ARTICLES
Justice in the Shadowlands: Pretrial Detention, Punishment, & the Sixth Amendment
Laura I. Appleman
Beyond War: Bin Laden, Escobar, and the Justification of Targeted Killing
Luis E. Chiesa
Alexander K.A. Greenawalt
The CARD Act on Campus
Jim Hawkins
Congress, the Constitution, and Supreme Court Recusal
Louis J. Virelli III

NOTES
“Waiving” Goodbye to Arbitration: A Contractual Approach
Paul Bennett IV
FATCA: Toward a Multilateral Automatic Information Reporting Regime
Joanna Heiberg
The Virtual Water Cooler and the NLRB: Concerted Activity in the Age of Facebook
Lauren K. Neal
Do End-Users Get the Best of Both Worlds?—Title VII of Dodd–Frank and the End-User Exception
Carney Simpson
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. McTENIA, JR., A.B., M.A., LL.B.
James P. Morefield Professor of Law, Emeritus

JAMES M. PEMISTER, 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. MOLLERTNO, B.S., J.D.
Vincent Bradford Professor of Law

BRNAN 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-RAFAEL, L.L.B., LL.M., S.J.D.
Visiting Assistant Professor of Law

DONALD EARL CHILDESS 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 MUHR, 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.

Forthcoming Symposium

Gideon at 50: Reassessing the Right to Council
(November 8–9, 2012)

Previous Major Symposia

Regulation in the Fringe Economy
(November 10–11, 2011)

Restitution Rollout: The Restatement (Third) of Restitution & Unjust Enrichment
(February 25, 2011)

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

Protecting Virtual Playgrounds: Children, Law, and Play Online
(October 3, 2008)

The Restitution Roundtable
(December 14, 2007)

Gender-Relevant Legislative Change in Muslim and Non-Muslim Countries
(March 30, 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.

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.
ARTICLES

Justice in the Shadowlands: Pretrial Detention, Punishment, & the Sixth Amendment
Laura I. Appleman

Beyond War: Bin Laden, Escobar, and the Justification of Targeted Killing
Luis E. Chiesa
Alexander K.A. Greenawalt

The CARD Act on Campus
Jim Hawkins

Congress, the Constitution, and Supreme Court Recusal
Louis J. Virelli III

NOTES

“Waiving” Goodbye to Arbitration: A Contractual Approach
Paul Bennett IV

FATCA: Toward a Multilateral Automatic Information Reporting Regime
Joanna Heiberg

The Virtual Water Cooler and the NLRB: Concerted Activity in the Age of Facebook
Lauren K. Neal
Do End-Users Get the Best of Both Worlds?—Title VII of Dodd–Frank and the End-User Exception

Carney Simpson
Justice in the Shadowlands: Pretrial Detention, Punishment, & the Sixth Amendment

Laura I. Appleman*

Abstract

In a criminal system that tips heavily to the side of wealth and power, we routinely detain the accused in often horrifying conditions, confined in jails while still maintaining the presumption of innocence. Here, in the rotting jail cells of impoverished defendants, lies the Shadowlands of Justice, where the lack of criminal procedure has produced a darkness unrelieved by much scrutiny or concern on the part of the law.

This Article contends that our current system of pretrial detention lies in shambles, routinely incarcerating the accused in horrifying conditions often far worse than those of convicted offenders in prisons. Due to these punitive conditions of incarceration, pretrial detainees appear to have a cognizable claim for the denial of their Sixth Amendment jury trial right, which, at its broadest, forbids punishment for any crime unless a cross-section of the offender’s community adjudicates his crime and finds him guilty. This Article argues that the spirit of the Sixth Amendment jury trial right might apply to many pretrial detainees, due to both the punishment-like conditions of their incarceration and the unfair procedures surrounding bail grants, denials and revocations. In so arguing, I expose some of the worst

* Associate Professor of Law, Willamette University. J.D., Yale University, B.A., M.A., English, University of Pennsylvania. Many thanks to W. David Ball, Shima Baradaran, Doug Berman, Stephanos Bibas, Jack Chin, Caroline Davidson, Cara Drinian, Roger Fairfax, David Friedman, Carissa Hessick, Arnold Loewy, Dan Markel, Caren Morrison, Judith Resnik, Giovanna Shay, Anne Traum, Rose Villazor, Norman Williams, and the participants at the 2011 ABA Fourth Annual Sentencing & Reentry Institute and Legal Educators Colloquium and 2012 Southwest Criminal Law Conference for comments on earlier drafts. Thanks also to Willamette for research support.
abuses of current procedures surrounding bail and jail in both federal and state systems. Additionally, I propose some much needed reforms in the pretrial release world, including better oversight of the surety bond system, reducing prison overcrowding by increasing electronic bail surveillance, and revising the bail hearing procedure to permit a community “bail jury” to help decide the defendant’s danger to the community.

Table of Contents

I. Introduction: Diplomats, Detention and Punishment..1299

II. Pretrial Detention as Punishment.................................1304
   A. Bail Bondsmen, Bounty Hunters, and Corrupt Incentives..........................1305
   B. Increased Numbers of Poor Indicted Offenders Denied Bail..........................1310
   C. The Punishing Conditions of Pretrial Detention ..................................1312
   D. Pretrial Incarcерative Harm ......................................................1318
   E. Punishment Before Conviction Violates the Spirit of the Sixth Amendment Jury Trial Right........1321

III. A Short History of Bailing and Jailing ..........................1323
   A. Colonial Practices .................................................................1324
   B. Bail Following the Constitution ...............................................1326
   C. Recent Bail Reforms ...............................................................1329

IV. Preventative Detention, Future Dangerousness, and the Sixth Amendment ..........................1336
   A. The False Promise of Preventative Detention ..................................1336
      1. Barefoot v. Estelle .................................................................1337
      2. Preventative Detention’s Binary Nature ..................................1338
      3. Adjudicating Dangerousness ..................................................1340
      4. Empirical Evidence and National Trends ................................1344
   B. Salerno, the 1984 Act, and the Sixth Amendment ..................................1349
      1. Salerno Did Not Close the Door ..............................................1349
      2. Pretrial Detention as Punishment Under the 1984 BRA ..................1350
JUSTICE IN THE SHADOWLANDS 1299

3. Applying the Sixth Amendment to the BRA .............................................................. 1351
C. Punishment, Community Rights, and Pretrial Detention ................................................................. 1353
1. Future Dangerousness in Adversarial Context .............................................................................. 1353
2. Post-Blakely, Community as Only Arbiter of Punishment .................................................... 1355
3. The Community Should Decide Danger to Itself ....................................................................... 1356
4. Laymen and Members of Community Are Better Predictors of Danger ................................... 1358
5. Attacking the Problems of Race & Gender in Pretrial Detention .............................................. 1359

V. Proposal: Reform and Revision ............................................................. 1361
A. Reforming Bail Bondsmen and Pretrial Release ........................................................................... 1361
B. Revising the Bail Hearing Procedure .......................................................................................... 1363
C. Potential Problems ...................................................................................................................... 1366
1. Lack of Historical Precedent ....................................................................................................... 1366
2. Impracticality/Cost/Delays .......................................................................................................... 1367

VI. Conclusion ............................................................................................................................... 1368

I. Introduction: Diplomats, Detention, and Punishment

Notwithstanding crime, the decision to imprison a defendant before trial all too often hinges on wealth and power. For example, a promiscuous foreign diplomat is halted at the airport, ready to flee the country, after allegedly sexually assaulting a hotel chambermaid, and initially denied bail, but then is permitted to reside in a posh penthouse while electronically monitored, serving an extremely upscale version of “house arrest.” A lifestyle maven charged with perpetuating insider trading pleads not guilty and is released without bail, along with her stockbroker. A well-known money manager, accused of

2. Richard Esposito, Betsy Stark & Ramona Schindelheim, Martha
running a $50 billion Ponzi scheme, is permitted basic freedom of movement within several states while awaiting trial, even after confessing to the crime and failing to live up to his original $10 million bail terms. A prominent governor is charged with serious corruption and is not only released on minimal bail, but is even allowed to take part in a Donald Trump reality show while charges are still pending. A wealthy couple charged with enslaving and brutally mistreating two young maids—including starving, beating, and torturing—are permitted pretrial release with electronic monitoring, due in part to their ability to afford a specialized security firm that functions as private bail guards for the very wealthy. And a well-known alleged Mafia boss, charged with various racketeering charges (and suspected of inducing a variety of violent crimes as acting boss), is released on a $10 million bail, an oath to wear an electronic bracelet, and a guard at his Oyster Bay, Long Island mansion.

---

7. Alan Feuer, Bail Sitters, N.Y. TIMES, Dec. 27, 2009, at MB1. These kinds of “bail sitting” jobs often require bullet-proof vests, electronic ankle bracelet monitoring, and “the deployment of a chase car, a digital voice recorder, a broadband wireless router, several metal door bars and a high-resolution, vandal-resistant Nuvico day/night camera—the one with the plastic dome and manual zoom lens.” Id.
In contrast, a New Jersey barber is pulled over and arrested for a backlog of unpaid parking tickets and failing to register his new car. After his arrest, unable to make his $1000 bail, he is sent to serve his pretrial detention at a local halfway house, where he is robbed and murdered by three inmates for the three dollars in his pockets. His predicament is all too common. The average defendant in pretrial detention has either committed a minor crime and cannot afford to pay the set amount of bail, or has somehow triggered a preventative detention hold—despite the fact that the science of predicting dangerousness can be dubious. Incarcerated, this impoverished defendant has little ability to contact an attorney or plan a defense. And this impoverished defendant is captive to a justice system that regularly allows commercial bail bondsmen to lobby against pretrial release based on inexpensive electronic monitoring, simply to increase their profits.

Although most convicted offenders are incarcerated at state or federal prisons, detainees are typically housed in local or municipal jails where “resources are scarcer, the staff is ‘less professionalized,’ classification of inmates is haphazard, and rapid turnover makes for generally chaotic conditions.” Once the average, nonprivileged, indicted defendant is detained, he is subject to all sorts of punitive conditions, as the state of many halfway houses and metropolitan and rural jails are truly reprehensible, even when measured against prisons. Frequently, these detention centers are vastly overcrowded. Abuse and even murder of pretrial detainees, either by guards or other prisoners, is endemic. Various infections and serious illnesses all too often rage unabated in local and county jails, with minimal health services provided because of the transient nature of the population. Often, not only are the jails themselves older and decaying, but they also have various

11. See AMANDA PETTERUTI & NASTASSIA WALSH, JUSTICE POLICY INSTITUTE, JAILING COMMUNITIES: THE IMPACT OF JAIL EXPANSION AND EFFECTIVE PUBLIC SAFETY STRATEGIES 15 (2008) (citing studies from the 1990s that show that 700 jails in the United States are older than fifty years old and 140 jails are
dangers associated with them, including mold, poor ventilation, lead pipes, and asbestos.\(^\text{12}\)

In a criminal system that tips heavily to the side of wealth and power, routinely detaining the accused in often horrifying conditions, justice is frequently nowhere to be found. Here, in the rotting jail cells of impoverished defendants—still innocent before proven guilty—are the Shadowlands of Justice: the murky corners of the criminal justice system, where the lack of criminal procedure has produced a darkness unrelieved by much scrutiny or concern on the part of the law.

Our current framework of constitutional criminal procedure has primarily focused on the treatment of offenders once the trial or plea proceeding has begun and, to a lesser extent, once these offenders have been convicted and sent to prison. But until very recently, little attention has been paid to the plight of those pretrial defendants languishing in the intermediate world of jails. In part, this is due to the classification of a pretrial offender’s treatment as “detention,” as opposed to “punishment.” As I will argue, however, the conditions of most pretrial detention differ little from punitive incarceration, subjecting these offenders to the worst of conditions without even a guilty verdict.

As such, these pretrial detainees would appear to have a cognizable claim for the denial of their Sixth Amendment jury trial right, which, at its broadest, forbids punishment for any crime unless a cross-section of the offender’s community adjudicates his crime and finds him guilty. This Article explores how the animating principles of the Sixth Amendment community jury trial right would apply to defendants who are held under pretrial detention. In doing so, I look specifically at the procedures surrounding indicted offenders who are denied bail and confined in jail.

In *Blakely v. Washington*,\(^\text{13}\) the Supreme Court clarified that a jury must determine any imposition of punishment.\(^\text{14}\) The

\(^{12}\) Id.

\(^{13}\) See *Blakely v. Washington*, 542 U.S. 296, 305 (2004) (holding that a state trial court’s sentence of more than three years above the statutory maximum “did not comply with the Sixth Amendment”).

\(^{14}\) See id. at 304 (“When a judge inflicts punishment that the jury’s verdict alone does not allow, the jury has not found all the facts ‘which the law makes
JUSTICE IN THE SHADOWLANDS

realities of bail and jail in today’s criminal justice system, however, dictate that punishment is often imposed by nonjury, nonjudicial, and occasionally, private actors, such as bail bondsmen, probation officers, and correction officials. In other words, conditions tantamount to punishment are imposed, far from the oversight imagined by the Framers of the Constitution, and violating the true spirit of the Sixth Amendment jury trial right.

Although bail and detention was a popular scholarly topic a generation ago, only a few contemporary legal academics have scrutinized the current machinations of pretrial release, with existing scholarship primarily focusing on Fourth, Fifth, or Eighth Amendment violations. None, however, have analyzed the results of the changes in pretrial release standards and the increasing relevance of the Sixth Amendment. This Article aims to fill that gap by studying the problems of our current pretrial detention system through a Sixth Amendment lens.

I contend that the spirit of the Sixth Amendment jury trial right might apply to many pretrial detainees, due to both the punishment-like conditions of their incarceration and the unfair procedures surrounding bail grants, denials, and revocations. In doing so, I also expose some of the worst abuses of current procedures surrounding bail and jail in both federal and state systems.

This Article proceeds in four parts. Part II of this Article exposes the often intolerable and primitive conditions of state and local jails, which end up punishing all those incarcerated in them, whether convicted or not. Part III traces the history of pretrial detention, focusing as well on the resurgence of the Sixth Amendment jury trial right. Part IV explores how pretrial confinement has become a kind of punishment imposed inconsistently, by fluctuating actors, and without proper predictive basis. This part focuses on both the failures of the 1984 Bail Reform Act as well as the complete lack of predictability that

essential to punishment.

the current dangerousness formula possesses. Part IV also discusses how the current system of bail and jail entirely bypasses the community role in both imposing punishment and creating a safer living area. Finally, Part V introduces some much needed reforms in the pretrial release world, including reforming the surety bond system, reducing prison overcrowding by increasing electronic bail surveillance, and revising the bail hearing procedure to permit a community “bail jury” to help decide the defendant’s danger to the community.

Only recently has the national and local media shined a spotlight on both bail and jail procedures and their conditions and failures, exposing a dark corner of the criminal justice system where procedural fairness and due process are limited and sometimes nonexistent. This Article hopes to add a scholarly dimension to these troubling exposés, illustrating how pretrial detention violates the spirit of the Sixth Amendment and creates a Shadowlands within criminal justice.

II. Pretrial Detention as Punishment

Pretrial detention in the twenty-first century has evolved from a brief containment for a few accused deemed exceptionally dangerous to punishment for large numbers of accused awaiting trial. The combination of inhumane and degrading conditions, a corrupt and unregulated system of bail surety, bail bondsmen, and bounty hunters, and rising numbers of detainees, with the general absence of criminal due process in the pretrial realm, has resulted in a criminal justice system that punishes before it convicts. This contradicts the requirements of even our minimal pretrial protection for defendants, which holds that punishment can only occur after a conviction.16 Punishing the accused before she is proven guilty violates every theory of punishment, but particularly retributive justice, which requires that punishment can only be imposed after a cross-section of the community has pronounced guilt—a far cry from the pretrial detention system we have now. Although the abuses of pretrial detention are

beginning to garner media attention, only a little scholarly attention has been paid.\textsuperscript{17} It is time to remedy this oversight.

\textbf{A. Bail Bondsmen, Bounty Hunters, and Corrupt Incentives}

Our current bail system is by-and-large unregulated and plagued with corruption. First, numerous offenders languish in local jails for weeks for committing mere misdemeanors, simply because they lack the funds to post bail.\textsuperscript{18} In New York City, for example, most of these charges are for minor quality-of-life offenses, such as smoking marijuana in public, jumping a subway turnstile, or shoplifting, and bail was set at $1,000 or less.\textsuperscript{19} Yet, the overwhelming majority of defendants are unable to muster funds and are sent to jail, where they remain, “on average, for more than two weeks.”\textsuperscript{20} In a 2010 study, eighty-seven percent of the low-income defendants who were not released on their own recognizance were unable to post bail and went to jail to await guilty pleas or trial.\textsuperscript{21}

What is even more disturbing is that many of the poorer defendants may have pled guilty at arraignment for sentences with no jail time, simply to avoid being behind bars while awaiting trial.\textsuperscript{22} Impoverished defendants will often accept a guilty plea, even if innocent, in order to gain release from pretrial detention, even if this injures their long-term prospects.\textsuperscript{23}

\begin{itemize}
\item \textsuperscript{17} See, e.g., Baradaran & McIntyre, supra note 15; Baradaran, supra note 15.
\item \textsuperscript{19} Id.
\item \textsuperscript{20} See id. (finding that 87\% of defendants whose bail was set at $1,000 or less did not post bail).
\item \textsuperscript{22} See id. at 2–3 (“Most persons accused of low level offenses when faced with a bail amount they cannot make will accept a guilty plea; if they do not plea at arraignment, they will do so after having been in detention a week or two.”).
Inability to muster the appropriate funds for bail is not just a problem for misdemeanor felony defendants. Judges often set money bail at an amount the defendant cannot afford. In New York, for example, only ten percent of defendants in all criminal cases in which bail is set are able to post it at arraignment.24 This is despite the fact that many states have laws that establish a preference for nonfinancial conditions of release or unsecured bonds.25

The existence of commercial surety bonds, or secured bonds, does not help low-income defendants. Commercial bondsmen rarely lend bail money of $1,000 or less, and their services are usually too expensive for low-income or indigent offenders.26 Likewise, secured bonds are often not accessible for poor defendants, who usually do not have the property available to secure such bonds, or friends with such assets.27 Under one bondsman’s system, to obtain bail for even a minor felony or misdemeanor charge an indicted defendant must pay cash out of pocket, sign a twenty-page contract, and initial eighty-six separate paragraphs.28

If a defendant is fortunate enough to even qualify for secured bonds, then she must face the web of complex and innumerable fees charged for simple regulation. For example, in New York, a bondsman often charges the defendant a fee of $250 if the defendant misses a weekly check-in, and as much as $375 per

---

26. See Mary T. Phillips, N.Y.C. Criminal Justice Agency, Making Bail in New York City: Commercial Bonds and Cash Bail 6 (2010) (interviewing fourteen detainees with low bail amounts who were turned down by the bondsman). Bond agents typically charge a 10% fee for the first $3,000, 8% for the next $7,000, and 6% for amounts over $10,000. The fee is not refunded. Bondsmen also require collateral, typically cash, which is refunded unless bail is forfeited for failure to appear. See id. at 3.
27. In New York, for example, criminal procedure law authorizes the use of secured bail bonds secured by personal and real property; the surety may be provided by the defendant himself or someone other than the defendant.
hour for obscure tasks like bail consulting and research. These specified fees can grow much greater when bail bondsmen are assigned to other tasks, such as obtaining court documents or delivering release papers to jail.

Even if an indicted defendant can afford a commercial surety bond, these commercial bonds are almost entirely unregulated and often corrupt. At the frontiers of criminal justice, bail bondsmen hold an immense amount of power over the bailees, despite the bondsmen’s lack of legal, political, or police authority. Far from having a jury or a judge decide whether an indicted defendant should be incarcerated and punished, these bail bondsmen make such decisions in a completely unstructured universe, where they are both judge and jury. This kind of unauthorized decision-making surely violates the spirit of the Sixth Amendment, which at its very core requires legal conviction before punishment.

More troubling are the vast amounts, sometimes thousands of dollars, that a bail bondsman may charge if he makes the decision to revoke bail and return the defendant to jail. These decisions, made entirely on the bondsman’s own accord, with no regulation from any judicial, police, or legal authority, end up not only returning the defendant to jail but also costing him and his family a large percentage of the deposited bond, which is forfeited when the defendant is surrendered on the sole decision of the bondsman.

There are few state laws regulating when it is permissible for a bondsman to surrender a defendant, which leaves the bail system open to manipulation. New York bondsmen, among others, have been returning defendants to jail for questionable or unspecified reasons, and then withholding thousands of dollars to which the bondsmen may not be entitled.

