121. See, e.g., Domeny v. Comm’r, T.C. Memo. 2010-9, at 3 (2010) (concluding summarily that the taxpayer’s action was based upon tort or tort-type rights).
122. See, Treas. Reg. § 1.104-1(c) (as amended in 2012) (excluding from gross income damages “on account of personal injuries or sickness”).
123. See e.g., O’Gilvie v. United States, 519 U.S. 79, 82–83 (1996) (holding that “on account of” required a stronger causal connection than “but for,” and that only “damages that were awarded by reason of, or because of, the personal injuries” could be excluded under § 104(a)(2)).
124. See Hobbs, supra note 23, at 81 (“The Court did not provide a test for determining when an amount was received for personal injury, in effect saying, we will know it when we see it.”).
125. See infra Part III (discussing court opinions interpreting “on account of”).
126. See Amos v. Comm’r, T.C. Memo 2003-329, at 4 (2003) (“The 1996 amendment does not otherwise change the requirements of section 104(a)(2) or the analysis set forth in Commissioner v. Schleier; it imposes an additional requirement for an amount to qualify for exclusion from gross income under that section.” (citation omitted)).
127. O’Gilvie v. United States, 519 U.S. 79, 82–83 (1996) (holding that “on account of” required a stronger causal connection than “but for,” and that only “damages that were awarded by reason of, or because of, the personal injuries” could be excluded under § 104(a)(2)).
128. See id. at 82 (resolving a circuit split on the proper interpretation of Schleier’s “on account of” requirement).
injuries and thus not excludable under § 104(a)(2). The Court rejected the petitioner's argument that § 104(a)(2) only required a "but for" connection between the damages received and the underlying personal injury. Instead, the Court agreed with the Government that § 104(a)(2) was "applicable only to those personal injury lawsuit damages that were awarded by reason of, or because of" the underlying personal injury.

Although O'Gilvie did not explicitly determine what damages fell within its interpretation, the Court narrowly interpreted § 104(a)(2)'s on account of requirement in accord with customary Tax Code interpretation. After O'Gilvie, lower courts continued to refine § 104(a)(2)'s requirement.

A. The "Direct Causal Link"

In Banaitis v. Commissioner, the U.S. Court of Appeals for the Ninth Circuit considered whether a taxpayer's economic and punitive damages recovered in his wrongful termination suit fell within § 104(a)(2)'s exclusion. In Banaitis, the taxpayer received $8,728,599 as settlement for his various claims against

---

129. See id. at 81 ("We conclude that the punitive damages here were not received 'on account of' personal injuries; hence [section 104(a)(2)] does not apply, and the damages are taxable.").

130. See id. at 82 (rejecting the petitioner's argument for "but for" causation).

131. See id. at 83 (agreeing with the Government's interpretation of the statute).

132. See id. at 82 (rejecting petitioner's broad reading of § 104(a)(2)'s "on account of" requirement).

133. See United States v. Burke, 504 U.S. 229, 248 (1992) (Souter, J., concurring in judgment) (noting the "default rule of statutory interpretation that exclusions from income must be narrowly construed").

134. See infra Part III.A (discussing the lower courts' interpretation of § 104(a)(2)'s "on account of" requirement).

135. Banaitis v. Comm'r, 340 F.3d 1074, 1080–81 (9th Cir. 2003) (holding that the taxpayer's economic and punitive damages recovered in his wrongful termination suit against his former employer were not "on account of" personal injuries and thus not excludable under § 104(a)(2)), rev'd on other grounds, Comm'r v. Banks, 543 U.S. 426 (2005).

136. See id. at 1079–81 (considering whether a taxpayer's economic and punitive damages recovered in his wrongful termination suit fell within § 104(a)(2)'s exclusion).
his former employer. The taxpayer claimed that § 104(a)(2) excluded the full amount, but the Ninth Circuit concluded that the entire amount was taxable.

Banaitis alleged that his employers’ conduct resulted in a number of physical injuries. Those injuries included “headaches, insomnia, gastrointestinal disorders, bleeding gums, and various orthopedic problems.” Addressing whether the award satisfied the “on account of” requirement, the court stated that to satisfy § 104(a)(2)’s second prong, the taxpayer must show the “damage award [was] more than only proximately caused by [his employer’s] tortious conduct; it must also be directly causally related to [his] personal injuries.” Further, the court stated that the on account of requirement “can only be satisfied if there is ‘a direct causal link’ between the damages and the personal injuries sustained.” The court noted that in the ordinary personal injury tort action, it is relatively easy to discern what damages are on account of a personal injury:

The tortious act causes personal injuries which, in turn, cause further damages, such as economic loss due to physical inability to work. Thus, in the paradigmatic personal injury case, both non-pecuniary damages (such as pain and suffering) and economic damages (such as wage loss, diminished work capacity, etc.) may be excluded from gross income because the losses are “on account of” personal injury.

When damages arise from “economic or commercial tort” actions, on the other hand, it is more difficult to discern what damages are on account of personal injuries. Put another way,
it is more difficult to establish the requisite “direct causal link” between the damages received and the personal injuries sustained. The Ninth Circuit stated that “[i]n such economic or commercial tort cases, economic damages are often caused solely by the tortious action itself, rather than as a consequence of personal injury.” The court gave the typical wrongful discharge lawsuit as an example, stating “wage loss is typically caused by the tortious employment termination, not by any physical injury that may also have been caused by the wrongful discharge.” The court determined that Banaitis’s alleged physical injuries “did not cause his wage loss,” and that his damages “were not causally related to [his] alleged personal injuries.” This analysis is nearly identical to Schleier’s analysis in which the Supreme Court determined that age discrimination caused both lost wages and a personal injury, but that the two were not linked. Accordingly, the Ninth Circuit concluded that § 104(a)(2)’s exclusion did not apply and that Banaitis’s damages award was fully taxable.

Banaitis and O’Gilvie illustrate that § 104(a)(2) excludes only damages directly caused by a “personal injury or sickness.” Notably, the two cases addressed § 104(a)(2)’s required causal connection between the taxpayer’s damages received and an

---

146. See id. (noting that “economic or commercial tort actions present a different circumstance,” and that “the ‘direct causal link’ question requires a fact-specific analysis”).

147. Id.

148. Id.; cf. Comm’r v. Schleier, 515 U.S. 323, 330 (1995) (“In age discrimination, the discrimination causes both personal injury and loss of wages, but neither is linked to the other.”).

149. See Banaitis v. Comm’r, 340 F.3d 1074, 1080 (9th Cir. 2003) (“The personal injuries Banaitis alleges (e.g., headaches, insomnia, gastrointestinal disorders, bleeding gums, and back aches) did not cause his wage loss.”); see also Banaitis v. Comm’r, T.C. Memo 2002-5, at 4 (2002) (determining that Banaitis’s economic damages were not received “on account of” his alleged personal injuries because “he was not forced to leave his job because of those injuries”).

150. Supra note 105 and accompanying text; see also Schleier, 515 U.S. at 330 (“In age discrimination, the discrimination causes both personal injury and loss of wages, but neither is linked to the other.”).

151. See Banaitis, 340 F.3d at 1081 (affirming that § 104(a)(2)’s exclusion did not apply and that Banaitis’s damage awards “should have been included in his gross income”).
already-established personal injury.\textsuperscript{152} When Congress amended § 104(a)(2)’s exclusion to add a “physical” requirement,\textsuperscript{153} establishing § 104(a)(2)’s qualifying “personal physical injury or physical sickness” became a much more difficult task.