Because most state laws allow bondsmen to enter into private contracts with the people they bail out, it is hard for judges to regulate their behavior. And even in states that afford

29. Id.
30. Id.
31. See id. (noting that one bondsman returned eighty-nine bailees over a four-month period and pocketed 15% of the bail when doing so).
32. See id.
some regulatory supervision, judges rarely take advantage of it. For example, in New York, although state law allows judicial consideration of the bondsman's background, character, and reputation when deciding whether to accept a bond, a judge rarely denies bond because of improper behavior.33

Instead of helping defendants stay out of jail, some bail bondsmen take advantage of the situation and disadvantage defendants to an even further degree.34 In other words, as described by a New York attorney familiar with these sorts of abuses, an indicted offender “can be ordered imprisoned by a court based solely on the unsworn, untested word of a non-law-enforcement civilian, a civilian who stands to profit financially if the defendant is incarcerated.”35 This type of entirely unregulated, potentially improper bond revocation, requiring the defendant to return to jail on the whims of a private actor, is yet another example of how the world of pretrial detention operates at the fringes of justice.

If bailees fail to appear for their hearings, the bondsman owes the entire bail amount to the court.36 This kind of financial liability has led to many bondsmen employing recovery agents, usually known as bounty hunters, to ensure that these indicted defendants appear for their court dates.37 The last time the Supreme Court addressed the role of bounty hunters and bondsmen—one hundred and fifty years ago38—it acknowledged the historical common law privileges of both bondsmen and bounty hunters, holding that the right to apprehend a fleeing defendant originates from the contract relationship between bondsmen and their clients.39 Despite vast changes in both

---

34. See Eligon, Bail System, supra note 28 (recounting abuses of N.Y.C. bail bondsmen that have been investigated by the New York State Insurance Department).
35. Id.
37. See id.
39. See id. at 370–71. The Taylor Court noted that bounty hunters could
criminal law and procedure, however, the Court has not addressed the topic since.

Bounty hunters do have a few regulations on their behavior, mostly codified in state law, though they vary widely from state to state.\textsuperscript{40} Despite persistent effort, however, attempts to control bounty hunters through federal legislation have failed.\textsuperscript{41} As should not be surprising in such a "wild west," behavior of some bounty hunters can be reprehensible. Beyond the showy brutality spotlighted in such reality shows as "Dog the Bounty Hunter,"\textsuperscript{42} the rules and prohibitions that constrain the police do not generally apply to bounty hunters.\textsuperscript{43} As a result, misconduct often occurs as bounty hunters take advantage of their legal privileges.\textsuperscript{44}

Bounty hunters look and act like police, but lack the screening and training that law enforcement provides its

\textsuperscript{40} Forty-seven states have some sort of statute regulating bounty hunting. \textit{See} Freeland, \textit{supra} note 36, at 210 n.70. For example, in 1997, Indiana, Nevada, and North Carolina started licensing bounty hunters, and Texas began to require warrants for bounty hunters as well as the assistance of licensed private investigators/security guards. Additionally, Florida, Illinois, Kentucky, and Oregon have banned commercial bail bond systems, and hence have no bounty hunters. \textit{See id.} at 210 nn.66, 70.

\textsuperscript{41} \textit{See id.} at 211–12 (noting that two bills that would have made bounty hunters subject to the same laws and constitutional constraints as police failed in the House of Representatives).

\textsuperscript{42} “Dog the Bounty Hunter” is a reality television show on A&E, which follows Duane Chapman as he hunts down defendants who have missed court appearances. \textit{See} DOG THE BOUNTY HUNTER, http://www.dogthebountyhunter.com/ (last visited Sept. 24, 2012) (on file with the Washington and Lee Law Review). The website for the show proclaims, “Considered the greatest bounty hunter in the world, Duane ‘Dog’ Chapman has made more than 6,000 captures in his twenty-seven-year career.” \textit{Id.}

\textsuperscript{43} \textit{See} JACQUELINE POPE, BOUNTY HUNTERS, MARSHALS, AND SHERIFFS: FORWARD TO THE PAST 4 (1998) (“Rules of law and conduct under which police function have no relevance for bounty hunters.”).

\textsuperscript{44} \textit{See} Freeland, \textit{supra} note 36, at 212 (collecting stories of incidents).
recruits. As a result, these recovery agents use dangerous and sometimes illegal tactics to retrieve defendants, property, or both, including confrontations at gunpoint, forced entries into homes without a search warrant, and the commandeering of vehicles on public roadways. Indeed, in states where they are allowed to practice, bounty hunters can legally use stun guns, mace, and firearms while apprehending bailees. As a result, in the process of “recovering” the bailee, bounty hunters often complicate and endanger public safety. Moreover, although bounty hunters play a police-like role, there is little constitutional protection against poor behavior because they are usually not considered state actors.

In sum, the unregulated private actors, unsupervised and unaccountable bail bonding companies, complex and unfair fee structures, tremendous pressure to plead guilty, over-incarceration for minor offenses, and disproportionately high bail all combine to make our system of pretrial detention a nightmare to navigate and constitutionally questionable. That this system disproportionately affects the poor makes our current pretrial detention system all the more disturbing.

B. Increased Numbers of Poor Indicted Offenders Denied Bail

At any given moment, a large proportion of jail dwellers consist of felony and nonfelony pretrial detainees who are in jail because they have not posted bail. Of the nation’s jail population, 60.2% are detainees awaiting trial. Nationally, taxpayers spend $9 billion annually to incarcerate defendants held on bail.

46. Id.
47. Id.
48. See Freeland, supra note 36, at 228 (discussing the inadequacy of the logic behind distinguishing police officers and bounty hunters).
The rate of pretrial incarceration (and incarceration in jails and other nonprison detention places in general) has continued to rise over the past ten years even though prison growth rates have been leveling off.\textsuperscript{51} Indicted offenders are less likely to be released pretrial.\textsuperscript{52} This includes not only those who have been indicted for violent offenses but also those who are awaiting trial for property, drug, and public-order-related charges.\textsuperscript{53} Additionally, fewer indicted offenders are being released from jail on their own recognizance, and those who have been granted bail are often unable to afford it.\textsuperscript{54} These types of high bail requirements make it very difficult for indicted defendants to obtain pretrial release, despite the fact that the vast majority of these offenders have been arrested for low-level, nonviolent offenses.\textsuperscript{55}

Despite this increasing reliance on incarcerating indicted defendants before trial, communities are not necessarily any safer. The places with the highest incarceration rates have not necessarily seen violent crime rates fall.\textsuperscript{56} In fact, quite to the contrary, New York City decreased its jail population and experienced a drastic reduction in crime rates.\textsuperscript{57}

As Human Rights Watch has astutely noted in discussing the problems with pretrial detentions in New York City:

\begin{quote}
Time in jail before one has had one’s day in court is particularly troubling for the one in five detained non-felony defendants who . . . will not be convicted. It is also disproportionate in light of sentences typically imposed when there is a non-felony conviction: data from the New York State Division of Criminal Justice Services, for example, indicates that eight out of ten convicted misdemeanor arrestees receive sentences that do not include jail time.\textsuperscript{58}
\end{quote}

\textsuperscript{51} Petteruti & Walsh, supra note 11, at 2.
\textsuperscript{52} Id. at 3.
\textsuperscript{53} Id. Indeed, as the report notes, three-quarters of those pretrial detainees charged with property, drug, and public order related charges are “significantly less likely” to be released. Id.
\textsuperscript{54} Id. As the report explains: “Once, more than half of those jailed received bail amounts of $5,000 or less; today, just about half of the people in jail receive the highest bail amounts ($10,000 to the maximum).” Id.
\textsuperscript{55} Id.
\textsuperscript{56} See Human Rights Watch, supra note 21, at 4.
\textsuperscript{57} See id.
\textsuperscript{58} Id. at 2.
This is disturbing in regards to the large number of pretrial detainees who are indicted and held on misdemeanors, as detailed above. Incarcerating poor defendants for nonfelony offenses (primarily misdemeanors, but also violations and infractions)\footnote{Id.} is perhaps the most troubling aspect of this trend. This kind of jailing is “uniquely difficult to reconcile with the fundamental notions of fairness and equality that should be the cornerstones of criminal justice.”\footnote{Id.} For these cases especially, pretrial detention is a disproportionate abbreviation of rights, particularly in light of the nonthreatening, petty nature of most of the charged nonfelony crimes.\footnote{Id.}

The increasing rate of pretrial detention is worrying, however, even for those defendants charged with more serious crimes, due to the conditions of the actual detention centers housing pretrial defendants. Although state and federal prisons are not generally known for their plush accommodations, the general state of the jails that hold pretrial detainees is so bad that simply to be incarcerated in them rises to a punitive experience.

C. The Punishing Conditions of Pretrial Detention

The substandard conditions of today’s pretrial detention centers—our halfway houses and local and municipal jails—have transformed the detainee’s experience into a punishing one. It is a little-known but unfortunate truth that pretrial detainees often undergo harsher conditions of confinement than those defendants who are convicted.\footnote{Gorlin, supra note 10, at 419.} While state and federal prisons house most convicted prisoners, jails and county lockups house the accused who have either been denied bail or cannot afford to pay it.

Moreover, pretrial detainees are often incarcerated alongside the ten percent of convicted criminals who are housed in jails rather than prisons.\footnote{Petteruti & Walsh, supra note 11, at 3.} This indicates that the holding conditions for pretrial detainees are, at minimum, punishment-like, as it is

\footnote{Id.}

\footnote{Id.}

\footnote{Id.}
precisely the same as that for some convicted offenders. The lines between prison and jail are becoming increasingly blurred, and not for the better.64

Historically, determining whether certain conditions rise to the level of punishment requires objective indicia. Specifically, in *Bell v. Wolfish*, the Supreme Court held that courts can decide “whether [the detainee’s condition] has historically been regarded as punishment.”66 In addition, *Youngberg v. Romeo* held that those forcibly committed to state institutions retain their substantive rights under the Due Process Clause.68 As others have noted, the Court’s holding in *Youngberg* indicates that, by the same token, criminal pretrial detainees’ arguments should be analyzed objectively, instead of subjectively.69

There are numerous objective indicia of confinement that rise to the level of punishment for pretrial detainees. By far the most concerning—but not isolated—examples come from Rikers Island in New York, the municipal lockup for pretrial detainees, immigrants, juveniles, and any prisoner subject to rehearing or resentencing.70 The jail on Rikers Island has been sued in recent years by more than a half-dozen Rikers inmates, all claiming to have been the victims of beatings by prisoners while guards ignored it, or worse, ordered the attacks.71

64. See id. (noting that criminals are increasingly sent to jail, not prison).

65. See *Bell v. Wolfish*, 441 U.S. 520, 537 (1979) (holding that courts may decide “whether [the detainee’s condition] has historically been regarded as punishment”).

66. Id.

67. See *Youngberg v. Romeo*, 457 U.S. 307 (1982) (holding that a mentally handicapped prisoner had “constitutionally protected liberty interests under the Due Process Clause of the Fourteenth Amendment to reasonably safe conditions of confinement, freedom from unreasonable bodily restraints, and such minimally adequate training as reasonably may be required by these interests”).

68. Id. at 315–16.


71. Benjamin Weiser, *Lawsuits Suggest Pattern of Rikers Guards Looking*
For example, in 2009, two Rikers guards were accused of recruiting inmates over a three-month period to assist in maintaining order in a housing unit for teen boys, including training the inmates in how to restrain and assault their victims, as well as deciding where and when attacks would occur.\textsuperscript{72} The housing unit was run much like a Mafia organization, in which two correction officers were the bosses.\textsuperscript{73} Even more troubling, the recent pattern of cases involving Rikers guards indicates that the management of the jail is, if not complicit in the abuses, at least marginally aware of it.\textsuperscript{74}

The types of punitive violations occurring at Rikers do not just involve misconduct by guards. For years the jail had a policy of strip-searching all prisoners, even nonviolent ones charged with minor crimes, and, over an eight-year period, roughly 100,000 people were strip-searched after being charged with misdemeanors and taken to Rikers Island and other city correction facilities.\textsuperscript{75} A majority of the strip-searched detainees were charged with trespassing, shoplifting, jumping turnstiles, or failing to pay child support,\textsuperscript{76} minor offenses on even a misdemeanor scale. Considering most federal circuits have upheld laws banning strip-searches for detainees charged with minor offenses,\textsuperscript{77} this course of action violated the pretrial


\textsuperscript{72} Id.

\textsuperscript{73} John Eligon, \textit{Correction Officers Accused of Letting Inmates Run Rikers Island Jail}, N.Y. TIMES, Jan. 23, 2009, at A20. As the article explained, “guards reputedly sent inmates to intimidate, threaten and silence uncooperative prisoners with brute force. Inmates were ordered to turn over money, and their every move, including when they could use the bathroom, was controlled. If word of an assault got out, the guards would allegedly orchestrate a cover-up.” Id.

\textsuperscript{74} Weiser, \textit{supra} note 71.

\textsuperscript{75} Michael Schmidt, \textit{City Reaches $33 Million Settlement over Strip Searches}, N.Y. TIMES, Mar. 23, 2010, at A22.

\textsuperscript{76} See id.

\textsuperscript{77} See, e.g., Evans v. Stephens, 407 F.3d 1272, 1290–91 (11th Cir. 2005) (noting other cases where officers were required to have reasonable suspicion to strip search minor offenders); Doe v. Burnham, 6 F.3d 476 (7th Cir. 1993) (“Law enforcement officers may not strip search an individual for contraband unless the officers have a reasonable basis to believe at the time of the search that the individual is concealing contraband on his or her body.” (citing Bell v. Wolfish, 441 U.S. 520, 559 (1979); Mary Beth G. v. City of Chicago, 723 F.2d 1263, 1273 (7th Cir. 1983))).
detainees’ Fourth Amendment rights and created an atmosphere of fear and punishment.

Jails in other major metropolitan areas are rife with similar abuses. The Los Angeles County Men’s Central Jail is notorious for overcrowding, forcing inmates to sleep on dirty floors, devoid of natural light, where inmates get only one two-hour outdoor recreation session per week on the jail’s roof. In Maricopa County’s Tent City Jail, in Phoenix, Arizona, inmates are housed outdoors in military tents without air conditioning (despite over 100 degree temperatures in the summers), fed 15-cent meals only twice a day to cut costs, are forced to wear humiliating prison gear, and have very few amenities. And in Washington, D.C.’s Central Detention Facility, there is overcrowding, serious health and sanitation issues (including broken showers, no running water in cells, and animal feces throughout the facility), and inadequate healthcare. All this is in addition to the abuse that pretrial detainees can be subjected to in jails, ranging from violent treatment by other prisoners to more institutionalized practices.

On a somewhat less deadly scale, but still punitive in nature, are the fees now imposed on poor detainees by many jails. In Florida, for example, fees are imposed for the use of the public defender for misdemeanors, which tend to include a large section of pretrial detainees who often cannot afford the fees to make bail. Adding a $50 fee to even consult a public defender undoubtedly has a chilling effect on many of these indicted offenders, who may go without counsel due to an inability to afford the fee. Moreover, failure to use the public defender can then lead to a cascade of effects, particularly for driving violations including court-ordered fees, followed by failure to pay, which can lead to more fees, more unlicensed driving, and sometimes incarceration. Under Florida and North Carolina law, there are

---

78. Petteruti & Walsh, supra note 11, at 19.
79. Id.
80. Id.
82. Id. at 4.
83. Id. at 5–6. Specifically, there is a $50 “application fee.” See id.
84. Id. at 4.
no exceptions or waivers for the indigent, and the waiver is rarely utilized in Georgia.

Likewise, in Louisiana, fees for using the public defender are imposed after the bail hearing. All of the criminal courts of Orleans Parish impose fines and fees regardless of an indicted defendant’s ability to pay them, and waiver is rarely granted. For very poor defendants who often cannot come up with bail money, the imposition of another fee on top of the bail fee, for simply consulting the public defender, results in a fee-based punishment. This is exacerbated by the fact that when defendants are unable to pay their fines, fees, and costs, they may be incarcerated—even if they have been found not guilty of their original crime. Once incarcerated, the indigent defendant is even less likely to be able to pay the fees, which compound, leading to higher debt and longer incarceration. Because they can only afford to pay small amounts of their incomes to redeem their fees, many of these men and women can remain caught up in the criminal justice system for years, and they may find themselves back in jail when their legal debts become overwhelming.

In impoverished states such as Michigan, some pretrial detainees are assessed a $12 jail entry fee, $60 per day for jail room and board, and additional reimbursement to the correctional facility for medical and other services. One jail in Michigan requires the defendant to pay a $12 fee to be released from jail. So for an indigent pretrial detainee who cannot afford

---

85. Id. at 7.
86. ACLU, IN FOR A PENNY: THE RISE OF AMERICA’S NEW DEBTOR PRISONS 62 (2010). Georgia requires a $50 fee from all poor criminal defendants, including pretrial detainees, who simply request the services of a public defender. See id. Most of such defendants never get a chance to demonstrate their indigence and simply waive their right to counsel. See id.
87. Id. at 17. The public defender fee is usually $40. Id.
88. Id.
89. Id.
90. Id. at 18. Late fees of $100 per payment are often imposed. Id. Defendants may also be sent to jail for fifteen to thirty days for failure to pay. Id.
91. See id. at 21.
92. See id. at 30.
93. See id. The aforementioned facility is the Saginaw County Jail, which is
to meet bail, extra costs are imposed, even if the original charges ultimately are not proven or are dropped.

Similarly, states such as Ohio frequently impose “pay-to-stay” programs on pretrial detainees. It took a federal lawsuit to stop an Ohio municipal jail from requiring detainees to pay a daily fee for their preconviction jail time. A comparable payment program for pretrial detainees also existed in Georgia until recently, in which the sheriff of Clinch County routinely charged pretrial detainees for the costs of room and board well before conviction. In certain cases, the sheriff even forced detainees to choose between signing a promissory note (to be later enforced) or being returned to jail. This practice only stopped with the initiation of a federal lawsuit, which settled in 2006.

These practices, taken together, have done much to transform pretrial detention into a modern-day debtor’s prison, and transformed it from a regulatory to a punitive experience.

---

94. See ACLU, supra note 85, at 53. The Hamilton County Jail, in Ohio, routinely charged its pretrial detainees a “pay to stay” fee until 2002, when the District Court for the Southern District of Ohio found this violated pretrial detainees’ right to the Fourteenth Amendment’s guarantee of due process, because the detainees were not given a predeprivation opportunity to be heard. See Allen v. Leis, 213 F. Supp. 2d 819, 832 (S.D. Ohio 2002) (holding jail’s policy of appropriating cash immediately upon pretrial detainee’s arrival at jail to cover “booking fee” violated defendants’ due process rights under the Fourteenth Amendment).

95. ACLU, supra note 86, at 56.

96. Id. at 56–57.

D. Pretrial Incarcerative Harm

The offenders who are incarcerated pretrial suffer unquestionable harm from this detention.\textsuperscript{98} In general, incarceration in jail negatively impacts the mental and physical health, employment, and family and community interactions of those incarcerated.\textsuperscript{99} Pretrial incarceration is also particularly difficult for those indicted offenders who suffer from poor health, as jails rarely have adequate resources available to treat people with physical or mental health problems.\textsuperscript{100} Additionally, the poor have a much thinner safety net keeping them from homelessness and abject poverty, and being incarcerated for potential crimes, even for a short time, can have a devastating effect.\textsuperscript{101}

Moreover, jails can be dangerous and unhealthy environments, even more so than prisons. First, the jail buildings themselves are often old and decaying.\textsuperscript{102} These old buildings can have various dangers associated with them, including mold, poor ventilation, lead pipes, and asbestos, all of which can be very detrimental to the health of pretrial detainees.\textsuperscript{103} Second, the concentration of prisoners, wardens, and visitors in a jail make it a vector of contagious diseases.\textsuperscript{104} This is in large part because serious infections and sexually transmitted diseases are highly concentrated and easily transmitted in jails,\textsuperscript{105} and the ever-changing detainee population means the residents are constantly in flux.\textsuperscript{106} For example, the MRSA drug-resistant superbug has been thriving in jails, with the potential to infect those detainees who are there even for only a short time.\textsuperscript{107} Many jails are not

\textsuperscript{98} See Gorlin, supra note 10, at 419.

\textsuperscript{99} See Petteruti & Walsh, supra note 11, at 3.

\textsuperscript{100} Id.

\textsuperscript{101} See ACLU, supra note 86, at 6.

\textsuperscript{102} See Petteruti & Walsh, supra note 11, at 15 (citing studies from the 1990s that show that 700 jails in the United States are older than fifty years old and 140 jails are more than 100 years old).

\textsuperscript{103} See id.

\textsuperscript{104} See id.

\textsuperscript{105} For example, HIV, antibiotic-resistant TB, and bacterial infections all exist at much higher rates in jails than in the general population. See id.

\textsuperscript{106} See id.

\textsuperscript{107} See Silja J.A. Talvi, Deadly Staph Infection ‘Superbug’ Has a Dangerous Foothold in U.S. Jails, ALTERNET.COM (Dec. 4, 2007), available at
properly equipped to treat serious health problems in their detainees, and what healthcare is available is hard to provide to jails’ often short-term visitors.108

Drug and alcohol addiction and mental illness are an equally large problem in jails. There is minimal, if any, drug treatment, although as many as half or more of those arrested and detained have had issues with drug addiction, alcohol addiction, or both.109 And ever since the wave of deinstitutionalizing the mentally ill thirty years ago, a large number of the jails’ residents are mentally unstable.110

This is troubling for a number of reasons. First, incarceration—even short-term incarceration such as pretrial detention—tends to further traumatize people with mental illness, making them more at risk of harming themselves or others.111 This is particularly true in regards to suicide, which is the second-highest reason for death after illness in jails.112 These high suicide rates are closely linked with untreated depression, all too common in all correctional facilities.113 To further complicate the situation, many jails lack the institutional mental health resources required to serve the needs of their detainees.114

Pretrial detention also exerts a burden on an indicted offender’s family. Children of indicted offenders often end up in foster care or are otherwise taken away from their families and

http://www.alternet.org/story/69576 (last visited Sept. 24, 2012) (detailing how Superbug spreads all too easily in jails because of problems like poorly ventilated living and sleeping quarters; overcrowded rooms; shared mattresses, toilets, and showers; and a preponderance of people who arrive with poor health, drug problems, and severely compromised immune systems) (on file with the Washington and Lee Law Review).

108. PETTERUTI & WALSH, supra note 11, at 15.


110. PETTERUTI & WALSH, supra note 11, at 3.

111. See MAEGHAN GILMORE & MARY-KATHLEEN GUERRA, NATIONAL ASSOCIATION OF COUNTIES, CRISIS CARE SERVICES FOR COUNTIES 1 (2010) (noting the detrimental effects of detention on juveniles with mental health disorders).

112. See PETTERUTI & WALSH, supra note 11, at 16 (finding that the suicide rate in jails is 42 per 100,000 compared to 11 per 100,000 for the general population).

113. Id.

114. Id. at 15.
are far more likely to fall into poverty.\textsuperscript{115} Family members of the person in jail experience not only emotional and economic hardships, but some have also reported experiencing physical ailments and declining health.\textsuperscript{116} In Michigan, courts have gone so far as to incarcerate a mother for being unable to pay for her child’s incarceration costs, locking her up without even any crime charged.\textsuperscript{117}

Pretrial detention also augments the possibility of conviction.\textsuperscript{118} Incarcerated defendants before trial are more likely to be found or plead guilty and serve prison time than those released pretrial.\textsuperscript{119} The mere possibility of pretrial imprisonment often compels defendants to plead guilty and give up their right to trial.\textsuperscript{120} The prospect of being incarcerated, even for a short time, can look ruinous to poor defendants, as this often means the loss of their livelihood, severe disruptions to their family lives, or both. Accordingly, when confronted with an unaffordable bail, a large number of pretrial detainees simply plead guilty.\textsuperscript{121} This rush to a guilty plea is often exacerbated by the application fee to use a public defender in some states, as detailed above.\textsuperscript{122}

Finally, pretrial detention can, in some cases, be literally deadly. Privatization of jails, prisons and halfway houses in states such as New Jersey have resulted in the housing of violent

\textsuperscript{115} Dolovich, supra note 109, at 247.

\textsuperscript{116} See Petteruti & Walsh, supra note 11, at 18 (finding that 48% of those related to a person in jail experienced declining health after the person was jailed and 27% reported that their children’s health had dropped).

\textsuperscript{117} ACLU, supra note 86, at 35. The mother in question was also charged a room-and-board fee, a drug test fee, and a booking fee for her imprisonment in the jail. Id. at 35–36.

\textsuperscript{118} Human Rights Watch, supra note 21, at 2.

\textsuperscript{119} See Baradaran & McIntyre, supra note 15, at 555. As the authors expound, “Detention leads to the loss of employment and other negative financial conditions, less likelihood to obtain private counsel, which harms defendant’s chances to be acquitted or at sentencing.” Id.

\textsuperscript{120} Human Rights Watch supra note 21, at 3; see also Baradaran & McIntyre, supra note 15, at 555 (“[L]iving conditions in jail are often poor and have been shown to have a negative influence on defendant’s trial demeanor.”).

\textsuperscript{121} Human Rights Watch, supra note 21, at 3. For example, guilty pleas account for 99.6% of all convictions of New York City misdemeanor defendants. Id.

\textsuperscript{122} See supra Part II.C.
JUSTICE IN THE SHADOWLANDS

convicted offenders with nonviolent pretrial detainees. More than once, this mixture of low-level detainees with dangerous convicted felons has resulted in injury or death for those who cannot make bail for misdemeanor charges.

All of these practices are transforming our imposition of pretrial detention from its original incarnation as brief confinement based on risk of flight to punitive incarceration decided by a fragmented and inconsistent variety of private and public actors. This creates two major problems: not only do our procedures for imposing jail and denying bail disproportionately affect the poor and disenfranchised, but, as I contend below, they also violate the spirit of the Sixth Amendment. Taken together, these practices have created a Shadowlands where the normal promises of substantive and procedural criminal justice do not apply.

E. Punishment Before Conviction Violates the Spirit of the Sixth Amendment Jury Trial Right

This prolonged pretrial incarceration feels troubling because it seems to punish accused offenders before conviction by members of the community, violating the very spirit of our criminal justice system. These offenders are considered innocent at this phase of the criminal process. Any discomfort we feel with such practices logically stems from the Sixth Amendment jury trial right, which holds that the accused “shall enjoy the right to a speedy and public trial, by an impartial jury of the state and district wherein the crime shall have been committed.” Our current pretrial detention practices violate the very tenets of the Sixth Amendment: the accused are incarcerated for lengthy periods, suffering punitive and dangerous conditions in jails and county lockups, based on decisions made by unaccountable private actors, harried magistrates, or line prosecutors.

123. See Dolnick, supra note 9, at A1 (finding that Delaney Hall, a new Jersey halfway house, housed at least one person accused of murder).

124. See id. (detailing the murder of a man, arrested for unpaid parking tickets and failure to purchase car insurance, for the three dollars in his pocket).

125. U.S. CONST. amend. VI.
Punishment, in other words, is being imposed on those not yet convicted, without the imprimatur of the jury.

Our pretrial detention practices are even more questionable when contrasted against the Supreme Court’s recent spate of opinions highlighting the Sixth Amendment right to a jury trial. Specifically, in the Apprendi-Blakely line of cases, the Supreme Court reinvigorated the Sixth Amendment jury right, concentrating on the need for the community, as jury, to impose punishment on those found guilty. By focusing on this basic idea—a valid conviction requires all aspects of a crime be determined by a jury—the Court “provided the basis for [its] . . . decisions interpreting modern criminal statutes and sentencing procedures.” The Court relied heavily on the historical role of the community as an arbiter of punishment to support its contention that only the jury could find facts that increased a convicted offender’s penalty. In holding that a court can sentence a defendant only on facts found by the jury beyond a reasonable doubt or admitted by the defendant himself, the Blakely Court gave strong support to the idea that the community must have the final word on criminal punishment. Thus, the basis of the Court’s new focus on the rights of the jury in criminal adjudication rested on the importance of the community’s determination of punishment.

Despite these recent decisions, bail and jail determinations still take place far from the community and the jury room, taking place in the barrens of procedural justice. The Court’s focus on community participation in criminal adjudication was not limited to criminal trials, as is illustrated in its recent opinion in Southern Union v. United States, holding that a jury must decide on the imposition of a criminal fine. The Court’s refusal to limit Blakely/Apprendi to jury trials leaves an opening to integrate the community jury right into the pretrial detention sphere. If the Supreme Court has focused on the jury as a

127. See id. at 313 (“If a State makes an increase in a defendant’s authorized punishment contingent on the finding of a fact, that fact—no matter how the State labels it—must be found by a jury beyond a reasonable doubt.” (quoting Ring v. Arizona, 536 U.S. 584, 588–89 (2002))).
129. See id. at 2357 (holding that Apprendi applies to criminal fines).
representative of the community, as the only appropriate body to impose punishment on a convicted offender, then how much more important is it that the community have a say in determining whether punitive conditions fall upon an unconvicted offender? The Apprendi-Blakely line of decisions, forbidding imposition of punishment until the jury has decided guilt or innocence, must inform our practices governing pretrial detention.

Imposing our bail and jail procedures upon pretrial detainees results in the imposition of unjustified punishment, taking place virtually unnoticed and unremedied. As such, the spirit of the Supreme Court’s recent Sixth Amendment jurisprudence should also apply to pretrial detention in a variety of circumstances. Whenever detention turns from regulatory to punitive, the community must have a say in the punishment imposed.

Looking back at both our constitutional and historical understanding of bail, it is difficult to understand how we got here, routinely meting out punishment to an accused not yet convicted of a crime. Our historical bail practices differed greatly from the complicated and often bewildering array of rules that govern pretrial detention today. Thus, a thorough understanding of the history of pretrial detention is critical to fully comprehending the problems we face today.

III. A Short History of Bailing and Jailing

Although the history of Anglo-American bail procedures has been well-covered, a brief review of how bail and jail evolved in this country both before and after the American Revolution will prove helpful in showing how far we have departed from our original understanding of both. Since the Supreme Court has shown a great fidelity to how bail was originally granted in deciding pretrial detention cases, we should strive to comprehend the actual working customs during the nation’s earliest days.
A. Colonial Practices

The practice of bail came over from England with the first colonists. The system of bail developed to free untried prisoners. Like so many of our current criminal procedures, the bare bones of colonial bail were originally quite simple: the accused had a friend or neighbor take a pledge, backed by property, and assume responsibility for him until trial. In determining bail, the judge usually considered such factors as likelihood of conviction, risk of flight, severity of sentence, and the character of the accused. Many of these provisions were aimed at limiting judicial discretion in bail decisions, not at providing liberty for the defendant. This is unsurprising, considering how much criminal justice in the Anglo-American world was focused on community justice. Early colonial communities were loath to allow a visiting magistrate to make any major decisions about one of their own offenders.

In 1628, the English Petition of Right, thought by many to be the indirect progenitor of colonial bail law, held that bail was to obtain "the liberty of the subjects" from pretrial imprisonment.

---

130. See W. Thomas, Bail Reform in America 11 (1976) ("The American system of bail is derived from practices that originated in medieval England.").
131. See Betsy K. Wanger, Note, Limiting Preventive Detention Through Conditional Release: The Unfulfilled Promise of the 1982 Pretrial Services Act, 97 Yale L.J. 320, 323 n.19 (1987). Wagner goes on to note that this system developed largely because magistrates in medieval England traveled among different counties, and permitting defendants to be released into the custody of friends or neighbors as a surety helped avoid their prolonged detention in jail. Id. Originally the surety had to deliver himself if the defendant absconded; later, the surety could forfeit money instead of his own person. Id.
134. See Hermine Meyer, Constitutionality of Pretrial Detention, 60 Geo. L.J. 1140, 1162–63 (1972) (indicating that the purpose of these provisions was to limit the "admittedly unlimited discretion" of the judges regarding bail to ensure that noncapital defendants had a right to bail).
135. See id.
137. 3 How. St. Tr. 80–224.
It was this understanding of the right to bail that the colonists brought with them from the mother country. Excluding capital cases, defendants were guaranteed release on bail before trial.138

On the American continent, right to bail provisions existed in several colonial charters and was articulated as early as 1641 in the Massachusetts Body of Liberties.139 Other colonies such as Pennsylvania,140 Delaware,141 and New York142 followed suit. Of

138. See Shima Baradaran & Frank McIntyre, supra note 15, at 499 n.1 (citing 2 Matthew Hale, The History Of The Pleas Of The Crown 289 (1676)).

139. See Massachusetts Body of Liberties, art. 18 (1641), available at http://www.winthropsociety.com/liberties.php. Article 18 provided:

No man’s person shall be restrained or imprisoned by any authority whatsoever, before the law hath sentenced him thereto, if he can put in sufficient security, bail, or mainprise, for his appearance and good behavior in the meantime, unless it be in capital crimes, and contempts in open Court, and in such cases where some express act of Court doth allow it.

Id.


141. Delaware adopted the Pennsylvania Frame of Government, including its bail provision, when it became a colony in 1702. Verrilli, supra note 136, at 337.


THAT In all Cases whatsoever Bayle by sufficient Sureties Shall be allowed and taken unless for treason or felony plainly and specially Expressed and menconed in the Warrant of Commitment provided Always that nothing herein contained shall Extend to discharge out of prison upon bayle any person taken in Execucion for debts or otherwise legally sentenced by the judgment of any of the Courts of Record within the province.

Id. However, there is some evidence that New York’s right to bail provision was honored more in the breach than in the execution. Few acknowledged a right to bail in the eighteenth century. See J. Goebel & T. Naughton, Law Enforcement in Colonial New York 502–03 (1944) (finding that neither defendants nor courts in New York viewed bail as a matter of right).
course, as many major crimes were still classified as capital felonies during the colonial era, any provisions granting bail in noncapital cases still excluded numerous defendants.

B. Bail Following the Constitution

Despite these specific discussions of bail in colonial documents, however, the Framers did not explicitly include a right to bail in the Constitution, only mentioning it in the context of the Eighth Amendment. Depending on how colonial history is interpreted, the lack of an explicit right to bail in the Constitution can be seen either as a historical accident or a deliberate decision. There is minimal documentary evidence of the Framers’ intent to support either position. In contrast, the Judiciary Act of 1789 did specifically provide a right to bail for all noncapital cases, although there is no evidence of any debate on that provision either.

143. See Todd R. Clear, George F. Cole & Michael D. Reisig, American Corrections 72–73 (2008) (noting that the Anglican Code in force during the mid-Eighteenth Century listed thirteen capital offenses). As the authors note, slightly more than twenty percent of felonies were capital ones in New York. See id. at 73.

144. The Eighth Amendment provides, among other things, that “excessive bail shall not be required.” U.S. CONST. amend. VIII.

145. See, e.g., Caleb Foote, The Coming Constitutional Crisis in Bail, 113 U. PA. L. REV. 959, 968–69 (1965) (arguing that the failure to include a right to bail in the Constitution was a historical accident).

146. See, e.g., William F. Duker, Right to Bail: A Historical Inquiry, 42 ALB. L. REV. 33 (1977) (arguing that nothing in the colonial history of bail or the history of the Bill of Rights evidences any intent to have a right to bail).

147. See Verrilli, supra note 136, at 338 n.58. According to the 1788–90 Annals of Congress, the discussion of the bail clause in the Eighth Amendment was limited to one comment. See id.

148. The Judiciary Act provided, in regards to bail:

[U]pon all arrests in criminal cases, bail shall be admitted, except where punishment may be by death, in which cases it shall not be admitted but by the supreme or a circuit court, or by a justice of the supreme court, or a judge of a district court, who shall exercise their discretion therein, regarding the nature and circumstance of the offense, and of the evidence, the usages of law.


149. Verrilli, supra note 136, at 338 n.58.
Congress in 1787, also contained a right to bail. On the whole, then, it is difficult to determine the particular intent of the Framers in regards to the right to bail.

Following the ratification of the Constitution, the federal judiciary made clear that bail was the norm following indictment, due to the presumption of innocence and due process.

The right to bail after 1789 also solidified through the vehicle of state constitutions. Specifically, although only two of the original colonies—North Carolina and Pennsylvania—retained a specific right to bail in their state constitutions, every state that joined the Union after 1789, excluding West Virginia and Hawaii, included a right to bail. This right managed to survive the “frequent redrafting of state constitutions that occurred during the nineteenth century.” Viewed another way, it was truly in the state constitutions that the American right to bail reached its full fruition.

This fully articulated state right to bail is important for a variety of reasons. First, most criminal law is state law, not federal law, despite the scholarly and popular focus on federal law enforcement.

150. See Northwest Ordinance § 14, art. 2 (1787), reprinted in Documents Illustrative of the Formation of the Union of the American States (Charles C. Tansill, ed., Government Printing Office 1927), H.R. Doc. No. 398, available at http://avalon.law.yale.edu/18th_century/nworder.asp. The bail provision in Article 2 was identical to the Pennsylvania Frame of Government’s bail provision, providing that “[al]l persons shall be bailable, unless for capital offenses, where the proof shall be evident or the presumption great.”

151. See Verrilli, supra note 136, at 350 (finding that it is impossible to determine the Founders’ intent by evidence drawn from before 1789).

152. Granted, many felonies during this time were classified as capital offenses. See Act of Apr. 30, 1790, ch. 9, 1 Stat. 112 (including as capital crimes treason, murder, piracy, counterfeiting, and robbery on the high seas).

153. See Ex parte Milburn, 34 U.S. 704, 710 (1835) (holding that bail is not “designed as satisfaction for the offense, when it is forfeited and paid; but as a means of compelling the party to submit to the trial and punishment, which the law ordains for his offense”); see also Taylor v. Tainter, 83 U.S. 366, 371–72 (1872).

154. Verrilli, supra note 136, at 351.

155. Id. at 352.

156. See id. (detailing the history of bail provisions in initial state constitutions and right-to-bail amendments).

157. This is particularly true with bail, as state and local detention practices
most state criminal justice codes eliminated many felonies from the list of capital crimes, generally leaving only murder and treason.\textsuperscript{158} Third, the historical right to bail, as articulated by the states, has only denied bail for reasons involving risk of flight, rejecting the newer preventative detention theories.\textsuperscript{159} Finally, this development shows that the right to bail, although not firmly rooted in a specific constitutional provision, has been part of the American criminal justice system since the founding of the country.

Moreover, how a specific right developed in state statutes has often been important to the Supreme Court when analyzing the scope of rights in the federal constitution. For example, in \textit{Jones v. United States},\textsuperscript{160} the Supreme Court looked at how the states treated certain aspects of an aggravated crime as either an element or a sentencing factor in determining whether the reach of the Sixth Amendment jury trial right required defining serious bodily harm as an element of federal carjacking, as opposed to a sentencing factor.\textsuperscript{161} Similarly, in \textit{Duncan v. Louisiana},\textsuperscript{162} the Supreme Court used the long history of jury trial rights in state constitutions to bolster its support for the jury trial right in the federal constitution.\textsuperscript{163}

\begin{itemize}
\item \textsuperscript{158} See Verrilli, \textit{supra} note 136, at 352 (noting that the constitutional amendments giving a right to bail occurred at a time where many states were pruning the definition of capital crimes to include only murder and treason).
\item \textsuperscript{159} Granted, many states have recently amended their constitutions to allow detention. Miller & Guggenheim, \textit{supra} note 15, at 345.
\item \textsuperscript{160} Jones v. United States, 526 U.S. 227, 236 (1999) (stating that the court’s finding of serious bodily injury after trial and consequent sentence enhancement was error since serious bodily injury was an element of the crime).
\item \textsuperscript{161} See \textit{id.} at 236–37 (reviewing how “many States use causation of serious bodily injury or harm as an element defining a distinct offense of aggravated robbery”).
\item \textsuperscript{162} Duncan v. Louisiana, 391 U.S. 145, 149 (1968) (holding that the Fourteenth Amendment guarantees a right of jury trial in all state criminal cases which would come within the Sixth Amendment’s guarantee if they were tried in a federal court).
\item \textsuperscript{163} \textit{Id.} at 153 (discussing how every state joining the Union subsequent to formation had the right to a jury trial articulated in its constitution); see also Verrilli, \textit{supra} note 136, at 354.
\end{itemize}
Thus the historical right to bail, both as commonly practiced around the time of the Founding and afterwards in the various states, has much to teach us about how bail rights today should be understood and defined, particularly in light of the preventative detention proposals that are currently fashionable.

C. Recent Bail Reforms

The basic form of bail, relying on the personal surety as the custodian of the defendant, remained unchanged until the mid-nineteenth century. However, because these bail custodians had to be both known and acceptable to the courts, the personal surety system eventually morphed into the commercial bondsman system. This switch “substantially reduced the courts’ ability to assess the risks of pretrial release,” and—combined with the decreasing number of nonbailable crimes—added to the general trend whereby judges set high bails exceeding a defendant’s ability to pay. The imposition of high bail remained the status quo in the state bail world.

Federal bail remained relatively unchanged from the Judiciary Act of 1789 until 1966, when Congress passed the first bail reform act. The 1966 Bail Reform Act marked a return to conditional release, and was designed in large part to reduce the high bails imposed by judges to prevent release of certain defendants. The 1966 Bail Reform Act relied heavily on custodial supervision to ensure proper behavior, requiring judges to consider a variety of release conditions and release defendants under the most minimal

165. See id. (finding that the increased urbanization of American society made finding a custodian known and acceptable to the court far more difficult).
166. See id.
168. Id.
release strictures possible.\textsuperscript{170} A major goal of the 1966 Act was to reduce pretrial flight.\textsuperscript{171} Pretrial detention based on future dangerousness was not envisioned.\textsuperscript{172}

The difficulties of successfully implementing the 1966 Bail Reform Act, such as setting the terms of release and ensuring that conditions were met, along with worries about the crimes committed by defendants out on conditional release, led to the passage of the 1984 Bail Reform Act (BRA).\textsuperscript{173} This federal statute was paralleled on the state level by no fewer than thirty-four states articulating specific statutory provisions allowing detention based on a defendant’s dangerousness, as opposed to a risk of flight.\textsuperscript{174}

The BRA was predicated on protection of the public and community safety, making this factor one of the most critical in the determination of whether to release or detain defendants before trial.\textsuperscript{175} Most states have followed the path of the BRA, with forty-five states and the District of Columbia specifically permitting the determination of dangerousness as a predicate for denying pretrial release.\textsuperscript{176}

\textsuperscript{170} See 18 U.S.C. § 3142 (1966) (allowing release upon conditions such as personal recognizance; execution of an unsecured appearance bond; third party custody; travel, association or living restrictions; execution of an appearance bond; and/or execution of a bail bond).