\textbf{IV. The Small Business Job Protection Act of 1996}

The Small Business Job Protection Act\textsuperscript{154} (the 1996 Act) became law on August 20, 1996.\textsuperscript{155} Section 1605 of the 1996 Act amended § 104(a).\textsuperscript{156} Section 1605(a) specifically amended § 104(a)(2), providing that gross income does not include the “amount of any damages (other than punitive damages) received (whether by suit or agreement and whether as lump sums or as periodic payments) on account of personal physical injuries or physical sickness.”\textsuperscript{157} Section 1605(b) also amended § 104(a) to expressly provide that “emotional distress shall not be treated as a physical injury or physical sickness” for purposes of § 104(a)(2).\textsuperscript{158}

While prohibiting emotional distress from constituting a physical injury or physical sickness, § 104(a) provides that the prohibition does not apply to “damages not in excess of the amount paid for medical care (described in subparagraph (A) or (B) Section 213(d)(1)) attributable to emotional distress.”\textsuperscript{159}

\begin{flushright}
152. \textit{See supra} Parts III.A–B (discussing O’Gilvie and Banaitis).
157. \textit{Id. § 1605(a), 110 Stat. at 1838 (emphasis added)}.
158. \textit{See id. § 1605(b), 110 Stat. at 1838 (“Section 104(a) is amended by striking the last sentence and inserting the following sentence: ‘For purposes of paragraph (2), emotional distress shall not be treated as a physical injury or physical sickness.’”)}.
159. \textit{See I.R.C. § 104(a) (2006) (“The preceding sentence shall not apply to an amount of damages not in excess of the amount paid for medical care (described in subparagraph (A) or (B) of section 213(d)(1)) attributable to emotional distress.”)}.
Section 213(d)(1)(A) defines “medical care” to include amounts paid “for the diagnoses, cure, mitigation, treatment, or prevention of disease, or for the purpose of affecting any structure or function of the body.”\textsuperscript{160} And § 213(d)(1)(B) provides that “medical care” includes amounts paid “for transportation primarily for and essential to medical care referred to in subparagraph (A).”\textsuperscript{161} Presumably, this language permits a taxpayer to exclude damages received solely for emotional distress from gross income under §104(a)(2) up to the amount the damages compensate for “medical care” expenses otherwise deductible under § 213(d)(1)(A) or (B).\textsuperscript{162}

In the 1996 Act’s legislative history (the House Report), Congress noted courts’ prior broad extensions of §104(a)(2) to nonphysical personal injuries and reiterated that all punitive damages should be included in income.\textsuperscript{163} Congress did provide, however, that punitive damages received in a wrongful death suit may be excluded from gross income if applicable state law provides that “only punitive damages may be awarded in a wrongful death action.”\textsuperscript{164}

Most importantly, under the heading “Include in income damage recoveries for nonphysical injuries,”\textsuperscript{165} the House Report provided guidance on when damages would be excludable under § 104(a)(2), as amended. The House Report stated that “[i]f an action has its \textit{origin} in a physical injury or physical sickness, \textit{then} all damages (other than punitive damages) that flow therefrom are treated as payments received on account of physical injury or physical sickness whether or not the recipient of the damages is the injured party.”\textsuperscript{166} Congress’s language indicated that a taxpayer’s action must first have its “origin” in a physical injury

\textsuperscript{160} Id. § 213(d)(1)(A).
\textsuperscript{161} Id. § 213(d)(1)(B).
\textsuperscript{162} Id. § 213(d)(1)(A)–(B).
\textsuperscript{163} See H.R. REP. No. 104-737, at 300–01, \textit{reprinted in} 1996 U.S.C.C.A.N. 1677, 1792–93 [hereinafter HOUSE REPORT] (noting that courts have interpreted the exclusion broadly “to cover awards for personal injury that do not relate to a physical injury or sickness” and reiterating that punitive damages should be included in income).
\textsuperscript{164} Id. at 301.
\textsuperscript{165} Id.
\textsuperscript{166} Id. (emphasis added).
or physical sickness before proceeding to determine whether an amount is received “on account of a personal physical injury or physical sickness.”

The House Report expressly stated, however, that emotional distress, including symptoms thereof, is not considered a physical injury or physical sickness for purposes of the § 104(a)(2)’s exclusion. As an example, the House Report provided that “the exclusion from gross income does not apply to any damages received . . . based on a claim of employment discrimination or injury to reputation accompanied by a claim of emotional distress.” When an action has its origin in a physical injury or physical sickness, however, and the physical injury or physical sickness causes the taxpayer to suffer emotional distress, § 104(a)(2) does exclude damages allocated to such emotional distress. The House Report explained that “[b]ecause all damages received on account of physical injury or physical sickness are excludable from gross income, the exclusion from gross income applies to any damages received based on a claim of emotional distress that is attributable to a physical injury or physical sickness.” Lastly, the House Report stated that the gross income exclusion applies to medical care expenses attributable to emotional distress, affirming the presumption created by § 104(a)’s text.

167. See Robert W. Wood, Waiting to Exhale: Murphy Part Deux and Taxing Damage Awards, ALI-ABA COURSE OF STUDY, at 5 (Feb. 7–8, 2008) (“Congress require[s] that the action have its origin in a physical injury or sickness.”).

168. See HOUSE REPORT, supra note 163, at 301 (stating that the House bill “specifically provides that emotional distress is not considered a physical injury or physical sickness” and also that the term emotional distress includes symptoms resulting from such emotional distress).

169. Id.

170. Id.

171. See id. (“[T]he exclusion from gross income specifically applies to the amount of damages received that is not in excess of the amount paid for medical care attributable to emotional distress.”).

172. See supra notes 159–62 and accompanying text.
V. Section 104(a)(2)’s Exclusion After the 1996 Act

Because Congress prohibited emotional distress and its resulting symptoms from being excluded from gross income under § 104(a)(2), after the 1996 Act, taxpayers faced the additional burden of establishing that their personal injury or sickness was physical rather than emotional. Courts, however, continued to only apply Schleier’s two-part test to § 104(a)(2)’s exclusion, failing to accommodate the added physical requirement. Although their reasoning left open the potential for misapplication, the circuit courts, amidst allegations of physical injuries or physical sicknesses, refused to apply § 104(a)(2)’s exclusion to taxpayers’ recoveries for nonphysical or emotional injuries. Perhaps the most notable—or infamous—case to do so is one the U.S. Court of Appeals for the D.C. Circuit decided in 2007.

A. The Circuits Reach the Right Outcome

In Murphy v. Internal Revenue Service, after determining the taxpayer’s compensatory damages were not income within the Sixteenth Amendment’s meaning, then sua sponte vacating its judgment and rehearing the case, the U.S. Court of Appeals for the D.C. Circuit held that Murphy’s award could not be excluded under § 104(a)(2). Additionally, the court held that Murphy’s

173. See Amos v. Comm’r, T.C. Memo 2003-329, at 4 (2003) (“The 1996 amendment does not otherwise change the requirements of section 104(a)(2) or the analysis set forth in Commissioner v. Schleier; it imposes an additional requirement for an amount to qualify for exclusion from gross income under that section.” (citation omitted)).

174. See infra Part V.A (discussing circuit court cases following the 1996 Act).

175. See generally Murphy v. Internal Revenue Serv., 493 F.3d 170 (D.C. Cir. 2007).

176. See Murphy v. Internal Revenue Serv., 493 F.3d 170, 171 (D.C. Cir. 2007) (holding that the taxpayer’s “compensation was not ‘received . . . on account of personal physical injuries’ excludable from gross income under § 104(a)(2)” and that § 61(a)’s gross income definition included the taxpayer’s award (alteration in original)).

177. See id. (“We hold, first, that Murphy’s compensation was not ‘received . . . on account of personal physical injuries’ excludable from gross income under § 104(a)(2).” (alteration in original)).
compensation fell within § 61(a)’s definition of income and admitted that there is no constitutional impediment to taxing personal injury awards.

Murphy alleged that her former employer violated various whistleblower statutes and “blacklisted” her. After a hearing before the Secretary of Labor, an Administrative Law Judge (ALJ) recommended compensatory damages totaling $70,000 to Murphy. The Department of Labor Administrative Review Board (Board) affirmed the ALJ’s findings and granted the award. $45,000 of the damages were allocated to “past and future emotional distress” and $25,000 of the damages were allocated to “injury to [Murphy’s] vocational reputation.” The court noted that “[n]one of the award was for lost wages or diminished earning capacity.” After including the full amount in her gross income on her 2000 tax return, Murphy filed an amended return seeking a $20,665 refund. She argued, among other things, that § 104(a)(2) excluded her award from her gross income. This time the D.C. Circuit rejected Murphy’s argument in full.