\textsuperscript{172} See 18 U.S.C. § 3142(b) (1988).

\textsuperscript{173} See Miller & Guggenheim, supra note 15, at 344 (stating that crime committed by persons on pretrial release was a major concern for legislators after the 1966 Bail Reform Act).

\textsuperscript{174} Id.

\textsuperscript{175} The Act provides, among other things, that defendants should be granted bail “unless . . . such release will not reasonably assure the appearance of the person . . . or will endanger the safety of any other person or the community.” 18 U.S.C. § 3142(b) (1988).

\textsuperscript{176} See Baradaran & McIntyre, supra note 15, at 507. As Baradaran and McIntyre note,

In determining whether the accused is too dangerous to release prior to conviction, state courts consider three main categories: (1) the circumstances surrounding the present offense charged, (2) the defendant’s past conduct, and (3) judicial discretion regarding the defendant’s circumstances and character. Many states use the first two categories in an attempt to objectively determine which defendants pose a risk to public safety.

\textit{Id.}
Technically, the 1984 Bail Reform Act reaffirmed the idea that pretrial release was to continue to be the norm. However, the 1984 Act lacked neutrality regarding the determination of future dangerousness. For example, the Act contains little to balance out the reliance on predicting dangerousness for the defense side. Evidence of other crimes may be presented as hearsay, which is not subject to cross-examination. There is no notice to defendants that prosecutors may seek pretrial detention based on prior crimes or behavior. Additionally, the 1984 Act does not require that there be any confrontation between the defendant and the prosecutor who proffers the evidence.

Moreover, purporting to be deeply concerned with “community safety,” the 1984 Bail Reform Act allows federal prosecutors to request pretrial detention for any felony committed after two or more convictions of federal or state crimes of violence. The Act contains a rebuttable presumption favoring detention whenever a defendant has a prior conviction of a violent crime less than five years prior and was arrested while on conditional release pending trial for another offense. The BRA grants authority to the courts to confine an indicted individual based on “the danger a person may pose to others if released.”

Additionally, under the 1984 BRA, there is no requirement of evidence of a substantial possibility of the defendant’s guilt. And there are no limits on the length of detention beyond the requirements of the Speedy Trial Act and the vague limits of the Sixth Amendment. In sum, the requirements for pretrial detention under the 1984 Act only require a quick hearing with a

179. See Miller & Guggenheim, supra note 15, at 347.
184. See Miller & Guggenheim, supra note 15, at 348 (stating that Congress expressly rejected the District of Columbia Act’s requirement of evidence of a substantial possibility of the defendant’s guilt).
few limited procedural protections, primarily focusing on whether the defendant’s prior acts or convictions make it “necessary” to deny bail for community safety purposes.\textsuperscript{185}

Finally, applying the seven-part test laid out by the Court in \textit{Kennedy v. Mendoza-Martinez}\textsuperscript{186} to determine whether an Act of Congress is penal or regulatory in character proves that the 1984 BRA, as it operates today, is fairly punitive in nature. The \textit{Mendoza-Martinez} test has seven distinct questions to determine the character of a Congressional Act:

\begin{quote}
[W]hether it involves an affirmative disability or restraint, whether it has historically been regarded as a punishment, whether it comes into play only on a finding of \textit{scienter}, whether its operation will promote the traditional aims of punishment—retribution and deterrence, whether the behavior to which it applies is already a crime, whether an alternative purpose to which it may rationally be connected is assignable for it, and whether it appears excessive in relation to the alternative purpose assigned.\textsuperscript{187}
\end{quote}

Applying \textit{Mendoza-Martinez} to the 1984 BRA illustrates its largely punitive nature. Five out of its seven factors point to this conclusion. First, the Act involves an obvious restraint: pretrial detention. Second, detention has been traditionally regarded as a punishment in this country,\textsuperscript{188} so much so that bail was historically required in all but the most heinous of charged crimes. Third, pretrial detention cannot promote either retribution or deterrence, because it is imposed before conviction. Fourth, the behavior to which the Act applies may be a crime, but that fact has not yet been determined, by either a jury or a judge.

\textsuperscript{185} See \textit{id.} at 349 (detailing the shortcomings of the act).
\textsuperscript{187} \textit{Id.} at 168–69.
\textsuperscript{188} Two glaring exceptions to this general rule, of course, are the administrative detentions of immigration holds and continuing civil incarceration of sex offenders. See, e.g., \textit{Matter of Sanchez}, 20 I&N Dec. 223, 225 (BIA 1990) (characterizing an immigration detainer as “merely an administrative mechanism to assure that a person subject to confinement will not be released from custody until the party requesting the detainer has an opportunity to act”) (citing \textit{Moody v. Daggett}, 429 U.S. 78, 80 n. 2 (1976); see also \textit{United States v. Comstock}, 130 S. Ct. 1949 (2010) (holding that the federal civil-commitment statute authorizes the Department of Justice to detain a mentally ill, sexually dangerous federal prisoner beyond the date the prisoner would otherwise be released).
On the other hand, the Act applies with no regard to scienter, just a mere showing of probable cause by the prosecutor. This disregard for criminal intent points to a more neutral, non-punitive rationale for the Act. Moreover, it can be argued that the stated alternative purpose for pretrial detention—incapacitation—is rational, although tremendously overused. Overall, however, putting the Act through the seven Mendoza-Martinez questions shows that it is currently penal in character.

Despite the Act’s flaws, the Supreme Court upheld the 1984 Bail Reform Act against a substantive due process facial challenge in United States v. Salerno. Decided on narrow grounds, Salerno concluded that the Act was not unconstitutional in its determinations weighing the defendant’s interest in liberty against the government’s interest in community safety. Rejecting the Southern District of New York’s reasoning that our criminal justice system can only hold persons accountable for past actions, not anticipated future ones, the Salerno Court found that merely detaining a person “does not inexorably lead to the conclusions that the government has imposed punishment.” The Salerno Court’s conclusion was based on its belief that the regulatory goal that Congress sought to achieve in the 1984 BRA was not punishment, but public safety.

The Salerno Court carefully noted that it was only looking at the 1984 Bail Reform Act as created by Congress, not as actually applied. As such, it reserved the right to decide the point “at which detention in a particular case might become excessively prolonged, and therefore punitive, in relation to Congress’ regulatory goal.” It is possible that by doing so, the Court was signaling that it would prefer to wait and strike down a particular detention order when the defendant could show his interest in liberty outweighed the state’s interest in community safety.

190. See id. at 741 (“We hold that, as against the facial attack mounted by these respondents, the Act fully comports with constitutional requirements.”).
191. Id. at 745.
192. Id. at 746.
193. Id. at 747.
194. Id. at 747 n.4.
safety. Nonetheless, the Salerno Court’s upholding of the Bail Reform Act struck a blow to concepts of retributive criminal justice (the belief that a wrongdoer can only be punished for crimes he or she has actually committed).

The Salerno Court, however, used some sleight of hand between the actual language of the 1984 Bail Reform Act and the way it justified the Act. One of the ways the Court defended the idea of pretrial detention exclusive of the risk of flight was by arguing that the government had a legitimate and compelling interest in preventing crime by arrestees.

Preventing future crime, though, is a different endeavor than public safety for the community, the purported reasons behind the 1984 BRA. Although of course crime prevention does, in a very general sense, enhance public safety, few crimes are so dangerous that their very potential requires detention of a suspect. And the most dangerous of them, murder, is usually barred from bail release in any case. Nonetheless, the Salerno Court easily conflated future crime prevention and community safety into one amorphous concept. Moreover, as other scholars have noted, the Court evaded the BRA’s underlying problem of identifying the actual circumstances that transform detention into punishment.

Salerno also specifically addressed two constitutional claims involving the Fifth and Eighth Amendments, quickly dismissing them both. Regarding the Fifth Amendment substantive due process claim—that pretrial detention constituted impermissible punishment before trial—the Court held that because the

---

195. See Miller & Guggenheim, supra note 15, at 351 (“[T]he Court would likely prefer instead to wait and declare unconstitutional any particular detention order in which the defendant could show that the state’s interest in community protection failed to outweigh his or her interest in liberty.”).

196. See United States v. Salerno, 481 U.S. 739, 748–49 (1987) (stating that the government may utilize pretrial detention when its regulatory interest in community safety outweighs an individual’s right to liberty and giving examples).

197. See Miller & Guggenheim, supra note 15, at 353 (“The Court avoided, however, the underlying problem of identifying circumstances that make detention punishment.”).

198. See Salerno, 481 U.S. at 746, 751–52 (rejecting a facial challenge under the Fifth and Eighth Amendments).

199. See id. at 746 (summarizing respondents’ argument).
legislative intent of Congress in the Bail Reform Act was not punitive, the pretrial detention was not punishment, but regulation.200

As for the Eighth Amendment claim—that the 1984 Bail Reform Act violated the Excessive Bail Clause because this clause grants a defendant the right to bail based solely on the considerations of flight201—the Salerno Court flatly rejected this argument, holding that nothing in the text of the Bail Clause limits bail decisions solely to questions of flight.202 However, the Salerno Court did not provide any historical evidence to support this conclusion about the Eighth Amendment, simply leaving the assertion to stand alone.203 As discussed briefly above,204 although the specific intent of the Framers regarding bail cannot be conclusively determined, all the available evidence points to the fact that pretrial detention, both under English common law and at the time the Constitution was written, was limited to flight risks. Thus, the Salerno Court’s rejection of the Eighth Amendment challenge on this basis is undersupported at best.

Given the changes made to the historical right to bail, our newfound reliance on preventative incarceration, and the Supreme Court’s recent focus on the Sixth Amendment’s jury trial right, our current pretrial detention procedures may require some substantive changes. In Part IV, I explore the problem with our current system’s reliance on future dangerousness to routinely imprison indicted offenders and contend that this violates our understanding of the role of punishment as dictated by the Sixth Amendment.

200. Id. at 747.
201. See id. at 752 (summarizing respondents’ argument).
202. Id. at 754.
203. Id. at 753. The Salerno Court did carve out a space to decide later whether “the Excessive Bail Clause speaks at all to Congress’ power to define the classes of criminal arrestees who shall be admitted to bail,” but maintained the validity of the Bail Reform Act even then. Id. at 754.
204. See supra Part III.B.
IV. Preventative Detention, Future Dangerousness, and the Sixth Amendment

There is convincing evidence that modern-day bail and jail practices result in punishment for the indicted defendant incarcerated before trial. Although usually the underlying reasoning for the imposition of pretrial detention is not based on a punishment rationale, the consequences of such decisions are often so severe that the end results are punitive.

This punishment before conviction creates numerous problems for our current bail and jail structure. First, many offenders are denied bail based on the relatively new field of preventative detention, which is one riddled with errors, both in theory and in practice. Second, although Salerno has seemingly closed off both due process and Eighth Amendment attacks against the 1984 Bail Act, there have not yet been any challenges based on the Sixth Amendment ban on punishment imposed before a conviction and without a jury’s imprimatur. Third, and relatedly, if we follow the dictates of the Sixth Amendment jury trial right, the community must take part if any punishment is to be imposed on an offender. Below I explore how these factors—the mistaken reliance on preventative detention, the holes left by Salerno, and the requirements of the Sixth Amendment—all converge to make our current pretrial detention hearings in need of reform.

A. The False Promise of Preventative Detention

In the federal system, preventative incarceration, or detaining the accused based on the potential of future crime, did not become popular until the 1984 Bail Reform Act. In the last twenty years or so, many states have followed suit, allowing their criminal justice systems to detain the indicted individuals based on their future Dangerousness. Both judges and academics have challenged the bases for this type of determination, however, putting the entire theory to question.

1. Barefoot v. Estelle

One year before Congress passed the 1984 Bail Reform Act, the Supreme Court discussed a similar issue regarding the propriety of nonjury actors determining and punishing for future dangerousness. In *Barefoot v. Estelle*, the Supreme Court agreed with the lower trial court that the use of psychiatric experts to discuss the potential dangerousness of the defendant *at trial* was permissible because this type of determination was for the jury to decide: “Such disputes are within the province of the jury to resolve. Indeed, it is a fundamental premise of our entire system of criminal jurisprudence that the purpose of the jury is to sort out the true testimony from the false, the important matters from the unimportant matters . . . .” Put another way, the majority of the *Barefoot* Court underlined the importance of the community’s role in deciding whether a fellow member of the public is so dangerous as to deserve incarceration on that basis. This decision, however, focused on the propriety of determining dangerousness during an actual trial, not during a pretrial detention hearing.

More important for our purposes, the dissent in *Barefoot* highlighted the role of the jury as the proper arbiter of decisions involving a defendant’s incarceration. Penned by Justice Marshall, the dissent noted that psychiatrists and other experts might actually be “less accurate predictors of future violence than laymen,” in part because the lay public lacks a personal bias leaning towards predicting violence, which can arise from being responsible for the erroneous release of a violent individual. If this is true for psychiatric experts, it is also likely to be true for magistrates and trial judges, especially those state court judges who must submit to the pressures of periodic re-election.

The dissent also focused on a key critique that applies widely to all determinations of potential future dangerousness: the

206. See *Barefoot v. Estelle*, 463 U.S. 880, 903 (1983) (finding that “[e]xpert testimony, whether in the form of an opinion based on hypothetical questions or otherwise, is commonly admitted as evidence where it might help the factfinder do its assigned job”).
207. *Id.*
208. *Id.* at 922 n.4 (Marshall, J., dissenting).
209. *Id.* at 922 n.4.
general unreliability of these predictions.\textsuperscript{210} Citing several studies, the dissent pointed out that long-term prediction of future violence by psychiatrists continues to be extremely inaccurate.\textsuperscript{211} This is due in part to the difficulty of identifying any sub-class of offenders who have a greater than fifty-fifty chance of re-engaging in assaultive conduct,\textsuperscript{212} the conduct with which future dangerousness is most concerned. Indeed, a ninety percent error rate is common.\textsuperscript{213}

All these concerns are focused, of course, on testimony given to the jury by expert psychiatric witnesses during an actual trial. How much more unreliable are those hasty predictions by an untrained magistrate or trial judge, determining future dangerousness of a person who has not yet been convicted? At the minimum, the incarceration of an accused individual due to his potential to commit more crimes should be based on stronger science than gut feelings or past conduct, and should admit some aspect of community participation.

\section*{2. Preventative Detention’s Binary Nature}

The unreliability of accurately determining future dangerousness, however, is not the only problem with current imposition of pretrial detention. As other scholars have argued, our present pretrial detention model is extremely binary, refusing to account for gray areas.\textsuperscript{214} Gray areas, however, repeatedly occur in determining eligibility for pretrial release: “[A] binary model requires a decision maker to round off the evidence and

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{210} See \textit{id.} at 920 (discussing how “the unreliability of . . . predictions of long-term future dangerousness is by now an established fact within the profession,” and noting that “two out of three predictions of long-term future violence made by psychiatrists are wrong” (citing Brief for American Psychiatric Association as Amicus Curiae Supporting Petitioner at 12, 463 U.S. 880 (1983) (No. 82–6080))).
\item \textsuperscript{211} \textit{id.} at 920.
\item \textsuperscript{212} \textit{id.} (citing Wenk, Robinson & Smith, \textit{Can Violence Be Predicted?}, 18 \textit{CRIME \& DELINQ.} 393, 394 (1972)).
\item \textsuperscript{213} \textit{id.} at 921 n.2.
\item \textsuperscript{214} See Jack F. Williams, \textit{Process and Prediction: A Return to a Fuzzy Model of Pretrial Detention}, 79 \textit{MINN. L. REV.} 325, 327 (1994) (“The prevalence of pretrial detention is largely a function of our bivalent system of law . . . .”).
\end{enumerate}
\end{footnotesize}
confine the case to total truth or no-truth to make a decision."215 Accordingly, our current bail system fails to account for the varying degrees of detainability, dangerousness, and culpability that indicted offenders present.216

Due to the limitations of the binary model, many of the judicial determinations that lead to pretrial detention are overly harsh or punitive.217 This is especially true when it comes to predicting future dangerousness. Neither experts nor courts have had much success in accurately determining if and when an offender might commit more crimes.218 Granted, there have been some more recent studies that have provided a far better prediction rate than the older evidence.219 Despite the existence of such new predictive materials, however, many harmless defendants are still unfairly detained as dangerous based on old or outdated beliefs.220

The term dangerous itself can be quite vague when it comes to detaining pretrial defendants. The 1984 Bail Reform Act failed to define the term at all.221 Thus the idea of determining dangerousness to the community is an incredibly broad concept, which could encompass almost anything, from physical danger to conspiracy. The problem with this imprecision of terminology is that accurately determining dangerousness requires a narrow focus, which the determination of bail so notably lacks.

215. Id.
216. Id. at 327–28.
217. See id. at 327 (criticizing the binary model for its inability to take into account "partial degrees of truth").
219. See generally Baradaran & McIntyre, Predicting Violence, supra note 15, passim.
All of this imprecision means that judging future dangerousness within our current pretrial detention scheme can be quite arbitrary. Forcing a black or white decision onto a mass of gray evidence (particularly since it is evidence that has not yet been proven beyond a reasonable doubt) makes many bail hearings unreliable and inconsistent.

3. Adjudicating Dangerousness

The problems with determining future dangerousness, however, do not end there. Even if future dangerousness can be accurately predicted, should that necessarily mean that dangerousness should always equal detention and punishment?

Although the question of dangerousness has been widely discussed in terms of sentencing, the scholarly exploration of the topic, both procedurally and jurisprudentially, has been rather limited as applied to pretrial detention. It is this gap that I aim to fill.

Norval Morris famously addressed using predictions of future behavior to determine whether a criminal should be imprisoned after conviction.222 As he noted, “as a matter of justice we should never take power over the convicted criminal on the basis of unreliable predictions of his dangerousness.”223 Morris was concerned that dangerousness was so expansive a concept that “the punitively minded” would use it to classify all offenders, deserving or not.224

Moreover, as other scholars have observed, dangerousness is “peculiarly seductive” because it can be ascribed as a personal characteristic of the offender, not a judicial imposition.225 And when dangerousness is seen as a personal trait, it leads to confusion between the determination of dangerousness and the determination of desert, when both usually animate the reasons underlying punishment.226 Likewise, Andrew von Hirsch has

223. Id. at 73.
224. Id. at 72.
226. Id.
argued that the moral argument against predictive detention “is that it is not deserved. This objection stands even were the prediction of future criminality accurate.”

If all these objections exist for the convicted defendant, how much more do they resonate for the offender who is only indicted? In 1968, before the passage of the 1986 Bail Reform Act, the American Bar Association rejected pretrial preventative detention, even for “dangerous” offenders, because too little was known about the actual need for this type of detention and of the predictive techniques used.

Although this ship has clearly sailed, the theoretical issues still remain.

Whether one’s take on desert is animated by limits, proportionality, or by the parity principle, the theory of desert is simply inapplicable to pretrial detention because no determination of crime has yet been made. Predictive dangerousness, when not based on accurate, up-to-date empirical evidence, has no place in any rational system of retributive justice, and yet it is a commonplace determination in American bail hearings, where it seems least appropriate.

Although the Supreme Court in Salerno rejected both the procedural and substantive due process claims, it failed to discuss the more theoretical problems with judging future dangerousness. The Salerno Court took for granted that dangerousness is a fixed term with a fixed meaning, as opposed to its actual amorphous nature, difficult to chart or pin down. As such, it is truly inappropriate for use at the pretrial detention hearing, at which judges have neither the time nor the

227. ANDREW VON HIRSCH, DOING JUSTICE: THE CHOICE OF PUNISHMENT 125 (1976). Granted, von Hirsch admits a small fraction of offenders should be confined by preventative detention, including those who have extensive violent records and who were convicted of serious assault crimes. Id. at 125–26.

228. See Zimring & Hawkins, Dangerousness and Criminal Justice, supra note 225, at 496 n.28 (citing ABA Standards Relating to Pretrial Release § 5.5 commentary at 69 (1968)).


230. See id. at 750–51 (rejecting the claim that an individual’s right to liberty always outweighs the government’s interest in protecting the community).
inclination to theorize about how desert might be applied to offenders who have not yet been convicted.231

This uncertainty and ambiguity of standards in pretrial detention matters because determining dangerousness is a central preoccupation of the criminal justice system.232 In part because of this, the reliability of preventative detention is a hotly debated topic. As Christopher Slobogin has noted, there is a two-part challenge to the reliability of predictive dangerousness: (1) persons should not be denied liberty on dangerousness grounds unless there is a high degree of certainty that the person will offend in the near future; and (2) this sort of proof of dangerousness is nearly impossible to obtain.233 Responses to these concerns have included the charge that even proof beyond a reasonable doubt has its own unreliabilities as an indicator of culpability and predictor of future dangerousness.234

For example, the Barefoot Court held that the inconsistencies inherent in predicting future dangerousness, even in regards to the death penalty, are largely eradicated by the adversarial process, which usually exposes erroneous views.235 As the

---


232. See Christopher Slobogin, A Jurisprudence of Dangerousness, 98 Nw. U. L. Rev. 1, 2 (2003) (stating that dangerousness assessments play a role in “death penalty determinations, non-capital sentencing, sexual predator commitment, civil commitment, pretrial detention, and investigative stops by the police”).

233. Id. at 3.

234. See id. at 7–8. As Slobogin explains in greater detail:

First, imposition of the reasonable doubt standard [for predictive detention] is overly stringent when the state’s goal is to prevent rather than to punish. Second, the belief that the criminal law permits conviction only when there is no reasonable doubt about blameworthiness is based on a misconception about the reliability of assessments made in criminal cases; in fact, the culpability determinations that provide the primary basis for criminal punishment are subject to serious inaccuracy. Third, requiring a high degree of danger is inconsistent with the fact that many of the crimes that penalize dangerous activity require very little in the way of predictive validity.

Id. at 6–7.

argument goes, if we can rely on the shaky credibility of predictive dangerousness to justify an execution, how could it not justify regular pretrial detention, where no life is lost?

The answer, of course, is specific to pretrial detention. First, there is no adversarial contest in the typical pretrial detention hearing, as there is no right to appointed counsel during bail determination. As the vast majority of those detained without bail are those who cannot afford counsel on their own, it is rare to see defense counsel appear at these hearings. Accordingly, the prosecutor usually presents her reasons why the indicted offender should not be granted bail, with no response by the defense, and the judge decides.

Second, there is a large difference between assessing future dangerousness for a convicted offender, who is subject to punishment of some kind, and assessing the same for someone who has not even been subject to conviction. Taking away the liberty of a person convicted beyond a reasonable doubt (or who has admitted his or her guilt) is a far more acceptable matter than imprisoning someone whose very guilt is still in doubt.

Finally, as Paul Robinson has convincingly argued, “[i]t is impossible to punish for dangerousness.” This is because, in both theory and actuality, deserved punishment can only exist in relation to actual wrongs done, not potential or imagined future wrongs. In other words, “one can restrain, contain, or incapacitate a dangerous person, but one cannot logically punish dangerousness.” Thus it is not only unfair but theoretically unsound to punish indicted offenders with pretrial detention for


238. Id.

239. Id. at 1432.
their potential future dangerousness. Dangerousness and desert are two very different concepts, and should not be conflated together in the realm of pretrial detention.

4. Empirical Evidence and National Trends

The actual work of determining dangerousness is a chancy business, filled with pitfalls and indeterminacy. To predict the dangerousness of the defendant, courts tend to analyze the nature of the charged crime and combine this information with their knowledge of his or her past conduct. The court then adds to this assessment its own determination of the accused's circumstances and character. In many states, a more subjective judicial assessment permits courts to consider the totality of the defendant's circumstances and character. This aspect of determining dangerousness, then, is highly influenced by the court's personal feelings and quirks. Despite the claim for scientific accuracy, the determination of dangerousness is far more based on subjectivity than objective factors.

As a whole, predicting pretrial crime is a dubious science. In fact, our assumptions about who might be mostly likely to reoffend before trial are often not borne out. For example, as demonstrated by one of the few large empirical studies done on defendants released before their trials, those charged with violent crimes are not necessarily more likely to be rearrested pretrial.

240. Id. at 1438.
241. See Baradaran & McIntyre, supra note 15, at 508–10. As Baradaran and McIntyre show, there are five states that allow for an even deeper look into a defendant's background by allowing judges to factor the defendant's past conduct into their determination. Id. at 511.
242. See id. (noting that this factor is much broader to allow for judicial discretion).
243. See id. at 510. As the authors note, some state statutes include a list of factors with an “including but not limited to” clause, or permit judicial officers to consider “any other factor” relevant to making a determination of dangerousness. Id. at 511.
244. Id. at 523.
245. See id. at 528. As the authors point out:
The highest rearrest rates pretrial are for defendants charged with drug sales or robbery (21%), followed by motor vehicle theft (20%), and burglary (19%). Those released who are charged with the 'more
Nationally, 16% of defendants released on bail are rearrested for any reason, and 11% are rearrested for a felony. Particularly important for our purposes, only 1.9% of all defendants released on bail before trial are rearrested for a violent felony. Thus, one prong of the furor over “dangerous” indicted defendants is blunted; if only a tiny percentage of all defendants released on bail go on to reoffend with violent crimes, then perhaps our fear over the danger posed by the pretrial defendant to the community is overblown.

Moreover, this same study shows that those charged with violent crimes are not necessarily more likely to be rearrested pretrial. Critically, the research done by Baradaran and McIntyre illustrates that although those defendants charged with violent crimes have the highest likelihood of being rearrested on bail, there is still huge variation in how dangerous these violent crime defendants can be, depending on the specific crime charged. Thus, simply being charged with a violent crime does not automatically create a presumption of dangerousness as many courts believe. Ultimately, despite the large variety of assessments of pretrial “dangerousness,” the defendants granted bail before trial are often far less threatening to public safety than most people would anticipate.

Interestingly, the greatest predictor for future pretrial crime, dangerous or not, is the existence of past arrests. It is dangerous crimes,’ such as murder, rape, and felony assault, have much overall lower rates of pretrial rearrest at 12%, 9% and 12% respectively.

Id. at 527.

Id.

Id. at 528. For example, those defendants charged with murder have a 6.4% violent crime rearrest rate, one of the highest. Similarly, those defendants charged with robbery have a 5.8% chance of rearrest for violent crime, those defendants charged with rape at 3.2% chance of rearrest for violent crime, and those defendants charged with assault reoffend at a rate of 2.9%. See id. at 528–29.

Id. at 529.

Id. at 536 (“A person’s number of previous arrests is a large predictor of future rearrest; however, whether or not that prior arrest turned into a conviction is largely irrelevant as an additional predictor.”).
important to note that analysis of prior convictions shows that even defendants with multiple prior convictions are still unlikely to be rearrested for a new violent crime while on release.\textsuperscript{252}

This information can help us find a better way of predicting “dangerousness” than our current fumbling in the dark. The charged crime, by itself, is a poor predictor of a threat to the community, except for the most violent ones (such as murder, which usually is statutorily prevented from bail release in any case). The existence of multiple past convictions for similar crime seems to have the most predictive effect, although even those pretrial recidivism rates are low.

Accordingly, relying on our current system of judicial and prosecutorial decision-making regarding pretrial release—which relies primarily on the charged crime—is not only unfair, but largely ineffective. And considering that roughly 62\% of the overall jail population consists of pretrial detainees,\textsuperscript{253} these decisions make a huge difference. Accordingly, if we are going to continue to use predictions of future dangerousness to determine the imposition of pretrial detention, as is likely, we should at least provide courts with the best and latest empirical evidence on the subject.

Finally, on the broadest level, post-9/11 case law has also raised important questions about the permissible scope of all pretrial confinement.\textsuperscript{254} Do the 9/11 detainee cases affect the law governing other detainees, whether held in state or in federal custody, either before or after conviction?\textsuperscript{255} At least one prominent scholar has answered the question affirmatively,

\begin{enumerate}
\item[252.] See id. (finding that only 5\% of all defendants have more than a 5\% chance of being rearrested on a violent felony charge when released on bail).
\item[253.] See id. at 37 (finding that in 2007, 62\% of the overall jail population consisted of pretrial detainees).
\item[254.] See Judith Resnik, \textit{Detention, the War on Terror, and the Federal Courts: An Essay in Honor of Henry Monaghan}, 110 COLUM. L. REV. 579, 586 (2010) (“The 9/11 case law has prompted diverse assessments, with arguments that the judiciary has done too much, or too little, or left unanswered important questions about the permissible scope of executive detention and surveillance powers.”).
\item[255.] See id. at 583. Resnik points out that “As [Henry] Monaghan noted, these questions are at the core of the shifting conception of federal courts jurisprudence, once preoccupied with the ‘relationship between state and federal law’ and the sometimes ‘irritating difficulty’ of sorting between the two kinds.” \textit{Id.} at 583 n.19. These issues arise with state courts as well.
\end{enumerate}
arguing that the law created in addressing the extraordinary
detainment of prisoners in Guantánamo Bay is “continuous with
judicial responses to the central challenges, faced daily, by
governments trying to maintain peace and security and, hence,
incapacitating some individuals feared likely to inflict grave
harm to the social order.”  This is because the criminal justice
system has always had to address uncertainty about how much
harm a detainee might do, whether in the context of 9/11 or of
more familiar kinds of criminality and border regulation.

Put another way, criminal charges range in a wide
continuum from minor crimes to terrorism. In all these cases,
however, the government “must still distinguish among and
classify detainees to justify why a particular subset is to be
confined in more restrictive conditions than others, and for longer
periods of time.” The crimes may differ, but the determinations
are still the same.

The six post-9/11 cases decided by the Supreme Court have
held that the Constitution requires some procedural justice for all
detainees, even those held at Guantánamo Bay. Applied to
pretrial detention, this conclusion signals that it is time to import
some fairness and procedural justice into the bail hearing, since it
is part of the same continuum of pretrial detention.

Integrating emerging 9/11 law with state and local laws
governing confinement is an important task. Doing so highlights,
among other things, the state’s job in addressing serious
challenges in securing safety, whether locally, nationally, or
worldwide. As Judith Resnik has pointed out, “[s]orting the
dangerous from the benign is a daunting task.” This is
particularly true because neither courts nor legislatures have

256. Id. at 584.
257. Id. at 585. As Resnik notes, “Governments regularly desire to obtain
information through intense interrogations aimed at preventing injuries and at
apprehending wrongdoers, and governments regularly detain various persons.
Courts in turn have, over the last several decades, ruled many times on the
legality of detention and of confinement conditions.” Id. at 584.
258. Id.
259. Id. at 581.
260. See id. at 587 (“[T]he state regularly faces tremendous challenges in
securing safety, at both local and global levels.”).
261. Id.
been able to come to any sort of consensus over the years regarding who should be detained and who may be freed.\textsuperscript{262}

For example, in \textit{Ashcroft v. Iqbal},\textsuperscript{263} the defendant, a Pakistani national, was detained for almost a year in the Metropolitan Detention Center in New York. Part of that time he was in solitary confinement, because of a claim that his identity papers were false.\textsuperscript{264} Iqbal was deported to Pakistan after his period of governmental detention.\textsuperscript{265} Iqbal sued the government for his treatment period of confinement, during which he alleged that he was subjected to cruel and inhumane conditions.\textsuperscript{266} The \textit{Iqbal} Court held, among other things, that courts could limit individual accountability and civil liability for the harms imposed during detention, thereby protecting those officials who were involved in such detention programs.\textsuperscript{267} Although \textit{Iqbal} is obviously a case that can be classified as a post-9/11 terrorism case, its disturbing lesson resonates for all pretrial detainees, and illustrates how the lack of proper bail procedures can extend to defendants both high and low.

Ultimately, similar problems plague both the general procedures of pretrial detention and the small body of 9/11 law. This includes the existence of only a tiny batch of procedural remedies instead of a more robust body of constitutional constraints in response to wide-ranging complaints of abuse.\textsuperscript{268} The lax oversight, minimal supervision, and extreme deference to governmental decisions and jailors continue to be issues for 9/11 detainees, detained aliens on immigration holds, and indicted defendants confined in pretrial detainment. All three are held with little procedural justice due to often nebulous fears of dangerousness. All three types of experiences linger in the arena of pretrial detention.

\textsuperscript{262} See id. at 588 (discussing the historical conflicts between legislature and judiciary).
\textsuperscript{263} See Ashcroft v. Iqbal, 556 U.S. 662, 672 (2009) (finding that “qualified immunity . . . shields Government officials”).
\textsuperscript{264} See id. at 667.
\textsuperscript{265} See id. at 668.
\textsuperscript{266} Id.
\textsuperscript{267} Resnik, supra note 254, at 633.
\textsuperscript{268} Id. at 635.
B. Salerno, the 1984 Act, and the Sixth Amendment

The problems with Salerno are larger than just its failure to grapple with the concept of future dangerousness. Viewed in the aftermath of the Court’s subsequent decisions in the Apprendi-Blakely line of cases, Salerno’s decision that a court may incarcerate a defendant on the basis of potential danger to the community seems to contradict the spirit of the Sixth Amendment. If, as I contend, the imposition of pretrial detention in today’s jails is a form of punishment, then the community must have some say in the matter. Although there have been numerous challenges to the 1984 Bail Reform Act, none have been based on the Sixth Amendment jury trial right. Below, I detail how and why such a challenge might be successful.

1. Salerno Did Not Close the Door

Most practitioners and scholars have concluded that the Supreme Court’s decision in Salerno pronounced the death knell for challenges to preventative detention. This is not entirely true. The Salerno Court specifically noted that it was foreclosing a facial challenge to the 1984 Bail Reform Act,269 not foreclosing all challenges for the future: “We hold that, as against the facial attack mounted by these respondents, the Act fully comports with constitutional requirements.”270 As such, an as-applied challenge to the 1984 Act is still entirely feasible.

The Salerno Court seemed to leave more than one door open to indicate its willingness to revisit its decision upholding the constitutionality of the 1984 Act. First, it began the analysis of the Act with a warning that a facial challenge to a legislative act was one of the most difficult at which to succeed, since the challenged “must establish that no set of circumstances exists under which the Act would be valid.”271 In other words, for the

---

269. As the Salerno Court held, “We are unwilling to say that this congressional determination . . . on its face violates either the Due Process Clause of the Fifth Amendment or the Excessive Bail Clause of the Eighth Amendment.” United States v. Salerno, 481 U.S. 739, 755 (1987).

270. Id. at 741.

271. Id. at 745 (emphasis added).
defendants to succeed in challenging the 1984 Act on facial grounds, they would have had to prove that the Bail Reform Act could never operate constitutionally.272 The Salerno Court even hinted that it might be willing to find some aspect of the 1984 Bail Reform Act unconstitutional under certain circumstances, but simply not here: “The fact that the Bail Reform Act might operate unconstitutionally under some conceivable set of circumstances is insufficient to render it wholly invalid.”273

Moreover, the Salerno Court carefully carved out space for future, as-applied challenges to the Act, noting in a footnote that they “intimate[d] no view on the validity of any aspects of the Act that are not relevant to respondents’ case.”274 Although not precisely an invitation to bring an as-applied challenge, the Supreme Court certainly allowed for the future possibility.

2. Pretrial Detention as Punishment Under the 1984 BRA

Even in its rejection of the defendants’ facial challenge, the Salerno Court carved out an exception to its holding: when pretrial detention becomes punishment. Of course, simply because a defendant is detained does not mean that he is being punished.275 Traditionally, we look to legislative intent to determine whether a restriction on liberty, like pretrial detention, is more like punishment than like regulation.276 Unless Congress intended on imposing punishment, whether a restriction on liberty is classified as punitive or regulatory depends on whether there is an alternative purpose related to the restriction and whether this restriction seems excessive in

272. Id.
273. Id.
274. Id. at 745 n.3.
275. See Bell v. Wolfish, 441 U.S. 520, 537 (1979) (“Not every disability imposed during pretrial detention amounts to ‘punishment’ in the constitutional sense.”).
276. See Schall v. Martin, 467 U.S. 253, 269 (1984) (“Absent a showing of an express intent to punish on the part of the State, that determination generally will turn on ‘whether an alternative purpose to which [the restriction] may rationally be connected is assignable for it, and whether it appears excessive in relation to the alternative purpose assigned.’” (citing Kennedy v. Mendoza-Martinez, 372 U.S. 144, 168–69 (1963))).
relation to the original goal.\footnote{277} In \textit{Salerno}, the Supreme Court decided that the 1984 Act, addressed facially, had a legitimate regulatory goal (preventing danger to the community) and that the incidents of pretrial detention were not excessive in relation to that regulatory goal.\footnote{278}

Whether the Act’s legislative intent was regulatory or not, however, the constitutionality of the scheme is far different when the effect of pretrial detention results in punishment. The \textit{Salerno} Court itself reserved the right to decide on the constitutionality of a situation in which “detention in a particular case might become excessively prolonged, and therefore punitive, in relation to Congress’ regulatory goal.”\footnote{279} This was particularly important in this case because, as a facial challenge, \textit{Salerno} only addressed theoretical pretrial detention, not actual pretrial detention as experienced by indicted defendants.

Much of pretrial detention, whether based on fears for community safety or flight risk, has an effect virtually indistinguishable from punishment. Thus, despite \textit{Salerno}’s decision facially upholding the 1984 Act, there is nothing precluding a court from finding pretrial detention unconstitutional under certain punitive circumstances.

3. Applying the Sixth Amendment to the BRA

\textit{Salerno} rejected the facial challenge to the 1984 Act under both the Fifth and Eighth Amendments, but did not evaluate any Sixth Amendment claims. As such, there is nothing prohibiting an attack on \textit{Salerno}’s defense of predictive dangerousness based on the argument that only a jury can make factual findings that result in the imposition of punishment on an offender.

In \textit{Salerno}, the Court first tackled the Fifth Amendment claim, in which the defendants argued that the Act violated the substantive Due Process Clause because the pretrial detention it authorized constituted impermissible punishment.\footnote{280} The Supreme Court rejected this argument, focusing on the

\footnotesize{
279. Id. at 747 n.4.
280. See id. at 746 (summarizing respondents’ argument).
}
regulatory nature of the Act. Next, the Court addressed the Eighth Amendment argument, in which the defendants argued that the Act violated the Excessive Bail Clause. The defendants argued that the Excessive Bail Clause granted them a right to bail calculated solely upon considerations of flight. The Salerno Court also rejected this argument, holding that “nothing in the text of the Bail Clause limits permissible Government consideration solely to questions of flight.”

Salerno, however, did not address the Sixth Amendment implications for the 1984 Act, as the defendants did not raise the issue on appeal. Additionally, a Sixth Amendment claim at the time of the Salerno decision, in 1986, would have been fruitless. Now, however, after the Apprendi-Blakely line of cases, an application of the current understanding of the Sixth Amendment jury trial right, focusing on the community’s right to be the arbiter of punishment, makes more sense.

Although it did not address the issue, Salerno left room for the application of our more recent understanding of the Sixth Amendment. First, in its discussion of the Due Process Clause challenge, the Salerno Court noted that even the Government had not argued that pretrial detention could be upheld if it were punishment. Second, as noted above, the Court carved out an exception for cases in which detention might become “excessively prolonged,” thereby converting to punishment.

In its discussion of the Eighth Amendment excessive punishment challenge, the Salerno Court was careful to note that its decision did not implicate the question of whether the Excessive Bail Clause affects the legislative power to delineate who might be eligible for bail. In this way, the Court left room to determine whether and how pretrial detention can be punishment. It is this space that I will explore through the dimensions of the Sixth Amendment community jury trial right.

281. See id. (concluding that the detention imposed by the Act is regulatory, not punitive).
282. See id. at 752–53 (summarizing respondents’ argument).
283. Id.
284. Id.
285. Id. at 746.
286. See id. at 749 n.4.
287. See id. at 754.
C. Punishment, Community Rights, and Pretrial Detention

In *Blakely v. Washington*, the Supreme Court held that the jury is the only body that can find facts that increase the maximum punishment for an offender. Put another way, *Blakely* contended that a criminal offender must have a jury, or the local community, make the determination to impose any type of punishment. Our revitalized understanding of the Sixth Amendment jury trial right applies to all aspects of our criminal justice system, from indictment to criminal fines post-prison release. In fact, our Sixth Amendment constitutional guarantees should apply with more force to the pretrial stage, since at this point the offender still maintains the presumption of innocence. Yet, the spirit of the Sixth Amendment jury trial right has yet to be applied to the pretrial detention determination stage.

1. Future Dangerousness in Adversarial Context

When we apply the spirit of the Sixth Amendment jury trial right to the pretrial detention hearing, many problems with our bail procedures are illuminated. One such problem is the lack of adversarial context when determining the future dangerousness of an indicted offender. Since the adversarial process is crucial for the proper community understanding and imposition of punishment, the lack of it raises equity and possibly even constitutional concerns.