Before the ALJ, Murphy submitted evidence that she suffered both “somatic”—relating to, or affecting, the body—and “emotional” injuries. Her injuries included “bruxism,” or

178. See id. at 180 (holding that gross income under § 61(a) included Murphy’s compensatory award).
179. See id. at 173 (agreeing with the Government’s argument that there is no constitutional problem with taxing personal injury awards).
180. See id. at 171–72 (describing the taxpayer’s original allegations).
181. See id. at 172 (“[T]he ALJ recommended compensatory damages totaling $70,000.”).
182. See id. (“In 1999 the Department of Labor Administrative Review Board affirmed the ALJ’s findings and recommendations.”).
183. See id. (describing the damages allocation) (alteration in original).
184. Id.
185. See id. (stating that Ms. Murphy originally included the amount in her gross income and then filed an amended return).
186. See id. at 171 (noting Ms. Murphy’s arguments that her award should be excluded or that taxing her award was unconstitutional).
187. See id. (“We reject Murphy’s argument in all aspects.”).
188. See id. at 174 (defining “somatic”).
189. See id. at 172 (“A psychologist testified that Murphy had sustained both ‘somatic’ and ‘emotional’ injuries . . . .”).
teeth grinding, and “anxiety attacks, shortness of breath, and dizziness.” Murphy argued that § 104(a)(2)’s exclusion did not require a “physical stimulus” and that “substantial physical problems caused by emotional distress” should be considered a physical injury or physical sickness. The D.C. Circuit did not agree, finding that § 104(a)(2)’s exclusion did not apply to Murphy’s award.

The court first cited § 104(a)(2)’s post-amble, which states that “emotional distress shall not be treated as a physical injury or physical sickness,” then addressed the Government’s contentions. The Government argued that O’Gilvie dictated a “strong causal connection” between Murphy’s received damages and her alleged physical injuries. This meant Murphy had to demonstrate that her damages were awarded “because of” her physical injuries, which the Government claimed she failed to do. Further, based on the Board’s failure to reference any of Murphy’s physical injuries—especially her bruxism—the Government argued that “there was no direct causal link between the damages award at issue and [Murphy’s] bruxism.” The court agreed that a strong causal link was required, finding that O’Gilvie’s analysis of § 104(a)(2)’s on account of requirement remained controlling after the 1996 Act.

---

190. See id. (noting Murphy’s alleged injuries).
191. See id. at 175 (noting Murphy’s contentions as to why her award should be excluded from her gross income by § 104(a)(2)).
192. See id. at 176 (concluding that § 104(a)(2) did “not permit Murphy to exclude her award from gross income”).
193. See id. at 174 (“[F]or purposes of this exclusion, ‘emotional distress shall not be treated as a physical injury or physical sickness.’” (citing I.R.C. § 104(a) (2006))).
194. See id. at 175 (addressing the Government’s arguments as to why Murphy’s award should not be excluded).
195. See id. (noting the Government’s argument that the Supreme Court in O’Gilvie read § 104(a)(2)’s “on account of” language to require a “strong causal connection”).
196. See id. (“The Government therefore concludes Murphy must demonstrate she was awarded damages ‘because of’ her physical injuries, which the Government claims she has failed to do.”).
197. See id. (stating that the Board made no reference in its award to Murphy’s physical injuries and noting the Government’s argument) (alteration in original).
198. See id. at 176 (finding that O’Gilvie’s analysis of the phrase “on account
Next, the court conceded that Murphy suffered “physical manifestations of emotional distress,” and that the ALJ may have at best “considered her physical injuries indicative of the severity of [her] emotional distress.” The court concluded, however, that because “her physical injuries themselves were not the reason for the award,” Murphy’s damages were not awarded “because of” her physical injuries. Thus, Murphy could not use § 104(a)(2) to exclude her award from gross income.

Other circuits have also refused to apply § 104(a)(2)’s exclusion to damage awards in which the taxpayers alleged to have suffered physical injuries. The U.S. Court of Appeals for the Tenth Circuit, in Johnson v. United States, refused to exclude any portion of a taxpayer’s award under the Americans with Disabilities Act (ADA) for being unlawfully terminated. The taxpayer suffered physical injuries while working as a juvenile guard for the State of Colorado’s Department of Corrections and attempting to restrain an inmate. After the injuries, the taxpayer was unable to perform the duties of a juvenile guard. Instead of accommodating him under the ADA with another job he could perform, Colorado terminated his

199. See id. (acknowledging that Murphy suffered physical injuries and that the ALJ may have considered them when recommending her award) (emphasis added).

200. See id. (concluding that Murphy’s “damages were not ‘awarded by reason of, or because of, . . . [physical] personal injuries’” (alterations in original) (citing O’Gilvie v. United States, 519 U.S. 79, 83 (1996))).

201. See id. (“Therefore, § 104(a)(2) does not permit Murphy to exclude her award from gross income.”).

202. See generally, e.g., Stadnyk v. Comm’r, 367 F. App’x 586 (6th Cir. 2010); Lindsey v. Comm’r, 422 F.3d 684 (8th Cir. 2005); Johnson v. United States, 76 F. App’x 873 (10th Cir. 2003).

203. Johnson v. United States, 76 F. App’x 873, 877–78 (10th Cir. 2003) (holding that § 104(a)(2) did not exclude the taxpayer’s ADA damages award for front and back pay from his gross income).


205. See Johnson, 76 F. App’x at 877–78 (concluding that none of the taxpayer’s damages could be excluded from his gross income under § 104(a)(2)).

206. See id. at 874 (noting that the taxpayer received physical injuries while attempting to restrain a juvenile inmate).

207. See id. (noting that the taxpayer’s injuries prohibited him from fulfilling his duties as a guard).
employment.\footnote{208} He sued the State of Colorado in state court and a jury returned a verdict in his favor.\footnote{209}

The taxpayer did not challenge the taxability of the portion of the jury verdict allocated to “emotional distress, pain, suffering and mental anguish.”\footnote{210} Instead, the taxpayer sought to exclude the $293,400 amount awarded for back and front pay from his gross income.\footnote{211} He argued that § 104(a)(2) excluded the damages because his physical injuries led to his unlawful termination.\footnote{212} The court refused to apply § 104(a)(2)’s exclusion to the award, however, reasoning that because the unlawful termination caused the loss of income, and not any alleged personal physical injury,\footnote{213} the damages were not received “on account of personal physical injuries.”\footnote{214}

When faced with similar issues, the Sixth and Eighth Circuits reached results similar to the Tenth Circuit’s.\footnote{215} In Stadnyk v. Commissioner,\footnote{216} the U.S. Court of Appeals for the

\footnotesize{\begin{itemize}
\item 208. See id. (noting that Colorado terminated his employment).
\item 209. See id. (noting that the taxpayer sued).
\item 210. See id. (noting that the taxpayer only sought a refund for the “tax liability on the back and front pay portions of the award”).
\item 211. See id.
\item 212. See id. at 876 (noting the taxpayer’s argument that § 104(a)(2) excluded his award because his physical injuries led to his unlawful termination).
\item 213. See id. at 877 (“[T]he actual cause of the loss of income and the ADA action was the unlawful termination, not the personal physical injury.”); cf. Comm’r v. Schleier, 515 U.S. 323, 330 (1995) (“In age discrimination, the discrimination causes both personal injury and loss of wages, but neither is linked to the other.”); Banaitis v. Comm’r, 340 F.3d 1074, 1080 (9th Cir. 2003) (explaining that wage loss is directly caused by the employment discrimination, not by any physical injury).
\item 214. See Johnson v. United States, 76 F. App’x 873, 877–78 (10th Cir. 2003) (concluding that the front and back pay damages lacked a “direct causal link” to the taxpayer’s physical injuries and were therefore not received “on account of” such injuries).
\item 215. See Stadnyk v. Comm’r, 367 F. App’x 586, 594 (6th Cir. 2010) (concluding that the taxpayer’s award for false imprisonment claims could not be excluded from gross income under § 104(a)(2) because there was no “causal connection between any physical injury and the settlement award”); Lindsey v. Comm’r, 422 F.3d 684, 686, 689 (8th Cir. 2005) (holding that § 104(a)(2) did not exclude the taxpayer’s settlement for “claims for tortious interference with contracts, for personal injury including injury to [his] personal and professional reputation and emotional distress, [and] humiliation and embarrassment”).
\item 216. See Stadnyk v. Comm’r, 367 F. App’x 586, 594 (6th Cir. 2010) (holding that the taxpayer’s award for false imprisonment claims could not be excluded}
Sixth Circuit found § 104(a)(2)'s exclusion inapplicable to a $49,000 settlement award for claims relating to false imprisonment.\textsuperscript{217} Although the taxpayer had testified she suffered no physical injuries resulting from her arrest and detention, and nothing in the record suggested a physical injury, she nonetheless argued that her physical restraint alone constituted a personal physical injury.\textsuperscript{218} The court disagreed.\textsuperscript{219} The Sixth Circuit determined that false imprisonment did not necessarily involve a physical injury,\textsuperscript{220} and concluded there was no “direct causal link” between any other alleged physical injuries and the settlement award.\textsuperscript{221}

In \textit{Lindsey v. Commissioner},\textsuperscript{222} the U.S. Court of Appeals for the Eighth Circuit affirmed the Tax Court’s denial of § 104(a)(2)’s exclusion to the taxpayer’s (Lindsey) $2 million settlement for claims against the taxpayer’s former employer.\textsuperscript{223} Lindsey’s claims included “tortious interference with contracts, . . . personal injury including injury to [his] personal and professional reputation and emotional distress, [and] humiliation and embarrassment.”\textsuperscript{224} Lindsey alleged that he suffered physical stress-related symptoms including hypertension, periodic

\begin{itemize}
  \item \textsuperscript{217} See \textit{id.} at 594 (finding § 104(a)(2) inapplicable to the taxpayer's settlement award).
  \item \textsuperscript{218} See \textit{id.} at 592–93 (noting that “Mrs. Stadnyk testified that she did not suffer any physical injury as a result of her arrest and detention,” that “[n]othing in the record suggests that Mrs. Stadnyk suffered physical, as opposed to emotional, injuries,” and that she argued that her physical restraint alone constituted a physical injury).
  \item \textsuperscript{219} See \textit{id.} at 593 (determining that a false imprisonment victim is not necessarily physically injured).
  \item \textsuperscript{220} See \textit{id.}
  \item \textsuperscript{221} See \textit{id.} at 594 (“Petitioners have failed to offer any concrete evidence demonstrating a causal connection between any physical injury and the settlement award.”).
  \item \textsuperscript{222} Lindsey v. Comm’r, 422 F.3d 684, 686, 689 (8th Cir. 2005) (holding that § 104(a)(2) did not exclude the taxpayer’s settlement for “claims for tortious interference with contracts, for personal injury including injury to [his] personal and professional reputation and emotional distress, [and] humiliation and embarrassment”).
  \item \textsuperscript{223} See \textit{id.} (“Therefore, the tax court properly denied the exclusion.”)
  \item \textsuperscript{224} \textit{Id.} at 685.
\end{itemize}
impotency, insomnia, fatigue, occasional indigestion, and urinary incontinence. Before the Eighth Circuit, Lindsey argued that the Tax Court erroneously concluded that § 104(a)(2) did not apply to any portion of the award.

In the Tax Court opinion under review, the court looked both to the settlement’s terms and the payor’s intent when analyzing whether § 104(a)(2) excluded Lindsey’s settlement. First, the Tax Court stated that Congress explicitly excluded “emotional distress and related injuries” from the “definition of physical injuries or physical sickness.” The court then analyzed the settlement’s terms, which simply reiterated Lindsey’s claims against his employer. The court found that “[i]njury to reputation, humiliation, and embarrassment are akin to emotional distress” and that “tortious interference with contracts is an economic injury, not a physical injury.” Accordingly, the court found that § 104(a)(2) did not exclude damages received on account of such claims.

Next, the court considered whether Lindsey’s alleged physical injuries served as the basis for any portion of the settlement agreement. To do this, the court analyzed the intent of the payor. Lindsey’s physician testified that he suffered physical manifestations of stress, including hypertension, fatigue, occasional indigestion, and insomnia. Lindsey’s physician further testified that Lindsey’s hypertension could lead to “strokes, heart attacks, and kidney disease.”

---

225. See id. at 688 (noting Lindsey’s alleged physical symptoms).
226. See id. at 687 (“Lindsey[] also contend[s] the Tax Court erred in finding the physical sickness Lindsey suffered was a type not excludable under I.R.C. § 104(a)(2).”).
227. See Lindsey v. Comm’r, T.C. Memo 2004-113, at 5–6 (analyzing the settlement’s terms and discerning the payor’s intent).
228. Id. at 5.
229. See id. (analyzing the settlement’s terms).
230. Id.
231. See id. (finding that § 104(a)(2) does not exclude damages received on account of such claims).
232. See id. at 5–6 (considering whether any portion of the settlement was intended to compensate Lindsey’s physical injuries).
233. See id. (analyzing the intent of the payor).
234. See id. at 5 (excerpting the physician’s testimony).
235. Id.
that Congress intended Lindsey’s injuries to be “within the definition of emotional distress,” citing the House Report’s statement that emotional distress included symptoms.\footnote{236} Finally, the court found that even if Lindsey had suffered a “personal physical injury within the meaning of section 104(a)(2), such injury could not have been the basis for settlement” because there was no evidence that Lindsey communicated such injuries to the payor.\footnote{237}

On review, the Eighth Circuit mostly confirmed the Tax Court’s findings and reasoning.\footnote{238} Notably, though, the Eighth Circuit implied that § 104(a)(2) might have excluded Lindsey’s settlement had he done two things.\footnote{239} First, the court noted the importance of Lindsey’s failure to make the payor aware of his physical injury or physical sickness.\footnote{240} Second, the court stated that Lindsey failed to meet the on account of requirement because he did not demonstrate “what percentage of the settlement damages [was] allocable to physical injury or physical sickness,” and because the record’s evidence failed to do the same.\footnote{241} The Eighth Circuit’s implications provide a blueprint for a taxpayer to exclude physical symptoms of emotional distress if his award is not expressly allocated to nonphysical or emotional injuries.

Without focusing on whether the taxpayer’s action against the defendant had its origin in a physical injury or physical sickness, the circuit courts still refused to apply § 104(a)(2)’s exclusion amidst allegations of physical injuries. This result was

\footnote{236. See id. (noting that Lindsey’s injuries fell “within the definition of emotional distress” and citing the House Report).}

\footnote{237. See id. at 6 (“Even if petitioner had suffered a personal physical injury within the meaning of section 104(a)(2), such injury could not have been the basis for settlement because, as the parties stipulated, petitioner did not communicate any physical injury to [the payor] during the settlement negotiations.”).}

\footnote{238. See Lindsey v. Comm’r, 422 F.3d 684, 689 (8th Cir. 2005) (“We affirm the well-reasoned decision of the Tax Court.”).}

\footnote{239. See id. at 688–89 (implying that § 104(a)(2) might have excluded Lindsey’s award had he made the defendant aware of the injuries and been able to prove that a portion of the settlement was allocated to such injuries).}

\footnote{240. See id. (noting the importance of Lindsey’s failure to communicate his physical injuries to the defendant).}

\footnote{241. Id. at 689.}
possible because the damages at issue were expressly allocated to nonphysical injuries. Because of this allocation, the taxpayers were unable to prove that any portion of their award was intended to compensate their alleged physical injury. Thus, focusing on whether the received damages were “on account of a personal physical injury or sickness” was sufficient to reach the right outcome. As three conflicting Tax Court cases indicate, the correct result is not as readily ascertainable when the damages are not so neatly allocated.242

VI. The Tax Court Reaches Conflicting Results

A. Sanford v. Commissioner

In 2008, the U.S. Tax Court decided Sanford v. Commissioner.243 Sanford held that § 104(a)(2) did not exclude the taxpayer’s damages received from a legal action against her employer for employment discrimination and sexual harassment.244 The taxpayer filed complaints with the U.S. Equal Employment Opportunity Commission (EEOC) against her employer, the U.S. Postal Service (USPS).245 She alleged that the USPS discriminated against her and that “she was retaliated against for previously participating in EEOC activity.”246

242 Compare Sanford v. Comm’r, T.C. Memo. 2008-158, at 1 (2008) (holding that § 104(a)(2) did not exclude the taxpayer’s damages received from a legal action against her employer for employment discrimination and sexual harassment that caused emotional distress manifested by physical symptoms), with Domeny v. Comm’r, T.C. Memo. 2010-9, at 5 (2010) (holding that § 104(a)(2) excluded the taxpayer’s settlement award from her employer where she alleged that the employer’s conduct caused emotional distress manifested by physical symptoms), and Parkinson v. Comm’r, T.C. Memo. 2010-142, at 7 (2010) (holding that § 104(a)(2) excluded the taxpayer’s settlement award for his intentional infliction of emotional distress claim when his emotional distress was manifested by a second heart attack).

243 Sanford v. Comm’r, T.C. Memo. 2008-158, at 1 (holding that § 104(a)(2) did not exclude the taxpayer’s damages received from a legal action against her employer for employment discrimination and sexual harassment that caused emotional distress manifested by physical symptoms).