Although the 1984 Bail Reform Act does permit a federal indicted defendant the right to legal assistance, the right to cross-examine prosecution witnesses, and the right to testify and

---

288. See *Blakely v. Washington*, 542 U.S. 296, 296 (2004) (“Because the facts supporting petitioner’s exceptional sentence were neither admitted by petitioner nor found by a jury, the sentence violated his Sixth Amendment right to trial by jury.”).

289. Id.


present witnesses in her own defense,292 the accused does not have a right to compulsory process.293 Indeed, evidence submitted by the prosecutor may be submitted by proffer.294 Moreover, in the states, the amount of disclosure required varies widely; some districts comport with Brady requirements, and some do not. Moreover, the lack of the right to appointed counsel at the pretrial detention stage means that the vast majority of defendants do not have any legal representation.295

The failure to require full disclosure of evidence by the prosecutor at the pretrial detention stage means that indicted defendants often cannot challenge the evidence presented by the Government to establish the accused’s future dangerousness. Although the Supreme Court has held that the pretrial detention hearing does not generally require “the full panoply of adversary safeguards,”296 the establishment of potential dangerousness by the government is such a questionable area that perhaps this is one aspect of the hearing that the adversary process should apply. Especially because the standard of evidence required from the prosecutor at this stage is relatively low—only probable cause297—it seems only fair that the accused get a greater disclosure of evidence when she is subject to potential detention based on dangerousness.

More and better disclosure of prosecutorial evidence is important to help the offender challenge the request for pretrial detention.298 When the defense lacks knowledge of the evidence

---

294. Id.
297. See id. at 464 (stating that “the prosecutor is only required to show probable cause that the accused has committed the offense attributed to him in addition to proving the ground of detention: obstruction of justice or dangerousness”).
298. See id. at 465 (“Disclosure could be of paramount importance to the accused in contesting the evidence against him.”).
against him, the defendant cannot properly challenge the detention request, meaningfully participate in the hearing, or refute any secret evidence because the proceeding is one-sided.299

This lack of adversary context and one-sided nature of the detention hearing also implicates the spirit of the Sixth Amendment jury trial right. If, as the Apprendi-Blakely line of cases instructs us, the community is supposed to decide all potential punishments for an offender after observing the adjudicatory process, and punishment is being imposed on the offender during pretrial detention, then there should be some sort of adversarial process during the detention hearing. Although a full adversarial procedure would not be feasible, a detention hearing should provide a few more procedural safeguards. At the very least, this should include the requirement of a minimum proffer of prosecutorial evidence at the hearing, proved by a standard of probable cause, in any case in which the indicted defendant might potentially be detained due to community safety concerns. Of course, for this proffer to have meaning, defense counsel would have to be provided for those indigent offenders at the hearing.

2. Post-Blakely, Community as Only Arbiter of Punishment

Of even more concern to the application of the Sixth Amendment jury trial right is the belief that the community should be the primary arbiter of punishment for all offenders. The guarantee that “[i]n all criminal prosecutions, the accused shall enjoy the right to . . . trial, by an impartial jury” promises a criminal offender that “all the facts which must exist in order to subject the defendant to a legally prescribed punishment must be found by the jury.”300 This interpretation of the Sixth Amendment jury trial right ultimately became law in Blakely.301

This animating principle behind Blakely—that the community should determine all punishment to be meted out to

---

299. Id. at 466.
301. Blakely v. Washington, 542 U.S. 296, 301–02 (2004) (holding that to increase a sentence beyond the maximum suggested by the Guidelines, a jury must find the aggravating factors beyond a reasonable doubt).
the defendant—has import for not just the jury trial, but for other areas of criminal procedure too. *Blakely* applied first and foremost to sentencing, but there is no reason why it should not be applied to the pretrial detention hearing as well. This holds particularly true after the Supreme Court’s recent decision in *Southern Union*, holding that any punishment rendered by the courts, including criminal fines, should have its facts determined by a jury, beyond a reasonable doubt.302 *Southern Union* helps bolster the belief that *Blakely* is applicable to all types of criminal procedures, from the front-end to the back-end.

Recent conditions in pretrial detention centers have rendered any sort of time in them a form of punishment in addition to all the negative externalities that flow from a pretrial loss of liberty. Not only is this punishment imposed before a determination of guilt, but it is also imposed by a judge, not a jury—a violation of the Sixth Amendment jury trial right to determine facts relevant to punishment for a criminal offender. The question becomes, then, under what circumstances does the judicial determination of pretrial detention transform into punitive measures? When denial of bail results in an imposition of punishment, this might mean that the community should play a role, particularly in determinations of its own safety.

3. The Community Should Decide Danger to Itself

Considering that a very popular aspect of pretrial detention confines indicted offenders due to an alleged threat to community safety, it is ironic that this determination entirely lacks community imprimatur. In many cases, members of the community would be more familiar with who might be dangerous, particularly when it comes to nonviolent drug crimes. In contrast, a line prosecutor often has other concerns on his or her mind when determining whether an indicted offender may post bail before trial. Additionally, judges are not always part of the communities they govern, and may be more interested in uniformity (or in some cases, re-election) than individual justice. Thus there may be a true need for the community voice within

the bail hearing procedure. The prosecutor is both technically and legally the public's representative.\footnote{303. As the Supreme Court has noted, the prosecution's interest “is not that it shall win a case, but that justice shall be done.” Berger v. United States, 295 U.S. 78, 88 (1935).} The prosecutor's function as the people's representative, however, can be subsumed by a gradual inclination to align with the government and a desire to achieve a positive win-loss record. Too often, prosecutors have “their own agendas, both personal and administrative.”\footnote{304. Abraham S. Goldstein, *Converging Criminal Justice Systems: Guilty Pleas and the Public Interest*, 49 S.M.U. L. REV. 567, 569 (1996).} For prosecutors, the concepts of *public interest* or *justice* can be too diffuse and elastic to constrain them. A prosecutor's simultaneous representation of both the state and the people can get submerged in the everyday details of doing her job, particularly when the indicted offender is a high-profile defendant and there is considerable media scrutiny and pressure on the case.

Allowing some slice of the community to help determine whether the accused is truly a threat to community safety, then, makes both logical and ethical sense. First, the community often knows the offender far better than either the prosecutor or the judge, especially in a state, municipal or local forum. Despite the persistent fear of crimes perpetuated by strangers, people tend to commit most crimes within their communities.\footnote{305. See Thomas M. Mengler, *The Sad Refrain of Tough on Crime: Some Thoughts on Saving the Federal Judiciary from the Federalization of State Crime*, 43 U. KAN. L. REV. 503, 508, 516 (1995) (noting that “most crime is local in nature, and consequently, the local community feels the brunt of the offense”).} As such, some representation from the community might help make the best determination as to whether the accused might pose a danger if granted bail.

Additionally, granting the community some power to determine whether pretrial detention should be imposed would also foster a feeling of participation and investment in the criminal justice system, something that many members of the public lack. Being given the opportunity to make real decisions on community safety would hopefully make these citizens, and by association, their families and friends, feel far more connected to how the criminal justice system works. Put another way, instead of envisioning the criminal justice system as a faceless, remote
entity, the local public, through their participation in the pretrial detention hearing, would realize that they are part of the justice system as well. This would generally promote the legitimacy and public confidence in the justice system, something that has been lacking of late.

Finally, having the community become involved with pretrial detention determinations will result in the public’s increased understanding of the criminal justice system. Particularly with high-profile cases, the current nontransparent procedures determining pretrial detention can create both disappointment and a sense of helplessness in the local community. For example, when disgraced financier Bernie Madoff was granted bail after his confession to the police, the local and international public were angered and dismayed. The public’s integration into the pretrial detention hearing through participation in the procedure can eliminate some of the concerns inherent in imposing pretrial detention or granting bail by exposing the public to the actual discussion and debate over individual bail determinations.

4. Laymen and Members of Community As Predictors of Danger

Having an informed segment of the community take part in pretrial detention determinations will also help the accuracy of the “future dangerousness” predictions. Currently, judges tend to use their own intuition to determine how dangerous an indicted offender might be, or they rely on a prosecutor’s statement of disrupted community safety. This can result in an overprediction of future dangerousness. Naturally a prosecutor would have an interest in the defendant remaining in custody as long as

306. See Madoff Debate: Should He Be Free on Bail?, MSNBC.COM (Jan. 7, 2009), http://www.msnbc.msn.com/id/28540171/ns/business-us_business/t/madoff-debate-should-he-be-free-bail/ (last visited Sept. 24, 2012) (discussing how investors, editorial writers, and the general public expressed outrage regarding the granting of Madoff’s bail, while prosecutors argued Madoff should be thrown behind bars because he could flee or hide his assets) (on file with the Washington and Lee Law Review); see also Larry Neumeister, Bernie Madoff Escapes Jail, Judge Declines to Revoke Bail, HUFFINGTON POST (Jan. 12, 2009), http://www.huffingtonpost.com/2009/01/12/bernie-madoff-jail-hearin_n_157049.html (last visited Sept. 24, 2012) (discussing how “[t]here is a thirst for blood that transcends just those who have been victimized”) (on file with the Washington and Lee Law Review).
possible, both to ensure cooperation and to make sure no further crimes happen that might reflect negatively on the prosecutor's office itself. Although communities can also fall prey to overpredicting dangerousness, in many situations, particularly those not involving violent crime, the local public has a more nuanced view of the potential liabilities.

Judges can have a tendency to be biased in favor of predicting dangerousness, in part because they will be responsible if they erroneously release a violent individual. Additionally, judges, like all experts who routinely make these types of dangerousness determinations, may also have a tendency to generalize from experiences with past offenders on bases that have few, if any, relationships to future violence.

Members of the community, on the other hand, lack some of these pressures and biases. First, even if they participate in some sort of pretrial detention hearing, they would not bear any ultimate responsibility if the indicted offender were to commit another crime before trial. Equally important, the small cross-section of the public that would be involved in determining whether the accused obtained bail would not have the aforementioned tendency to generalize from prior experiences, allowing them to make the determination of pretrial detention from a fresh perspective.

5. Attacking the Problems of Race & Gender in Pretrial Detention

Like the rest of the criminal justice system, the pretrial detention procedure is rife with racial and gender-based disparities. First, people of color are disproportionately confined


308. See Saleem A. Shah, Dangerousness: A Paradigm for Exploring Some Issues in Law and Psychology, 33 Am. Psychologist 224, 229–30 (1978) (arguing that when prior probabilities of the outcome, or base rates, are very low, the predictions are of poor reliability). Of course, one way to combat this problem of the courts relying on intuition is to provide them with accurate empirical studies of which offenders are most likely to offend, such as that provided by Baradaran & McIntyre.
in jail, whether for pretrial detention or other reasons. Recent analysis has shown that both the race and gender of an offender can have a significant effect on the likelihood of pretrial detention.

For example, in federal cases, an offender’s sex affected pretrial custody for both black and white offenders; pretrial detention was less likely for both black and white females than for males of their respective races. Likewise, the likelihood of pretrial custody was substantially higher for black male offenders than for other offenders—twice those for white males, and over three times more likely than for black or white females. The author of the study has suggested that some of this may be due to judicial stereotyping of black or male defendants as more dangerous than their white or female counterparts, particularly when dealing with drug crimes. Finally, Latinos are most likely of all races to suffer the negative consequences of pretrial detention: they are most likely to have to pay bail (as opposed to being released on their own recognizance); courts tend to set them the highest bail amounts; they are least likely to be able to pay; and they are by far the least likely to be released prior to trial.

Although there is no guarantee that the local community or general public would hold less prejudicial attitudes toward

309. See Petteruti & Walsh, supra note 11, at 4 (“Latinos are more likely than are whites or African Americans to have to pay bail, and they have the highest bail amounts, are least likely to be able to pay, and are by far the least likely to be released prior to trial.”).

310. Casasia Spohn, Race, Sex and Pretrial Detention in Federal Court: Indirect Effects and Cumulative Disadvantage, 57 U. KAN. L. REV. 879, 893 (2009). The data analyzed in this article showed that race and gender affected pretrial detention determinations even after excluding the offender's dangerousness, community ties, financial resources, criminal history, and crime seriousness. Id.

311. Id. at 891.

312. Id. at 895–96.

313. Id. at 898–99. See generally Sara Steen, Rodney L. Engen & Randy R. Gainey, Images of Danger and Culpability: Racial Stereotyping, Case Processing, and Criminal Sentencing, 43 CRIMINOLOGY 435 (2005) (discussing their study of adult drug offenders in Washington and finding that those who most or least resemble a dangerous drug offender receive harsher or more lenient punishment, respectively).

314. Petteruti & Walsh, supra note 11, at 4.
defendants based on race or gender, in some communities, the presence of community members could make a difference in terms of expanding diversity and representation. Although strides certainly have been made to diversify the bar and the bench, the fact remains that magistrates and trial judges are primarily white\textsuperscript{315} and defendants are often minorities.\textsuperscript{316} In certain districts, such as the Bronx or East Los Angeles, the presence of a few community members would make some diversity that much more likely.

Additionally, the presence of representatives from historically minority communities would help make the ultimate decision to either grant bail or impose pretrial detention more understandable to the local public as well as explain how the pretrial detention decision might impact the community. Because many minority communities tend to feel alienated or distanced from the criminal justice system, the incorporation of some public representatives into the bail determination hearing would help reduce some of this distance.

\textbf{V. Proposal: Reform and Revision}

\textit{A. Reforming Bail Bondsmen and Pretrial Release}

One problem that plagues several states is the problem of commercial bondsmen. Some states have no commercial bondsmen at all, which results in serious problems for the poor, who often cannot afford even the small amount comprising their bail. Other states have bail bondsmen who do not provide loans for small bail amounts. There are a couple of ways to solve this problem.

First, states and counties could expand and better fund their pretrial release programs, changing their general policy to one that assumes the granting of bail unless there is a serious or


violent crime involved. Pretrial release programs often allow nonviolent or low-level offenders out on bail with supervision, using GPS, ankle bracelets, and monitoring.\footnote{Laura Sullivan, Bondsman Lobby Targets Pretrial Release Programs, NATIONAL PUBLIC RADIO (Jan. 22, 2010), http://www.npr.org/templates/story/story.php?storyid=122755849 (last visited Sept. 24, 2012) [hereinafter Sullivan, Bondsman] (on file with the Washington and Lee Law Review).} Allowing indicted offenders out on bail with the help of electronic monitoring permits them to save their jobs, pay their bills, keep their homes and see their families.\footnote{Id.} This would not only prevent nonviolent indicted offenders from suffering the dangers and indignities of pretrial detention, but also save counties and states thousands of dollars in incarceration costs.

However, pretrial release programs across the country are all too often fighting a futile battle with bail bond companies trying to either limit these types of programs or completely shut them down.\footnote{Id.} As one pretrial release program official notes, commercial bail bondsmen lobby to keep these programs as minuscule as possible, so that they do not siphon off any paying customers, even if that means thousands of inmates wait in jail at the taxpayers’ expense.\footnote{Id.}

Thus, one way to improve the current pretrial detention system is to increase local and county pretrial supervision programs in conjunction with much broader granting of bail, combined with providing appointed counsel for all indigent defendants in bail hearings.\footnote{As Yale Kamisar has long argued, the bail hearing is a critical phase, deserving of effective assistance of counsel. See, e.g., YALE KAMISAR ET AL., MODERN CRIMINAL PROCEDURE: CASES, COMMENTS AND QUESTIONS 872 n.8 (8th ed. 1994).} Although these services do cost money, in the long term they end up saving far more taxpayer dollars, as it is far more expensive to imprison those indicted offenders waiting for trial than to supervise them electronically at home.\footnote{Sullivan, Bail Burden, supra note 23.}
B. Revising the Bail Hearing Procedure

Some of the ills of our criminal justice systems can be traced to the dissociation between our local communities and their ability to effect change on the system. Insufficient local control is a serious issue, pervading all aspects of criminal justice. As Bill Stuntz persuasively argued,

To the suburban voters, state legislators, and state and federal appellate judges whose decisions shape policing and punishment on city streets, criminal justice policies are mostly political symbols or legal abstractions, not questions the answers to which define neighborhood life. Decisionmakers who neither reap the benefit of good decisions nor bear the cost of bad ones tend to make bad ones. Those sad propositions explain much of the inequality in American criminal justice.323

This is particularly true in pretrial detention hearings, at which the question of whether an indicted defendant is released or not before his trial often rests on decisions made by remote legislators or senior district attorneys. The local community has little or no say in the matter.

But this result is neither preordained nor necessary. One way to get the community more involved in the criminal justice process—thus making them feel more invested in the system—is to invite their participation in the pretrial detention hearing, allowing them to give their opinion on whether the suspect is a true threat to community safety.

Having the community actually involved in determining what would best serve community safety, and possibly being more lenient regarding the pretrial release of indicted, low-level offenders is not such a novel idea. Various scholars have noted in the past ten years the social consequences of mass incarceration: by incarcerating too many nonviolent criminals, either before or after conviction, not only are poor and minority communities increasingly harmed by the massive scale of incarceration, but many of these nonviolent offenders do become dangerous after being exposed to violent criminals in jail or prison.324 Thus, the

324. See, e.g., Jeffrey Fagan & Tracey L. Meares, Punishment, Deterrence and Social Control: The Paradox of Punishment in Minority Communities, 6
short-term decision to incarcerate pretrial offenders for purposes of “community safety” ends up backfiring on a number of levels in the long run.

However, incorporating the community into the bail hearing would have many positive results. These positive aspects include the public’s increased understanding of the criminal justice system, a restoration of criminal adjudication’s educative function, and a sense of investment and trust for the local community.

First, and most basically, the current bail hearing process—like so much else in the criminal justice system—functions out of sight from the average citizen. The local public has a meaningful interest in uncovering the procedures involved in denying or granting bail, especially because so many taxpayer dollars are being used to incarcerate those who have not yet been determined guilty.

Moreover, enhancing local, popular participation within an existing criminal justice institution,325 such as the bail hearing, combines the positives of community involvement without requiring new courts or immense change in the existing system. Additionally, through citizen involvement, the “cynicism and contempt” for the criminal justice system that is invariably created by more secret proceedings will be minimized.326 This is especially important for communities that have felt distanced and isolated by the criminal justice system; by allowing these communities to determine whether one of their own is “safe” enough to release pretrial, the local public may feel some investment or purchase into the workings of the system.

As Judith Resnik has powerfully argued, “[t]hird-party scrutiny illuminates the treatment of suspects, detaineest,327

---

325. See Adriaan Lanni, The Future of Community Justice, 40 HARV. C.R.-C.L. L. REV. 359, 363 (2005) (“The best way to introduce the community justice goal of greater citizen input into the administration of justice is not to scale up current community justice programs, but to provide for enhanced local, popular participation within existing criminal justice institutions.”).

prisoners, and immigrants, all reliant on government for their well-being."³²⁷ Allowing the community to observe and participate in the routine preventative detention hearing of a domestic defendant, then, would vindicate a number of rights, including, of course, the Sixth Amendment’s jury trial right. Resnik has made a similar suggestion concerning the imposition of detention on various detainees or prisoners, proposing a more audience role for the public.³²⁸ Her comment, however, applies equally to both her proposal and mine: “[L]aw ought to oblige open decision making when confinement is at stake.”³²⁹

Moreover, allowing a cross-section of the community to make decisions on questions of “dangerousness” along with the court would provide a less jaded sensibility to the pretrial detention determination. Although normally we defer to courts to make such decisions, having fresh eyes look at each individual situation, from a body that is not beholden to re-election or re-appointment processes, would inject both transparency and fairness into the proceedings. A bail jury, in other words, could make decisions informed by their own knowledge of the community and unburdened by judicial pressures and biases, providing a different view that could be added to the court’s determination to provide the fullest range of opinion before decision-making.

This kind of open decision making—and open participation by the local public—is most likely to take place in the state court system, which processes a vast percentage of pretrial detainees in the country. As Resnik notes, “[t]he last few decades have brought attention to state courts as a font of constitutional

³²⁷. Resnik, supra note 254, at 670. Although Resnik is primarily envisioning courts as this public oversight, she admits to a possible role for other third parties. As she notes, speaking of the general public’s right of audience during these impositions of detention, “[e]mpowered, participatory audiences can therefore see and then debate what legal parameters ought to govern.” Id.

³²⁸. Id. at 671. As Resnik argues, “law could require that some members of the public (subject to appropriate security screening) be permitted to observe decisions resulting in the long-term detention of persons, whether alleged to be terrorists, illicit migrants, or misbehaving prisoners.” Id.

³²⁹. Id. This is particularly true when, as now, the Court has “repeatedly insulated the federal judiciary from addressing the merits to decide when an individual is wrongfully convicted or detained in intolerable conditions.” Id. at 681.
jurisprudence that can be more rights-protective than federal precepts." Because of this, it would be easy enough to draw from the county jury rolls to create a body of local fact finders for bail hearings, who would be selected in a manner similar to those citizens selected for the grand jury.