244 See id.

245 See id. (detailing the taxpayer’s claims against the USPS).

246 Id.
Additionally, she alleged that a USPS coworker sexually harassed her.\footnote{Id.}

The EEOC found the taxpayer was discriminated against because of her sex, and that she was sexually harassed.\footnote{Id.} The USPS Final Agency Decision awarded the taxpayer “compensatory damages of $7,662 in past medical expenses and transportation, $14,033 for past benefits lost (leave without pay), and $12,000 in nonpecuniary compensatory damages.”\footnote{Id.} In 2003, the USPS paid the taxpayer the total damages of $33,695.\footnote{Id.} The taxpayer then appealed the USPS Final Agency Decision to the EEOC.\footnote{See id. at 2 (“Petitioner appealed the $33,695 USPS Final Agency Decision to the EEOC.”).} On appeal, the EEOC “noted that [the taxpayer] had provided sufficient documentation to substantiate or justify her request for additional compensatory damages,” with documentation consisting of statements from her friends, coworkers, and psychologist.\footnote{Id.} The EEOC also noted that the documentation showed the taxpayer “experienced physical symptoms” due to the emotional distress and psychiatric problems that the long-term harassment created.\footnote{Id.} These physical symptoms included “intensification of petitioner’s asthma, sleep deprivation, skin irritation, appetite loss, severe headaches, and depression.”\footnote{Id.} Accordingly, the EEOC modified the USPS Final Agency Decision, determining that “USPS should pay [the taxpayer] a total of $115,000 in nonpecuniary damages, $33,542 in future pecuniary losses, $7,662 for medical expenses, and $14,033 for use of annual leave.”\footnote{Id.} In 2004, the USPS paid the taxpayer the damages.\footnote{Id.}

The taxpayer failed to report any of the damages received in the legal action as income on her 2003 and 2004 tax returns.\footnote{Id.}
and the Service determined a deficiency. 258 The Tax Court considered whether § 104(a)(2) excluded the taxpayer’s received compensation for “nonpecuniary damages and future pecuniary losses.” 259 The court’s analysis began by noting that “gross income is broad in scope, while exclusions from income are narrowly construed.” 260 Then, the court reiterated § 104(a)(2)’s text, specifically stating that “emotional distress is not treated as a personal physical injury or physical sickness.” 261 Because the Service conceded the taxpayer’s underlying cause of action was “based in tort or tort-type rights,” the court proceeded to determine whether the taxpayer’s damages were received on account of her physical injuries or physical sickness. 262

The court concluded that the taxpayer’s nonpecuniary damages and future pecuniary losses “were not received on account of personal physical injury or physical sickness” and thus not excluded from her gross income under § 104(a)(2). 263 Because the taxpayer did not meet the requirements for “medical care” deductions under § 213, the court also concluded the taxpayer could not exclude the portion of the award attributed to medical expenses. 264 When analyzing the nonpecuniary damages and future pecuniary losses portion, the court looked to the EEOC and USPS decisions, which noted that the taxpayer’s sexual harassment “caused her emotional distress.” 265 Further, the court acknowledged that the taxpayer’s “emotional distress manifested

---

258. Id.
259. See id. at 3 (“We now consider whether petitioner must include in income the portion of the award for nonpecuniary damages and future pecuniary losses.”).
260. Id. (citing Comm’r v. Schleier, 515 U.S. 323, 328 (1995)).
261. See id. (“Emotional distress is not treated as a personal physical injury or physical sickness.” (citing I.R.C. § 104(a) (2006))).
262. See id. (noting that the Service conceded that the taxpayers met the “tort or tort-type rights” requirement and proceeding to determine whether the damages were received “on account of” her physical injuries or physical sickness).
263. See id. at 4 (“We conclude that the [damages] awarded to [the taxpayer] as a result of the legal action were not received on account of personal physical injury or physical sickness. [The taxpayer] therefore must include these damages in her income under section 104(a)(2).”).
264. See id. (concluding that § 104(a)(2) did not exclude taxpayer’s portion of the award for “past medical expenses and transportation”).
265. Id. at 3.
itself in [the] physical symptoms” noted in the EEOC decision but found that “[t]hese physical symptoms were not the basis of the award.”266 Instead, the taxpayer was awarded relief for sexual harassment and sex-based discrimination, and was “compensated . . . for the emotional distress she suffered because of the sexual harassment.”267 Importantly, the Tax Court stated that “[d]amages received on account of emotional distress, even when resultant physical symptoms occur, are not excludable from income under section 104(a)(2).”268 This statement is entirely consistent with § 104(a)(2)’s legislative history in which Congress provided that the term “emotional distress” included symptoms.

B. Domeny v. Commissioner

Less than two years later, in Domeny v. Commissioner,269 the Tax Court reached a conclusion directly contrary to Sanford.270 After being terminated, the taxpayer complained that her pre-existing multiple sclerosis (MS) “spik[ed],” and that she suffered “shooting pain up her legs, fatigue, burning eyes, spinning head, vertigo, and lightheadedness.”271 Before filing suit, however, the taxpayer’s attorney negotiated a settlement agreement with her former employer.272 In the settlement agreement, the taxpayer released potential claims under the ADA, the ADEA, and potential claims for invasion of privacy, defamation and misrepresentation, and infliction of emotional distress.273 The

266. Id.
267. Id.
268. Id. (citing Hawkins v. Comm’r, T.C. Memo. 2005-149 (2005)).
269. Domeny v. Comm’r, T.C. Memo. 2010-9, at 5 (2010) (holding that § 104(a)(2) excluded the taxpayer’s settlement award from her employer where she alleged that the employer’s conduct caused emotional distress manifested by physical symptoms).
270. See id. at 5 (concluding § 104(a)(2) excluded the taxpayer’s settlement award because her work environment exacerbated her physical illness).
271. Id. at 2.
272. See id. (noting that a settlement agreement was reached before any suit was filed).
273. See id. (listing the causes of action that the taxpayer released pursuant to the settlement agreement).
The settlement agreement made no mention of the taxpayer's alleged physical symptoms.\textsuperscript{274} The taxpayer's employer paid a total of $33,308 under the agreement.\textsuperscript{275} Of the total, $16,375 was sent directly to the taxpayer's attorney.\textsuperscript{276} The Tax Court addressed the sole question of whether § 104(a)(2) excluded the settlement agreement's remaining $16,933, which was sent directly to the taxpayer and not included as income on her tax return.\textsuperscript{277}

The court began by noting that the taxpayer believed she was being compensated “for physical injuries that occurred in a hostile work environment.”\textsuperscript{278} Next, the court noted § 104(a)(2)’s text, including the post-amble, which states that “emotional distress shall not be treated as physical injury or physical sickness.”\textsuperscript{279} After summarily concluding the settlement agreement was ambiguous, the court proceeded to analyze the payor’s intent.\textsuperscript{280} The court also concluded that the taxpayer’s claim against her employer was based on tort or tort-type rights.\textsuperscript{281} Finally, the court analyzed whether there was a “direct causal link” between the damages and the taxpayer’s alleged physical injuries.\textsuperscript{282}

The taxpayer’s employer did not withhold taxes on the $16,933 at issue and labeled the amount as “[n]onemployee compensation.”\textsuperscript{283} Based on this fact alone, the Tax Court inferred that the taxpayer’s employer “was aware that at least part of [the taxpayer’s] recovery may not have been subject to tax; i.e., was due to physical illness.”\textsuperscript{284} The court then coupled that inference

\begin{itemize}
  \item \textsuperscript{274} See \textit{id.} at 3 (noting that the settlement agreement contained only a list of “numerous possible causes of action that [the taxpayer] was releasing”).
  \item \textsuperscript{275} \textit{id.} at 2.
  \item \textsuperscript{276} \textit{id.} Half of the $16,375 was compensation due to the taxpayer and the taxpayer reported that as income on her 2005 tax return. \textit{id.} The taxpayer was not issued any tax forms on the other half sent to her attorney and did not report it as income. \textit{id.}
  \item \textsuperscript{277} \textit{id.} at 3.
  \item \textsuperscript{278} \textit{id.}
  \item \textsuperscript{279} \textit{id.}
  \item \textsuperscript{280} \textit{id.}
  \item \textsuperscript{281} \textit{id.}
  \item \textsuperscript{282} \textit{id.} (citing \textit{Banaitis v. Comm’r}, 340 F.3d 1074, 1080 (9th Cir. 2003)).
  \item \textsuperscript{283} \textit{id.} at 4.
  \item \textsuperscript{284} \textit{id.}
\end{itemize}
with the “likelihood that [the taxpayer’s] attorney represented [the taxpayer’s] circumstances to [her employer] in the course of the settlement negotiations.”285 Based solely on these bare inferences—and ignoring the settlement agreement’s terms—the Tax Court found that the taxpayer’s employer intended to compensate her for “her acute physical illness caused by her hostile and stressful work environment.”286 Thus, because the taxpayer had “shown that her work environment exacerbated her existing physical illness,” she had “shown that the only reason for the $16,933 payment was to compensate her for her physical injuries.”287