These “bail juries” could sit for a week or two at a time, focusing their decisions on the community safety issue; in other words, whether the indicted defendants brought before the court would be eligible for bail. If the “bail jury” deemed the defendant safe for release out into the community, perhaps using a preponderance of the evidence standard, and the court had no substantive objection, then barring any true evidence of danger to the public, he or she would be released either on their own recognizance or electronically monitored until trial or guilty plea date. If the bail jury’s determination differed from the judge’s, then the court could call for more information from both sides to further illuminate the issue, until an agreement is reached.

C. Potential Problems

The major possible critiques of a bail jury are two-fold. First is one of history: the question of bail has always been left to the court, and not the community. Second is the cluster of concerns centering around the implementation of a bail jury, its potential costs, complexity, and delay. I address both these concerns below.

1. Lack of Historical Precedent

Since the practice of granting bail began, the traditional arbiter of detention or release has been the court. However, as discussed infra Part II, the decision of whether to detain a defendant pending trial was always left to the community—literally, as a friend or neighbor had to stand surety for the defendant and house him or her in their own residence. The various rules that developed regarding bail were, as noted above,

330. Id. at 682.
331. Thanks to Dan Markel for flagging this issue.
332. See infra Part II.A.
aimed more at limiting judicial discretion than anything else—always a concern in both seventeenth-century England and colonial America. Thus even in the earliest of bail systems, the community’s voice was heard.

Additionally, pretrial detention as originally configured was based only on risk of flight. There was no incarceration, at a neighbor’s house or in a local jail, ever predicated on future dangerousness, or community safety. Thus, there is no historical precedent limiting the bail decision solely to the judge. On the contrary, returning some power to the community would be a return to the original bail granting practice, upon which our Constitution is based.

Moreover, the use of a bail jury meshes neatly with the recent interest in local control over local environments. Various advocates of localism argue that “local governments are more responsive to the specific needs of unique communities and that local institutions can provide better and increased services.” These arguments parallel the one that can be made for the bail jury—that the local input is critical in implementing our criminal justice system, particularly for issues of community safety.

2. Impracticality/Cost/Delays

A second, potentially more serious, critique of the bail jury is its impracticality, both in terms of cost and delay. In these fiscally stringent times, any procedure that would add to the cost structure of the criminal justice process is viewed dubiously at best. And incorporating a bail jury into the pretrial detention hearing would potentially increase costs; although the bail jury could be drawn from the same rolls as the grand jury, the more citizens drawn, the more money needed to fund per diems, reimburse transportation costs, and pay court staff to organize such juries.

333. See Richard C. Schragger, The Limits of Localism, 100 Mich. L. Rev. 371, 380 (2001) (finding an “increasing insistence on and institutionalization of local control over local environment”). Schragger points out that this new localism has arisen as a response to urban disorder and the problems of urban governance. Id.

334. See id. at 381.
As noted above, however, the potential amount of savings that a properly working bail jury would provide could be vast. Even a thirty percent reduction in those incarcerated before trial could save the state thousands of dollars. And while there are no guarantees that a bail jury would be more lenient on matters of pretrial release than the traditional judicial arbiter, it is likely that a more informed body of decision makers—i.e., a cross-section of the community—might not believe that every indicted offender is a threat to public safety.

As for delays, it is possible that the incorporation of the bail jury would slow down the pretrial detention somewhat. But that is not necessarily a bad thing. Part of the problem with our current bail hearings is the speed at which these critical determinations happen. Since there is no right to appointed counsel at a bail hearing, often they are very fast, as the prosecutor presents evidence regarding defendant’s incarceration and the court makes a decision. Reducing the haste of the process might also ensure that both the court and the prosecutor take the process more seriously; since the bail jury will not be criminal justice insiders, they will likely focus on each individual case more intensely, and require the prosecutor to more fully explain her reasoning for denying bail. The court, too, may spend more time on the decision, as it will need to incorporate the bail jury’s decision into its own. Thus, the delay in imposing pretrial detention or granting bail might be a positive one, opening up the procedure to some much needed sunshine and scrutiny.

Although the bail jury does not tidily solve all of the problems of our framework of bail and jail, it provides a partial solution to some of the glaring inequities of pretrial detention.

VI. Conclusion

Reform is desperately needed in the realm of pretrial detention to remove it from the Shadowlands of justice. As Resnik has persuasively argued, in protecting and preserving rights, “Article III judges—the exemplars of independent jurists—can never be enough.”335 This is equally true for state court judges

---

335. Resnik, supra note 254, at 685. As Resnik argues, “[a]s the constitutional law of detention makes painfully clear, if American law is to
and magistrates, many of whom are at the mercy of re-election cycles, overwhelming caseloads, and restrictive statutes.

Although there is no one perfect solution, the use of a bail jury, combined with increasing monitored pretrial release, would begin to solve both constitutional and procedural problems. First, the bail jury would help ameliorate the Sixth Amendment issue, allowing the community a chance to help decide any pretrial punishment imposed on indicted offenders. Additionally, giving local citizens a say in who is released back into their community has a practical aspect to it; instead of having the prosecutor and the judge, both representatives of the government, be the only ones to determine community safety, it makes sense to have some input from the very community that the government is trying to protect. Second, the addition of increased electronic monitoring of those offenders released before trial would be a relatively easy, cost-effective way to permit those accused who either cannot afford to make bail or about whom the community or judge still have some reservations to escape remaining in jail until their trial.

As both a practical measure and a fundamental matter of constitutional fidelity, the people should be involved in the machinations of criminal punishment. This includes the procedures that happen before trial. Our current system of pretrial detention lies in shambles, incarcerating those not yet convicted in punitive conditions often far worse than those existing in prisons. Allowing the community a say in the matter and broadening the ambit of those who can be released pretrial is one way to shine light into the darkness.

cherish human dignity, it will be because more than life-tenured judges make it do so.” *Id.*
Beyond War: Bin Laden, Escobar, and the Justification of Targeted Killing

Luis E. Chiesa*
Alexander K.A. Greenawalt**

Abstract

Using the May 2011 killing of Osama bin Laden as a case study, this Article contributes to the debate on targeted killing in two distinct ways, each of which has the result of downplaying the centrality of international humanitarian law (IHL) as the decisive source of justification for targeted killings.

First, we argue that the IHL rules governing the killing of combatants in wartime should be understood to apply more strictly in cases involving the targeting of single individuals, particularly when the targeting occurs against nonparadigmatic combatants outside the traditional battlefield. As applied to the bin Laden killing, we argue that the best interpretation of IHL would have required the SEALs to capture bin Laden in conditions short of surrender, if he was in fact manifestly defenseless or otherwise could have been readily captured with little risk.

Second, we take seriously the possibility that the law should tolerate some targeted killings under conditions that are justified neither by reference to IHL nor by reference to the traditional...
justifications available in peacetime. Drawing upon the example of Colombian crime family leader Pablo Escobar, who died in a police raid in 1993 in circumstances suggesting the authorities were not interested in capture, we suggest that targeted killings in these circumstances may be morally—if not legally—justified when (1) killing the targeted individual will protect society from a serious threat, (2) the individual is undeniably culpable for past atrocities, and (3) trying the individual is eitherlogistically impossible or extraordinarily dangerous.

Although we conclude that the bin Laden killing does not clearly satisfy the third criterion, this model nevertheless provides—in important ways—a superior framework for understanding public responses to bin Laden’s death than does the war paradigm.

Table of Contents

I. Introduction ................................................................. 1373

II. Justifying Targeted Killing in War .............................. 1383
   A. The Morality of War and the Limits of
      International Law .................................................... 1389
   B. IHL and Targeted Killing ........................................ 1396
      1. The Permission to Kill by Name ............................ 1396
      2. Targeted Killing in a Global Battlefield ............ 1400
         a. The Feasibility of Capture .............................. 1400
         b. The Harm of a Diffuse Battlefield ................. 1406
      3. The Problem of Guilt ....................................... 1413
   C. The War Paradigm and the bin Laden Killing .......... 1415

III. The Lawfulness of Targeted Killings Beyond War ...... 1420
   A. A Typology of Cases ............................................. 1420
   B. Targeted Killing as Traditional Defense of
      Self or Others ..................................................... 1422
   C. Targeted Killing as Expanded Defense of
      Self or Others ..................................................... 1425
   D. Targeted Killing as Law-Enforcement Action ........ 1428
   E. Targeted Killings as Necessity or Lesser Evils ....... 1431

IV. Another Paradigm for Targeted Killings? The Killing
    of Pablo Escobar ................................................... 1434
I. Introduction

On the evening of Sunday, May 1, 2011, President Barack Obama appeared on television to announce that a United States operation had killed Osama bin Laden following a raid upon the al Qaeda leader’s secret home in Abbottabad, Pakistan. In the United States, the news that the mastermind of the 9/11 attacks had finally been taken down after a decade of false leads and
intelligence failures was met, understandably, with widespread jubilation, and crowds gathered in New York City and Washington, D.C., to celebrate the event. As the New York Times reported the following morning, “Bin Laden’s demise is a defining moment in the American-led fight against terrorism, a symbolic stroke affirming the relentlessness of the pursuit of those who attacked New York and Washington on Sept. 11, 2001.”

News reports in subsequent days and weeks revealed additional details of the suspenseful operation against bin Laden: A Navy SEAL team flew clandestinely by helicopter deep into Pakistan and undertook its mission by cover of night in a sprawling compound. Analysts debated the significance of the killing for future anti-terrorism efforts and the implications for the United States’ relationship with Pakistan, given the startling revelation that bin Laden was not hiding in the remote tribal areas of the Northwest, where Pakistan has struggled to establish control, but in an affluent suburb only thirty miles outside Islamabad.

Comparatively little public attention, at least inside the United States, has focused on the killing’s legality. Was the

---


3. Id.


5. See Schmidle, supra note 4, at 36 (describing American efforts to locate bin Laden in Pakistan since 2001 and stating “it remains unclear how [bin Laden] ended up living in Abbottabad”).

BEYOND WAR

killing an exercise of individual self-defense on the part of Navy SEALs? Or were U.S. forces operating under more permissive rules of engagement that did not require bin Laden to pose an immediate threat? In his speech to the nation, the President did not address these questions explicitly. To the extent his remarks provided an answer, they gave mixed signals. On the one hand, President Obama emphasized the gravity of the al Qaeda leader’s criminality and appeared to portray the mission as an exercise in criminal justice. Bin Laden was “the leader of al Qaeda, and a terrorist who’s responsible for the murder of thousands of innocent men, women, and children.” The operation the President authorized was one “to get bin Laden and bring him to justice,” and by his killing, the President could “say to those families who have lost loved ones to al Qaeda’s terror: Justice has been done.” The President struck a similar chord in a television interview the following week, in which he explained that “the one thing I didn’t lose sleep over was the possibility of taking bin Laden out. Justice was done. And I think that anyone who would question that the perpetrator of mass murder on American soil didn’t deserve what he got needs to have their head examined.”

Few could doubt the President’s assessment of bin Laden’s individual desert, but desert alone does not supply a justification for killing. The law, after all, generally does not authorize law-enforcement officials to kill dangerous criminals in lieu of a criminal trial conducted in accordance with due process of law.

---

7. See Obama, supra note 1 (recounting bin Laden’s role in the 9/11 attacks, efforts to track bin Laden, and Obama’s order “to get Osama bin Laden and bring him to justice” but not addressing any legal grounds supporting the killing).
8. Id.
9. Id.
11. In U.S. law, this prohibition finds expression in the Fifth and Fourteenth Amendments to the Constitution. See U.S. CONST. amends. V, XIV (codifying due process rights). At the international level, equivalent protections are found in major human rights treaties among other sources. See Nils Melzer, Targeted Killing in International Law 91–139 (2008) (analyzing protections found in the International Covenant on Civil and Political Rights (ICCPR), the American Convention on Human Rights (ACHR), the African Charter on Human and Peoples’ Rights (ACHPR), and the European Convention
Accordingly, killing by law-enforcement officials is generally only permitted on urgent preventive grounds, to deflect an imminent lethal threat to self or others, or to prevent a dangerous suspect’s escape.12

In his speech, the President also indicated a separate consideration that has significant legal implications: the United States was at “war against al Qaeda to protect our citizens, our friends, and our allies.”13 Although the discourse of war received only passing mention in this context, it assumed prominence in later days as administration officials more explicitly addressed the legality of the bin Laden killing. In a May 19, 2011 blog post on the international law blog Opinio Juris, State Department Legal Advisor Harold Hongju Koh relied exclusively on the law regulating the conduct of war—the “law of war” in U.S. legal terminology, or international humanitarian law (IHL) in the terminology of international law—to justify the operation against bin Laden.14 He noted that, following 9/11, Congress authorized the use of force against al Qaeda, and he referred to the current existence of an “armed conflict with al-Qaeda, the Taliban and associated forces.”15 Pursuant to the applicable rules of war, argued Koh, U.S. forces were permitted to kill bin Laden as an enemy combatant without prior due process because he had not affirmatively provided a “genuine offer of surrender that [was] clearly communicated . . . and received by the opposing force,

12. See infra Part III (discussing permissive killings); MELZER, supra note 11, at 232 (providing the circumstances, based on a review of international legal sources, in which lethal force may be used). Melzer concludes that “[a]s a general rule, potentially lethal force should not be used except to: (1) defend any person against an imminent threat of death or serious injury, (2) prevent the perpetration of particularly serious crime involving grave threat to life, or (3) arrest a person presenting such a danger and resisting arrest, or to prevent his or her escape.” Id.


15. Id.
under circumstances where it [was] feasible for the opposing force to accept that offer of surrender."16

The invocation of the more permissive approach to killing applicable in wartime raises its own set of questions, concerning whether the war paradigm applies and which rules of engagement this body of law permits. The tension revealed in Obama’s speech—between bin Laden as a criminal suspect and bin Laden as an enemy combatant—is emblematic of the broader difficulties raised in recent years by the practice of targeted killing as a component of the United States’ efforts to combat al Qaeda in Afghanistan, Pakistan, Yemen, and elsewhere, and by Israel’s counterterrorism efforts in the West Bank and Gaza.17


17. Recent academic literature addressing these issues is substantial. For selected works of particular relevance to this Article, see generally TARGETED KILLINGS: LAW AND MORALITY IN AN ASYMMETRICAL WORLD (Claire Finkelstein, Jens David Ohlin & Andrew Altman eds., 2012); MELZER, supra note 11; Gabriella Blum & Philip Heymann, Law and Policy of Targeted Killing, 1 HARV. NAT’L SEC. J. 145 (2010); Michael L. Gross, Assassination and Targeted Killing: Law Enforcement, Execution or Self-Defence?, 23 J. APPLIED PHIL. 323 (2006); Mary Ellen O’Connell, Unlawful Killing with Combat Drones: A Case Study of Pakistan, 2004–2009, in SHOOTING TO KILL: THE LAW GOVERNING LETHAL FORCE
Gabriella Blum and Philip Heymann have aptly summarized how this emerging practice has tested the intuitions underlying the traditional legal paradigms that regulate the conditions under which governments may kill individuals as exercises of either law enforcement or war powers:

A targeted killing entails an entire military operation that is planned and executed against a particular, known person. In war, there is no prohibition on the killing of a known enemy combatant; for the most part, wars are fought between anonymous soldiers, and bullets have no designated names on them. The image of a powerful army launching a highly sophisticated guided missile from a distance, often from a Predator drone, against a specific individual driving an unarmored vehicle or walking down the street starkly illustrates the difference between counterinsurgency operations and the traditional war paradigm. Moreover, the fact that all targeted killing operations in combating terrorism are directed against particular individuals makes the tactic more reminiscent of a law enforcement paradigm, where power is employed on the basis of individual guilt rather than status (civilian/combatant). Unlike a law enforcement operation, however, there are no due process guarantees: the individual is not forewarned about the operation, is not given a chance to defend his innocence, and there is no assessment of his guilt by any impartial body.  

We do not attempt in this Article to resolve every problem associated with targeted killing. Nor do we attempt a definitive answer regarding the legality of bin Laden’s killing itself. However, the circumstances surrounding bin Laden’s demise bring into sharp focus the war paradigm’s limits as a justifying framework for the evolving practice of targeted killing. In this respect, this Article contributes to the debate on targeted killing in two distinct ways, both of which have the result of downplaying the war paradigm as the decisive justification for targeted killing.


In Part II, we argue that IHL, assuming it applies, does not play the decisive role in justifying targeted killing that is often assumed.19 The problem here is twofold. First, there is indeterminacy in IHL itself. It remains unclear how the requirements of IHL—in particular, the overarching mandate that killing pursue military necessity—apply to operations, such as the bin Laden killing, whose functional requirements do not resemble those traditionally associated with the battlefield, but rather mirror those of traditional law-enforcement operations.20 Second, the problem of targeted killing exposes a deeper uncertainty about the moral status of IHL. At its root, IHL is a body of law that relaxes the deep-seated rule against murder.21 Its expanded permission to kill most plausibly reflects realism in the face of the historical inevitability of war rather than a fully developed moral justification for killing.22 Because of this feature of IHL, the possibility that other law might similarly evolve to accommodate expanded forms of killing outside the war paradigm cannot be ruled out. However, it also urges deep caution with respect to the interpretation and the application of IHL in nonparadigmatic cases.

These considerations justify the intuition—reflected already, to a degree, in both U.S. policy23 and Israeli judicial doctrine24—that the rules governing killing should be stricter in cases involving the targeted killing of single individuals, particularly when the targeting occurs against nonparadigmatic combatants outside the traditional battlefield. They also indicate that IHL’s

19. See infra Part II (discussing moral and legal justifications for targeted killing in war).

20. See infra Part II.C (discussing the difficulties of applying traditional criteria for targeting killing to the circumstances involving bin Laden).

21. See MELZER, supra note 11, at 74–81 (defining basic principles of IHL and the legal framework it provides for justifying targeted killings).

22. See infra Part II.A (discussing how IHL’s justification for killing illustrates the law’s adapting to real needs).


generally expansive approach to killing is inappropriate in operations that are not supported by functional requirements or reasons that plausibly sound in military advantage.\textsuperscript{25}

Although these considerations alone cannot tell us whether it was permissible to kill bin Laden under the circumstances—much depends on facts that may never be known—they do call into question the breadth of the standard that the State Department applied in justifying the killing. It is not enough, we argue, to establish that bin Laden was a combatant engaged in armed conflict against the United States who had not affirmatively and clearly surrendered under conditions that would have facilitated a safe capture. We argue that the best interpretation of IHL required an attempt to capture bin Laden absent surrender unless the decision to kill in lieu of capture was justified by recourse to conventional understandings of military advantage, was in fact motivated by the pursuit of military advantage, and would have spared U.S. military personnel from risks greater than those generally expected of law-enforcement officials acting under like circumstances.\textsuperscript{26} Although it is correct, as commentators have noted, that IHL does not impose any general duty to capture non-surrendering combatants, the requirements we identify are implied in other rules of IHL and, more deeply, by the structure of IHL itself.\textsuperscript{27}

The remainder of the Article takes seriously the possibility that the law should tolerate some targeted killings under conditions that are justified neither by reference to an armed conflict nor by reference to the traditional parameters of the law-enforcement paradigm. Part III explores the limits of existing legal categories rooted in self-defense, lesser-evils, and law-enforcement authority as frameworks for justifying targeted killings.\textsuperscript{28} In Part IV, we identify what we believe to be one of the

\textsuperscript{25} See infra Part II.B (discussing the importance of military necessity to justify targeted killing).

\textsuperscript{26} See infra Part V.D (laying out criteria to justify killing rather than capture).

\textsuperscript{27} See infra Part II.B (arguing that IHL implicitly requires a duty of capture in some cases).

\textsuperscript{28} See infra Part III (discussing legal grounds for targeted killing outside of war, including defense of self or others, law enforcement, and a necessary or lesser evil).
most persuasive cases for justifying targeted killing beyond the currently accepted paradigms: that of late Colombian crime family leader Pablo Escobar, who died in a police raid in 1993 in circumstances suggesting the authorities were not interested in capture.29 As we explore in Part V, we are uncertain whether the lessons of the Escobar killing can or should be reduced to a judicially enforceable legal standard, but the intuition that Colombian authorities would have acted justifiably in ordering the killing is supported by three features that distinguish Escobar from other criminal suspects.30 First, at the time of his death, he remained a dangerous individual who posed demonstrable dangers to Colombian society and was almost certain to continue his involvement in serious crimes.31 Second, he was unquestionably guilty of horrendous crimes that rank among the worst of the late twentieth century.32 Third, his track record—including the murder of incorruptible judges and his prior escape from confinement—indicated that it was infeasible and unacceptably dangerous to try him.33

Might these criteria that support the killing of Escobar likewise justify the killing of Osama bin Laden? We think not. As we explain in Part VI, although the bin Laden case mirrors Escobar’s in the combination of grave, unquestionable guilt and continuing danger, the obstacles to trial remain speculative and lack the demonstrable urgency witnessed by the case of Escobar, who, by the time of his death, had already established a track record of corrupting and murdering the officials whom the legal system relied on to bring him to justice.34 There is no parallel in the case of bin Laden, and to justify his killing based on the more generalized fear that a trial would bring security risks would create a dangerous slippery slope.