_Domeny’s_ conclusion cannot be squared with _Sanford_. In _Sanford_, the EEOC and the court explicitly noted that the taxpayer’s emotional distress caused physical symptoms, including intensifying her preexisting asthma.288 In _Domeny_, the taxpayer’s emotional distress caused physical symptoms, including intensifying her preexisting MS.289 Yet § 104(a)(2) was applied to the award in _Domeny_ and not in _Sanford_.290 A Tax Court case decided later in 2010 is further indicative of this inconsistent application.

C. Parkinson v. Commissioner

In _Parkinson v. Commissioner_,291 the taxpayer asserted claims of intentional infliction of emotional distress and invasion of privacy against his former employer (medical center) and two

---

285. *Id.*
286. _Id._
287. _Id._ at 5.
288. See _Sanford v. Comm’r_, T.C. Memo. 2008-158, at 3 (2008) (“We further acknowledge, as did the EEOC, that the emotional distress manifested itself in physical symptoms such as asthma.”).
289. See _Domeny v. Comm’r_, T.C. Memo. 2010-9, at 2 (2010) (noting that the taxpayer suffered physical symptoms of her emotional distress, including exacerbating her preexisting MS).
290. See _supra_ Parts VI.A–B (discussing _Sanford_ and _Domeny_).
291. See _Parkinson v. Comm’r_, T.C. Memo. 2010-142, at 7 (2010) (holding that § 104(a)(2) excluded the taxpayer’s settlement award for his intentional infliction of emotional distress claim when his emotional distress was manifested by a second heart attack).
named coworkers. The taxpayer’s complaint alleged that the two named coworkers “harassed and harangued” him, causing him to suffer “severe emotional distress, manifested by permanent, irreparable physical harm in the form of his second heart attack.” For both claims, the “complaint sought $500,000 in compensatory damages, $500,000 in punitive damages, attorney’s fees, and costs.” The day after a jury trial began, the “medical center agree[d] to pay [the taxpayer] $350,000 ‘as noneconomic damages and not as wages or other income,’” and the taxpayer agreed to drop all claims. In 2005, the taxpayer received a $34,000 payment pursuant to the settlement agreement but failed to report this amount on his 2005 federal income tax return.

After considering the settlement agreement’s terms and the payor’s intent, the court first concluded that “the entire settlement payment [was allocable to] [the taxpayer’s] cause of action for intentional infliction of emotional distress.” Then, the Tax Court concluded that § 104(a)(2) excluded one-half of the taxpayer’s 2005 settlement payment because it “was made on account of [the taxpayer’s] physical injuries.” In support of its conclusion, the court looked to the 1996 Act’s House Report.

First, the Tax Court acknowledged the House Report provided that the term “emotional distress” included symptoms. Next, the court quoted the portion of the House Report that stated that “[b]ecause all damages received on account of physical injury or physical sickness are excludable from gross income, the exclusion . . . applies to any damages received based on a claim of emotional distress that is

---

292. See id. at 1–2 (detailing the taxpayer’s complaint).
293. Id. at 2.
294. Id.
295. Id.
296. See id. (“Pursuant to the settlement agreement [the taxpayer] agreed to drop all his claims.”).
297. Id.
298. Id.
299. See id. at 6 (finding one-half of the settlement amount excluded).
300. Id. at 5.
301. Id.
PHYSICAL CONSEQUENCES

attributable to a physical injury or physical sickness.” The court stated that this language, with respect to a claim for emotional distress, indicated Congress’s intent to distinguish “damages attributable to physical injury or physical sickness, which are excludable, from damages attributable to emotional distress or ‘symptoms’ thereof, which are not excludable.” The court took the House Report’s language out of context and read it incorrectly.

Parkinson dropped the origin of the action portion of the House Report, focusing instead only on the end result. The House Report provided that damages for emotional distress qualified for § 104(a)(2)’s exclusion only when a personal physical injury or physical sickness caused emotional distress. In other words, when a taxpayer’s personal physical injury or physical sickness causes the taxpayer to suffer emotional distress, then the damages allocated to such emotional distress will be on account of the personal physical injury or physical sickness. As Parkinson reads the House Report, damages allocated to a physical symptom of emotional distress would fall within § 104(a)(2)’s exclusion because they are on account of the physical symptom. Under that reading, damages allocated to emotional distress caused by a physical injury or physical sickness would be on account of emotional distress. This is directly contrary to the statute’s text and its legislative history. Section 104(a) explicitly states that damages on account of emotional distress should be included in income. Congress intended for damages allocable to emotional distress caused by a physical injury to fall within § 104(a)(2)’s exclusion only because all of those damages have their “origin in a physical injury or physical sickness.”

302. Id.
303. Id.
304. See HOUSE REPORT, supra note 163, at 301 (providing that § 104(a)(2) excluded damages allocated to emotional distress “on account of” a physical injury or physical sickness).
305. See I.R.C. § 104(a) (2006) (“[E]motional distress shall not be treated as a physical injury or physical sickness.”).
306. See HOUSE REPORT, supra note 163, at 301 (providing that § 104(a)(2) excluded damages allocated to emotional distress “on account of” a physical injury or physical sickness).
Next, the Tax Court stated it was “self-evident that a heart attack and its physical aftereffects constitute physical injury or sickness rather than mere subjective sensations or symptoms of emotional distress.” A heart attack is undoubtedly a “severe” physical symptom, but it is a “symptom” nonetheless. In fact, the taxpayer stated in his complaint that he “suffered severe emotional distress, manifest by permanent, irreparable physical harm in the form of his second heart attack.” Because the taxpayer’s emotional distress manifested itself with such a severe physical symptom, he received a large sum as settlement for his claim. This does not mean, however, that the taxpayer’s larger sum should not be taxed, while someone suffering a “milder” physical symptom—and consequently receiving a smaller settlement—should be taxed.

Lastly, the Tax Court relied on treatise excerpts, stating that when a plaintiff is compensated for emotional distress that is evidenced by physical symptoms, both the mental and physical elements have been compensated. The Tax Court is correct that the medical center compensated the taxpayer for both elements of his emotional distress. But simply because both the mental and the physical elements have been compensated does not mean each element should be taxed separately rather than taxed as one recovery for emotional distress. Under Parkinson’s reasoning, every taxpayer suffering a physical symptom of emotional distress would be entitled to exclude a court-determined amount of any potential recovery under § 104(a)(2). The analytical difficulties of drawing a line between the emotional and the physical elements would make judicial determinations, at best, impossible.

---

308. Id. at 2.
309. The taxpayer received $350,000. Id.
310. See, e.g., Murphy v. Internal Revenue Serv., 493 F.3d 170, 176 (D.C. Cir. 2007) (concluding that § 104(a)(2) did not exclude the taxpayer’s $70,000 award for emotional injuries manifested by physical symptoms, such as “bruxism”).
311. See Parkinson, T.C. Memo. 2010-142, at 5–6 (stating that, when a plaintiff is compensated for emotional distress accompanied by physical symptoms, both the physical and the mental elements have been compensated (citing the RESTATEMENT (SECOND) OF TORTS § 46(1)) (1965); W. PAGE KEETON ET AL., PROSSER & KEETON ON THE LAW OF TORTS, § 12, at 64 (5th ed. 1984)).
wildly speculative. As a result, the tax consequences of emotional distress recoveries would be entirely unpredictable.

Moreover, when Congress stated that the term emotional distress includes symptoms that may result from such emotional distress, it indicated that § 104(a)(2)’s exclusion does not consider emotional distress’s physical symptoms as a stand-alone physical injury or physical sickness. Instead, a recovery for “emotional distress” encompasses its physical symptoms, no matter their severity. As the Tax Court previously stated, “[d]amages received on account of emotional distress, even when resultant physical symptoms occur, are not excludable from income under section 104(a)(2).”

VII. Proposal: A New Test

Before the 1996 Act, the “origin” of the “personal injury” giving rise to damages was a nonissue for a few reasons. First, Burke’s “tort or tort-type rights” test ensured that § 104(a)(2) excluded only damages compensating for personal-injury torts, or for personal injuries where the full range of tort-type remedies were available. The test intended to “distinguish damages for personal injuries from, for example, damages for breach of contract.” Second, § 104(a)(2) excluded all personal injuries, physical, nonphysical, and emotional.
After the 1996 Act, on the other hand, Treasury amended its § 104(a)(2) regulations that served as the basis for Burke’s requirement, thus eliminating Burke’s tort or tort-type rights test.317 In the amendment’s proposal, Treasury stated that judicial and legislative developments—including the 1996 Act—have “eliminated the need to base the section 104(a)(2) exclusion on tort and remedies concepts.”318 The amended regulations extend § 104(a)(2)’s exclusion to “personal physical injuries or physical sickness not defined as torts under state and common law,” and to no-fault statutes.319 Section 104(a)(2)’s exclusion no longer “depend[s] on the scope of remedies available under state or common law.”320 Replacing the tort or tort-type rights requirement, the amended regulations define § 104(a)(2) “damages” as “an amount received (other than workers’ compensation) through prosecution of a legal suit or action, or through a settlement agreement entered into in lieu of prosecution.”321

Additionally, § 104(a)(2) now excludes only damages received “on account of personal physical injuries or physical sickness.”322 Because of the added “physical” requirement, defendants can be liable for “personal injury” compensatory damages that § 104(a)(2) no longer excludes. For example, a defendant’s conduct may directly cause a taxpayer to suffer emotional distress. Before the 1996 Act, the taxpayer’s emotional distress would have qualified as a “personal injury” for purposes of § 104(a)(2)’s exclusion. If the defendant then paid the taxpayer damages on account of such emotional distress, § 104(a)(2) would exclude those damages as being on account of a personal injury. Now, however, because § 104(a) prohibits emotional distress from being treated as a physical injury or physical sickness, § 104(a)(2)

that it was now settled that it extended to “nonphysical injuries to the individual”).

317. See Treas. Reg. § 1.104-1(c) (as amended in 2012) (eliminating the “tort or tort type rights” requirement).
318. Proposed Regulation, supra note 314, at 47,153.
319. Id.
320. Id.
321. Treas. Reg. § 1.104-1(c) (as amended in 2012).
does not exclude damages received on account of the “personal injury” of emotional distress. Thus, defendants can be liable for “personal injury” compensatory damages—damages that would have been excluded before the 1996 Act—without directly causing a “physical injury or physical sickness.”

This creates the problem that was illustrated in *Domeny* and in *Parkinson*. According to *Domeny* and *Parkinson*, damages allocated to a physical symptom of emotional distress may be excluded if the taxpayer could show that the damages compensated for, or were “on account of,” the physical symptom.323 In *Lindsey*, the Eighth Circuit implied that to do so, a taxpayer simply has to alert the defendant to the taxpayer's physical symptom and then prove that at least a portion of the award is intended to compensate for the physical symptom.324 This result directly contradicts the exclusion’s text, however, which states that “emotional distress shall not be treated as a physical injury or physical sickness” for purposes of § 104(a)(2)’s exclusion,325 and the House Report, which states that “emotional distress” includes resulting symptoms.326 To remedy this contradiction, it is necessary to focus on whether the taxpayer’s action against the defendant has its *origin* in a physical injury or physical sickness.

As Treasury noted when it proposed amending its regulations, judicial and legislative developments have affected § 104(a)(2)’s exclusion.327 Most importantly, the 1996 Act added the requirement that a taxpayer’s “personal injury or sickness” must be “physical.”328 To accommodate that added requirement, a new test needs to be implemented to prevent results—such as

323. *See supra* Parts VI.B–C (discussing *Domeny* and *Parkinson*).
324. *Supra* notes 239–41 and accompanying text.
325. I.R.C. § 104(a).
326. *See House Report, supra* note 163, at 301 n.56 (stating that the term emotional distress includes symptoms that may result from such emotional distress).
327. *See Proposed Regulation, supra* note 314, at 47, 152–53 (noting that judicial and legislative developments have affected § 104(a)(2)’s exclusion that Treasury’s proposed regulations reflect the statutory developments).
Domeny and Parkinson—that contradict § 104(a)(2)’s text and legislative history.

Under the new test, to be excluded from gross income under § 104(a)(2): (1) an amount must be received (other than workers’ compensation) through prosecution of a legal suit or action, or through a settlement agreement entered into in lieu of prosecution; (2) such legal suit or action must have its origin in a physical injury or physical sickness; and (3) the amount must be received on account of such physical injury or physical sickness.

The test’s three requirements must each be satisfied independently. The first requirement is taken from § 104(a)(2)’s text and Treasury’s proposed regulations. As Treasury intended, it replaces the tort or tort-type rights requirement and defines § 104(a)(2) “damages.” The requirement functions as Treasury detailed in its proposed regulations above329 and thus needs no additional explanation here. The test’s third requirement derives from § 104(a)(2)’s on account of text and incorporates all the case law interpreting that language. The cases interpreting on account of are detailed earlier in Part III of this Note.330 Accordingly, the third requirement also requires no further explanation here. The test’s second requirement derives from § 104(a)(2)’s text combined with Congress’s language in the House Report. The test’s second requirement works in unity with the first and needs additional explanation.

A. The “Origin” Requirement

When a statute is ambiguous on its face—as § 104(a)(2)’s language has always been331—courts should view the legislative history to determine Congress’s intent.332 Appropriately, then, the test’s second requirement is drawn from the House Report that

329. Supra notes 318–21 and accompanying text.
330. See supra Part III (discussing the case law refining § 104(a)(2)’s “on account of” text).
331. See O’Gilvie v. United States, 519 U.S. 79, 79 (1996) (acknowledging that the “on account of” phrase in § 104(a)(2) is ambiguous).
332. See Middlesex Cnty. Sewerage Auth. v. Nat’l Sea Clammers Ass’n, 453 U.S. 1, 13 (1981) (stating that the Court first looks to the statutory text and then reviews the legislative history to determine congressional intent).
gave guidance on how to apply the amended § 104(a)(2) exclusion.\footnote{See House Report, supra note 163, at 301 (providing the “origin” requirement).} According to the House Report, to qualify for § 104(a)(2)'s exclusion, the taxpayer's “action” against the defendant must have its “origin in a physical injury or physical sickness.”\footnote{See id. ("If an action has its origin in a physical injury or physical sickness, then all damages (other than punitive damages) that flow therefrom are treated as payments received on account of physical injury or physical sickness . . ."); see also Wood, supra note 167, at 5 (“Congress require[s] that the action have its origin in a physical injury or sickness.").} This requirement, combined with the regulations' use of “legal suit or action,” forms the test's second component.

The word “origin” is commonly defined as “the point at which something begins or rises” or as “something that creates, causes, or gives rise to another.”\footnote{Origin Definition, MERRIAM-WEBSTER.COM, http://merriam-webster.com/dictionary/origin (last visited Nov. 14, 2012) (defining origin) (on file with the Washington and Lee Law Review).} Thus, to qualify for § 104(a)(2)'s exclusion, a physical injury or physical sickness must create, cause, or give rise to the taxpayer's “legal suit or action” against the defendant. Except in the context of a wrongful death or survival suit, this origin requirement will be satisfied only when the defendant's conduct directly causes the taxpayer's personal physical injury or physical sickness, thus giving rise to the taxpayer's action against the defendant. This requires that there be a “direct causal link” between the defendant's conduct and the taxpayer's personal physical injury or physical sickness. For example, in Schleier's automobile accident illustration, the taxpayer's personal physical injuries suffered in the automobile accident gave rise to the taxpayer's negligence action against the defendant.\footnote{See supra note 101 and accompanying text (using an automobile accident example to illustrate a personal injury).} Without an intervening emotional injury, the defendant's negligent conduct directly caused the taxpayer's personal physical injuries, including bruises, cuts, or broken bones. Because the defendant directly caused the taxpayer's personal physical injuries, the defendant was liable for damages on account of the taxpayer's personal physical injuries. Thus, the taxpayer's “legal suit or action” against the defendant—that may have resulted in other claims such as lost wages, or
pain, suffering, and emotional distress—had its origin in a physical injury or physical sickness.

In the context of a wrongful death or survival suit, the taxpayer's action against the defendant has its origin in a physical injury or physical sickness because the taxpayer's action originated out of the decedent's death, which the defendant directly caused. Because the taxpayer could not have prosecuted the action against the defendant absent the decedent's personal physical injury of death, the decedent's death gave rise to the spouse's action. Therefore, the decedent's death was the origin of the spouse's wrongful death or survival action against the defendant. The taxpayer's action may then include claims for loss of consortium and pain, suffering, and emotional distress, but the action's origin was in a physical injury or physical sickness.337

On the other hand, the origin requirement will not be satisfied when the defendant's conduct directly causes the taxpayer to suffer emotional distress manifested by a physical symptom, even if the physical symptom alone could be considered a physical injury or physical sickness. With employment discrimination or wrongful termination, the defendant's conduct directly causes lost wages and nonphysical personal injuries, such as emotional distress, but does not directly cause a physical injury or physical sickness.338 Because the defendant's conduct only directly causes the taxpayer's emotional distress, there is no direct causal link between the defendant's conduct and the taxpayer's physical symptom. The taxpayer can still bring an action against the defendant for infliction of emotional distress and lost wages, but the action would not have its origin in a physical injury or physical sickness. Thus, the taxpayer would fail to meet the test's second requirement.

In *Parkinson*, for example, the defendant's conduct directly caused the taxpayer's emotional distress.339 The taxpayer's emotional distress then manifested itself with the severe physical

---

337. *Cf.* House Report, *supra* note 163, at 301 (noting that a taxpayer's claim for loss of consortium due to his spouse's physical injury or physical sickness would be excluded from gross income).

338. See *supra* note 101 and accompanying text (using an automobile accident example to illustrate a personal injury).

339. See *supra* Part VI.C (discussing Parkinson v. Comm'r, T.C. Memo. 2010-142 (2010)).
symptom of a heart attack. The taxpayer’s emotional distress severed the direct causal link between the defendant’s conduct and the physical symptom. Thus, the taxpayer’s heart attack was not the origin of his action against the defendant. Rather, the taxpayer’s severe emotional distress was the origin of his action against the defendant and the emotional distress then gave rise to his physical injury.

Importantly, the immediacy of the physical symptom’s onset does not affect this result. Consider a taxpayer-employee that is deathly afraid of snakes. If her employer places a rubber snake in her office and she immediately suffers a heart attack upon seeing the snake, her heart attack will not be the origin of her action against her employer. Instead, her cause of action for infliction of emotional distress will have its origin in her severe emotional distress, which was manifested by a heart attack. These consequential physical symptoms may, of course, indicate the severity of the defendant-inflicted emotional distress and thus increase the total damages sum. But this does not mean that a taxpayer suffering physical consequences of emotional distress should be better off from a tax perspective than a taxpayer whose emotional distress does not have physical consequences.

Admittedly, there can be confusion when the defendant simultaneously inflicts separate emotional and physical injuries. This scenario is best illustrated in the sexual harassment and false imprisonment contexts. As the Service explained in a private letter ruling in 2000, when sexual harassment takes the form of an “unwanted or uninvited physical contact[] resulting in observable bodily harms,” then § 104(a)(2) excludes damages allocable to such injuries. But when the sexual harassment is

---

340. This example was provided by Professor Brant J. Hellwig.
341. If the employee died as a result of her heart attack and her husband brought a wrongful death action against her employer, then the husband’s action would have its origin in the physical injury of death. Thus, § 104(a)(2) would exclude damages that he received “on account of” his wife’s death.
342. Cf. Murphy v. Internal Revenue Serv., 493 F.3d 170, 176 (D.C. Cir. 2007) (noting that the Administrative Law Judge may have considered the taxpayer’s physical injuries “indicative of the severity of [her] emotional distress”).
343. See I.R.S. Priv. Ltr. Rul. 200041022 (Oct. 13, 2000) (providing that § 104(a)(2) would exclude damages allocated to “unwanted or uninvited physical
nonphysical and causes the taxpayer to suffer emotional distress. § 104(a)(2) does not exclude any damages attributable to the emotional distress, even if accompanied by physical symptoms.\(^{344}\) Similarly, as the Sixth Circuit explained in Stadnyk, false imprisonment does not necessarily involve a physical injury.\(^{345}\) Presumably, if the false imprisonment took the form of an “unwanted or uninvited physical contact[] resulting in observable bodily harms,” then § 104(a)(2) would exclude damages attributable to such injuries.\(^{346}\) But § 104(a)(2) would not exclude damages attributable to emotional distress resulting from the false imprisonment’s nonphysical aspects, such as confinement alone.\(^{347}\)

In these scenarios, the physical injury or physical sickness does not have to be the action’s sole origin. Because the defendant directly caused the taxpayer’s personal physical injury, and that personal physical injury is one of the origins of the taxpayer’s action against the defendant, the test’s second requirement would be met. A court would then have to determine what portion of the damages is on account of the taxpayer’s personal physical injuries, and thus excluded by § 104(a)(2). Section 104(a)(2) would not, however, exclude the portion of damages flowing from, or directly caused by, the taxpayer’s emotional distress, except in the amount not exceeding qualifying medical care under § 213.

\(^{344}\) See id. (providing that § 104(a)(2) would exclude only damages allocated to “unwanted or uninvited physical contacts resulting in observable bodily harms”); see also Sanford v. Comm’r, T.C. Memo. 2008-158, at 1 (2008) (holding that § 104(a)(2) did not exclude the taxpayer’s damages received from a legal action against her employer for sexual harassment that caused emotional distress manifested by physical symptoms).

\(^{345}\) See Stadnyk v. Comm’r, 367 F. App’x 586, 593 (6th Cir. 2010) (determining that a false imprisonment victim is not necessarily physically injured).


\(^{347}\) Cf. Stadnyk, 367 F. App’x at 594 (holding that the taxpayer’s award for false imprisonment claims could not be excluded from gross income under § 104(a)(2) because there was no “causal connection between any physical injury and the settlement award”).
VIII. Conclusion

Of course, it still remains difficult to draw the line between physical and nonphysical. Some commentators have suggested that § 104(a)(2)'s exclusion be repealed due to the difficulty with drawing such a line. Others have suggested rewriting the exclusion to apply only to damages attributable to lost human capital, or that the Service adopt definitions from the Restatement (Third) of Torts. No matter where the line is drawn between physical and nonphysical, however, some taxpayers will be better off than others.

Although Congress did not explicitly define physical when it amended § 104(a)(2), it stated both in the statutory text and the House Report that “emotional distress shall not be treated as a physical injury or physical sickness” for purposes of the exclusion. Importantly, the House Report provided that the term emotional distress includes symptoms that may result from such emotional distress. By providing that emotional distress and its resulting symptoms are not to be treated as a physical injury or physical sickness, Congress excluded emotional distress from its definition of physical. Section 104(a)(2) does not exclude a taxpayer’s recovery for emotional distress, unless an originating physical injury or physical sickness causes the taxpayer’s emotional distress. To prevent the inconsistency of having § 104(a)(2)'s gross income exclusion extended to some physical symptoms of emotional distress, but not others, it is necessary to focus on whether the taxpayer’s action has its origin in a physical injury or physical sickness.

348. See Barret, supra note 13, at 1194 (“In order to promote predictability and consistency, the exclusion should be eliminated.”).

349. See Frank J. Doti, Personal Injury Income Tax Exclusion: An Analysis and Update, 75 DENVER U. L. REV. 61, 62 (1997) (proposing that § 104(a)(2) should be redrafted and only encompass “damages attributable to lost human capital and not lost wages and earning power”).

350. See Wright, supra note 7, at 215 (proposing that the Restatement (Third) of Torts’ definition be adopted).

351. See HOUSE REPORT, supra note 163, at 300–01 (failing to define “physical injury or physical sickness”); see also I.R.S. Priv. Ltr. Rul. 200041022 (Oct. 13, 2000) (“The term ‘personal physical injuries’ is not defined in either § 104(a)(2) or the legislative history of the 1996 Act.”).


353. See HOUSE REPORT, supra note 163, at 301 n.56 (stating that the term “emotional distress” includes symptoms).