29. See infra Part IV (discussing the killing of Pablo Escobar as well as reactions to and justifications for the killing).
30. See infra Part V (analyzing what Escobar’s killing teaches about the morality of targeted killing and its justifications).
31. See infra Part IV.A (describing Escobar’s threat to the justice system).
32. See infra Part IV.A (describing Escobar’s crimes).
33. Id.
34. See infra Part VI (comparing Escobar and bin Laden and considering whether justifications for killing Escobar extend to bin Laden’s killing).
Nevertheless, the Escobar precedent remains relevant to the bin Laden killing. As a descriptive matter, we believe that the Escobar precedent better explains the public reception of bin Laden’s killing—if not the government’s actual motives—than does a focus on wartime hostilities. In other words, the reason why bin Laden’s killing occasioned so little debate in the United States is that the public saw bin Laden much as the President presented him in his May 1 speech: not simply or even primarily as a military opponent, but as a dangerous and deeply culpable criminal.\textsuperscript{35} As a normative matter, moreover, the Escobar precedent arguably provides a superior—if ultimately unsuccessful—argument for why, assuming bin Laden could have been readily captured at minimal risk, it was nevertheless appropriate to kill him. Finally, Escobar’s case is informative for understanding the bin Laden killing as a precedent, one that may point toward a developing law in action that justifies a limited set of targeted killings based on considerations of culpability and danger, rather than a connection to armed conflict. 

Although we are concerned about where this precedent could lead and would sharply limit the evolving law to privilege criminal trials whenever feasible, there is also value to speaking honestly about the possible bases of government action. The temptation to rely on IHL to justify targeted killings understandably derives from the fact that this is the body of law that makes it easiest to justify killing. But relying on IHL comes with the risk that states will expansively interpret that body of law so as to justify policies that have little basis in the values or functional requirements that initially gave rise to IHL. If governments are in fact looking to culpability considerations in developing policies of targeted killing, then there is value to considering the circumstances under which such policies might be justified and to debating concrete cases pursuant to the criteria that are likely of most relevance to decision makers.

\textsuperscript{35} See Obama, supra note 1 ("[B]in Laden has been al Qaeda’s leader and symbol, and has continued to plot attacks against our country and our friends and allies.").
II. Justifying Targeted Killing in War

When people debate the legality of targeted killings, they usually have in mind one or more cases that they believe represent core examples of the practice. A common case—call it TK 1—takes place in the theater of war and invokes IHL\(^36\) to justify the killing.\(^37\) As a general matter, the world’s legal systems forbid the intentional killing of individuals except in narrowly defined circumstances, such as when necessary to prevent the imminent use of lethal force by a wrongdoer.\(^38\) But if TK 1 cases are properly subject to IHL, the legal landscape changes considerably. Pursuant to these war rules, the norm against killing is remarkably relaxed. Provided that a number of status-based requirements are met, combatants in armed conflicts are permitted to use lethal force against their opponents in ways that would otherwise violate the peacetime prohibition against murder. When combatants aim to kill their opponents, there is no requirement that those targeted pose an imminent threat, or indeed pose any sort of direct threat beyond their participation in the broader war effort.\(^39\) As Michael Walzer has

---

36. See Melzer, supra note 11, at 74–81 (defining basic principles of IHL and the legal framework it provides for analyzing targeted killings).
37. See id. at 55–58 (discussing the application of IHL to justify targeted killing in conflicts).
38. See infra Part III (addressing the scope of permissible killing in peacetime).
39. See, e.g., Blum & Heymann, supra note 17, at 146 (“[T]he enemy combatants belong to another identifiable party and are killed not because they are guilty, but because they are potentially lethal agents of that hostile party.”).
put it, the law of war permits the killing of “naked soldiers,” those who, for example, are eating their breakfast or are asleep in their beds. The potential for combatants to be attacked derives from the general danger posed by their status. So long as hostilities persist, the potential to be killed terminates only upon a combatant becoming incapacitated—or “hors de combat”—on account of his falling into the power of an adverse party; his clearly expressed surrender; or his being rendered unconscious or “otherwise incapacitated by wounds or sickness” such that he is “incapable of defending himself.”

Civilians, by contrast, may not be targeted “unless and for such time as they take a direct part in hostilities,” and the intentional targeting of civilians is a war crime under international law. Nor may belligerents employ weapons or

v. Gov't of Israel ¶ 4 [2005] (Isr.), available at http://elyon1.court.gov.il/files_eng/02/690/007/a34/02007690.a34.pdf (stating that Additional Protocol I “reflects the norms of customary international law, which obligate Israel”); see, e.g., JEAN-MARIE HENCKAERTS & LOUISE DOSWALD-BECK, 1 CUSTOMARY INTERNATIONAL HUMANITARIAN LAW: RULES 72 (2005) (mentioning an Israel–Lebanon ceasefire understanding relying on customary international law principles and outlining rules of customary international law governing the conduct of hostilities).

40. See MICHAEL WALZER, JUST AND UNJUST WARS: A MORAL ARGUMENT WITH HISTORICAL ILLUSTRATIONS 138 (1977) (“The first principle of the war convention is that, once war has begun, soldiers are subject to attack at any time (unless they are wounded or captured).”); Gabriella Blum, The Dispensable Lives of Soldiers, 2 J. LEGAL ANALYSIS 69, 71 (2010) (noting that “[t]he existing interpretation of the laws of war supports Walzer’s conclusion” that a soldier stripped naked and swimming in the lake is a legitimate target during an armed conflict); Yoram Dinstein, The System of Status Groups in International Humanitarian Law, in INTERNATIONAL HUMANITARIAN LAW FACING NEW CHALLENGES: SYMPOSIUM IN THE HONOUR OF KNUT IPSEN 144, 148 (Wolff Heintschel von Heinegg & Volker Epping eds., 2007) (“[O]rdinary combatants . . . can be attacked (and killed) wherever they are, in and out of uniform: even when they are not on active duty. There is no prohibition either of opening fire on retreating troops (who have not surrendered) or of targeting individual combatants.”).

41. See Blum, supra note 40, at 71 (stating that one’s liability to attack is based on his or her status as a combatant).

42. Additional Protocol I, supra note 39, art. 41. Thus, IHL prohibits the denial of quarter, including orders that there shall be no survivors. Id. art. 40; see also MELZER, supra note 11, at 367–71 (discussing the IHL rule against denial of quarter).

43. Additional Protocol I, supra note 39, art. 51(3).

means of warfare that are indiscriminate or that cause unnecessary suffering.45 Even so, civilians are also subject to reduced protection under the war rules. When combatants attack military targets, they may do so knowing that their actions will kill civilians, so long as the anticipated loss of civilian life is not disproportionate to the anticipated military advantage.46

These rules apply equally to all sides in conflict, without distinction based on whether the party invoking IHL is supported by a just cause.47 Thus, the rules are the same even if the killers are invaders engaged in an illegal aggressive conquest, and their opponents have taken up arms justly because self-defense requires it. In the case of international armed conflicts covered by the Geneva Conventions of 1949, states are affirmatively prohibited from criminally prosecuting as murderers those who have complied with IHL’s requirements.48 No corresponding

directing attacks against the civilian population as such or against individual civilians not taking direct part in hostilities” is a war crime).

45. Additional Protocol I, supra note 39, arts. 35(2), 51(4).

46. See id. art. 51(5)(b) (stating that attacks should be considered indiscriminate if it “may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated”).

47. See id. pmbl. (applying this protocol “without any adverse distinction based on the nature or origin of the armed conflict or on the causes espoused by or attributed to the Parties to the conflict”). Here as well, these rules are more permissive than those generally enforced by the criminal law. A bank robber who kills a police officer in self-defense may not justify the killing if the police officer herself was justified in using lethal force against the robber under the circumstances. See WALZER, supra note 40 at 127–28 (offering this hypothetical). Yet precisely such a privilege attaches to those who, fighting an illegal war of aggression, kill enemy combatants who fight a just war of self-defense. See id. at 128 (“The moral equality of the battlefield distinguishes combat from domestic crime.”).

48. This immunity appears, among other places, in the Geneva Convention Relative to the Treatment of Prisoners of War. Pursuant to Article 87, “[p]risoners of war may not be sentenced by the military authorities and courts of the Detaining Power to any penalties except those provided for in respect of members of the armed forces of the said Power who have committed the same acts.” Geneva Convention Relative to the Treatment of Prisoners of War, art. 87, Aug. 12, 1949, 6 U.S.T. 3316, 75 U.N.T.S. 135 [hereinafter POW Convention]. Article 99 in turn provides that “[n]o prisoner of war may be tried or sentenced for an act which is not forbidden by the law of the Detaining Power or by international law, in force at the time the said act was committed.” Id. art. 99.
combatant privilege applies in the case of internal armed conflicts (thus, governments may invoke their domestic laws to prosecute rebels for their rebellion), but even here the Second Additional Protocol to the Geneva Conventions endorses “the broadest possible amnesty to persons who have participated in the armed conflict.”

In light of this extraordinary permission that IHL affords, it is unsurprising that the legal debate over the United States’ targeted killing policy has focused predominantly on whether and how this body of law applies. These are the rules that the United States has invoked to justify its targeted killing policies in Afghanistan, Pakistan, and Yemen, and as already noted, to justify the specific targeting of bin Laden. Similarly, Israel’s Supreme Court invoked IHL when ruling on the permissibility of targeted killings in the West Bank and Gaza.

The invocation of IHL in these circumstances has triggered various points of controversy. One set of questions focuses on targeted killing’s *jus ad bellum* dimension. When the United

As one U.S. court has noted, “[t]hese Articles, when read together, make clear that a belligerent in a war cannot prosecute the soldiers of its foes for the soldiers’ lawful acts of war.” United States v. Lindh, 212 F. Supp. 2d 541, 553 (E.D. Va. 2002).


50. See Melzer, supra note 11, at 262–69 (examining the U.S. government’s arguments that the war on terrorism falls outside of traditional categories of “conflict” and therefore is not governed by IHL).

51. See Koh, supra note 14 (using IHL’s framework, including its principles of distinction and proportionality, in arguing that the targeted killing of bin Laden was legal); Johnson, supra note 16 (discussing when targeted killing is legal); Holder, supra note 16 (stipulating conditions under which targeted killing is legal); Brennan, supra note 16 (discussing legal principles behind targeted killing).

52. See HCJ 769/02 Pub. Comm. Against Torture in Israel v. Gov’t of Israel para. 18 [2005] (Isr.), available at http://elyon1.court.gov.il/Files_ENG/02/690/007/a34/02007690.a34.pdf (applying IHL to the armed conflicts in the West Bank and Gaza).

States launches an attack against an individual residing in Pakistan or Yemen, does the attack represent a legitimate use of force, or is the United States committing an act of aggression against the territorial sovereignty of the state in which the attack has taken place?54 Another set of questions focuses on whether IHL is the operative body of law governing a particular killing, the central issue being whether the use of lethal force reflects a requisite nexus to an armed conflict (either international or non-international in character).55 A third set of questions focuses on

54. See id. paras. 37–45 (describing the necessary conditions to allow a targeted attack on an individual in another state’s territory); Van Schaack, supra note 2, at 266–81 (surveying the *jus ad bellum* issues present in the killing of Osama bin Laden and Anwar al-Aulaqi); O’Connell, supra, at 13 (“The drones used in Pakistan are lawful for use only on the battlefield. The right to resort to them must be found in the *jus ad bellum*; the way they are used must be based on the *jus in bello* and human rights.”); Paust, supra note 17, at 279 (concluding that, when acting in self-defense, a nation may attack non-state actors in another state without that state’s permission); Kenneth Anderson, *Targeted Killing in U.S. Counterterrorism Strategy and Law,* in *LEGISLATING THE WAR ON TERROR: AN AGENDA FOR REFORM* 346, 369–70 (Benjamin Wittes ed., 2009) (discussing the targeting of terrorist suspects without the territorial state’s consent).

On this point, special attention has focused on the scope of the right to self-defense under international law, considering that non-state actors are often killed across international borders. The ICJ has denied that such a right exists with respect to armed attacks by non-state actors that are not imputable to a state. *See Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory, Advisory Opinion,* 2004 I.C.J. 136, paras. 138–39 (July 9) (refusing to apply the inherent self-defense right to permit armed attacks when the “attacks against [the U.N.-member state] are [not] imputable to a foreign State”); *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v. Uganda),* 2005 I.C.J. 168, paras. 235–36 (Dec. 19) (stating that international responsibility only arises if injurious acts are imputed to a state). Current U.S. policy dictates that the use of force can be permissible in those circumstances, absent consent of the third state, provided the state is unable or unwilling to deal effectively with the threat to the United States posed by the non-state group. *See Holder,* supra note 16 (“[T]he use of force in foreign territory [without consent] would be consistent with these international legal principles . . . after a determination that the nation is unable or unwilling to deal effectively with a threat to the United States.”); Brennan, supra note 16 (“There is nothing in international law that bans the use of remotely piloted aircraft for this purpose or that prohibits us from using lethal force against our enemies outside of an active battlefield, at least when the country involved consents or is unable or unwilling to take action . . . .”).

55. See Alston Report, supra note 53, paras. 46–56 (describing the existence and scope of an armed conflict); *Melzer,* supra note 11, at 76–81 (stating that “the applicability of IHL presupposes the existence of an international or non-
whether the targets of attack—often un-uniformed individuals suspected of membership in terrorist groups—are in fact subject to attack under IHL, either because they have the status of combatants who may be targeted at any time or because they are civilians directly participating in hostilities when they are targeted.56

These are vital questions whose resolution is critical to establishing both the permissibility of many targeted killings and the broader scope of anti-terrorism efforts. They are not, however, the primary focus of this Article. We will instead largely assume in this Part that IHL does in fact apply to bin Laden’s killing and to other instances of targeted killing. We nevertheless argue that IHL does not play as determinative a role in justifying targeted killing as the current debate would suggest.

Subpart A strikes a cautionary note about the moral status of the IHL rules.57 Understood in its best light, the scope of the permission to kill in wartime does not reflect a fully developed moral framework for killing, but instead reflects a body of law that is largely reactive to the historical experience of warfare as an inevitable evil that can at best be regulated on the margins so international ‘armed conflict’). The United States maintains that it is engaged in armed conflict with al Qaeda and that the conflict extends beyond the borders of Afghanistan. See Holder, supra note 16 (“Our legal authority is not limited to the battlefields in Afghanistan. . . . We are at war with a stateless enemy, prone to shifting operations from country to country.”); Johnson, supra note 16 (stating that the conflict against al Qaeda requires military authority for necessary and appropriate force to extend beyond the “hot” battlefields of Afghanistan”); Koh, supra note 14 (acknowledging “our armed conflict with al Qaeda”). Thus far, the Obama Administration has not specified whether it considers this to be an international or non-international armed conflict. For a discussion about the complexities associated with both views, see Craig Martin, Going Medieval: Targeted Killing, Self-Defense and the Jus ad Bellum Regime, in TARGETED KILLINGS: LAW AND MORALITY IN AN ASYMMETRICAL WORLD 223, 231 (Claire Finkelstein, Jens David Ohlin & Andrew Altman eds., 2012) (“There continues to be debate over the exact parameters of non-international armed conflict, which are relevant to the controversy over the validity of the claim that the United States is, as a matter of law, engaged in a ‘transnational armed conflict’ with [a]l Qaeda and others.” (citations omitted)).

56. See Blum & Heymann, supra note 17, at 146–47 (explaining that an attack on an individual not wearing a uniform but suspected of involvement in terrorism must be based on the target’s status as a combatant or as an agent of a hostile force).

57. See infra Part II.A (highlighting moral issues implicated by targeted killing and related international law).
as to temper its inhumanity. This animating feature of IHL cautions against efforts to interpret or expand the war rules to nonparadigmatic cases, but it also cannot preclude the possibility of such expansion. The law, after all, has already once accommodated the reality that states employ more permissive rules of killing than are generally permitted.

Subpart B explores problems of application that arise when justifying targeted killings under IHL.58 Although the mere fact of combatants being individually targeted need not raise unique problems for IHL, the evolution of targeted killing as a distinct and significant method of warfare puts pressure on IHL, testing how the core legal requirements of necessity and discrimination apply to operations whose mechanics depart significantly from the types of combat that have traditionally informed IHL.

Subpart C details how the killing of Osama bin Laden presents an especially difficult test case for the IHL rules given that the functional mechanics of the operation—an isolated raid on a single house in a noncombat area—resemble in many respects the types of operations associated with law enforcement.59 We are therefore skeptical that the scope of the permission to kill should depend upon whether we label the operation a wartime attack as opposed to, for example, an operation to apprehend an especially dangerous criminal. Our conclusion does not change, moreover, even assuming that the United States could have permissibly elected alternate, less discriminating means of killing bin Laden, such as by firing a missile from an unmanned aerial vehicle.

A. The Morality of War and the Limits of International Law

As we just described, to invoke IHL in order to justify targeted killing is to invoke a more permissive legal framework toward killing. Underlying, but often ignored, in the interpretation and application of IHL is a broader moral question concerning why the law recognizes multiple legal paradigms for

58. See infra Part II.B (discussing the application of IHL to targeted killings).

59. See infra Part II.C (examining the circumstances of bin Laden’s killing with respect to the IHL framework).
killing at the outset. If society maintains that killing is wrong except in very narrow circumstances—such as to prevent the imminent use of lethal force by a wrongdoer—why should our moral judgment change when dealing with wartime scenarios? Why, in particular, should combatants not be bound by the same rules that apply to law-enforcement officials in peacetime? The question is not merely theoretical, as evidenced by the ongoing debate over the application of IHL to targeted killing. Whether and how we apply IHL to borderline or nonparadigmatic cases should be informed, to some degree, by the moral arguments supporting IHL’s creation.

Although the morality of killing in wartime is the central focus of multiple intellectual traditions, including just-war theory, the United Nations charter system, and, of course, IHL itself, the nature of the permission to kill in war remains, in many ways, perplexing.60 The conventional approach to justifying this state of affairs generally proceeds along the following lines: A state’s decision to employ the barbaric tactics of war must first be justified only by overriding interests of utmost importance, most paradigmatically the interest of self-defense against an armed attack by another state.61 The nature of this interest further necessitates, under strictly defined circumstances, the tactics commonly associated with warfare. A state facing an invading army will likely find it impossible to fend off the attack if it commits itself to rules of engagement that limit killing to the circumstances permitted by the law-enforcement paradigm.62

60. See JEFF McMahan, KILLING IN WAR 104–07 (2009) (discussing the conflation of morality and legality in the body of international law governing armed conflicts).

61. The right of self-defense is enshrined in Article 51 of the U.N. Charter, which provides that “[n]othing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security.” U.N. Charter art. 51.

62. This reasoning finds expression in customary international law’s general requirement that the use of armed force in self-defense be both necessary and proportionate, precluding the use of force beyond that necessary to repel an attack or restore the status quo ante. See, e.g., Oil Platforms (Iran v. U.S.), 2003 I.C.J. 161, para. 74 (Nov. 6) (“[W]hether the response to the [armed] attack is lawful depends on observance of the criteria of the necessity and proportionality of the measures taken in self-defence.” (second alteration in
Therefore, IHL grants a broader permission to kill, albeit the permission is not absolute. Like the law-enforcement paradigm, IHL places certain non-negotiable limits on killing.\textsuperscript{63} However, it draws the line in a different, more permissive place.

Those who accept justifications of this nature still have difficulty explaining why international law should embrace the strict separation between \textit{jus ad bellum} and \textit{jus in bello}, such that combatants fighting for illegal aggressors benefit from the same permissive rules that apply to those acting in legitimate self-defense. Walzer has defended this “moral equality of soldiers” by appealing to the moral intuition that it is unfair to blame soldiers merely for fighting wars on behalf of their state, when doing so reflects their “routine habits of law-abidingness, their fear, their patriotism, their moral investment in the state.”\textsuperscript{64} Moreover, because most combatants will predictably perceive themselves to be fighting for the just side in war, a rule that privileges just combatants over unjust combatants could serve to increase the cruelty of war.\textsuperscript{65} There would be less incentive to comply with even minimal rules of humane treatment if combatants expect to be prosecuted by the opposing side merely for participating in combat.\textsuperscript{66} Yet, even if one accepts these arguments, it remains difficult to explain why the law should not

\textsuperscript{63} See MELZER, \textit{supra} note 11, at 176 (stating that IHL, like the law-enforcement paradigm, requires necessity, proportionality, and precaution).

\textsuperscript{64} WALZER, \textit{supra} note 40, at 39; see also GEORGE P. FLETCHER & JENS DAVID OHLIN, DEFENDING HUMANITY: WHEN FORCE IS JUSTIFIED AND WHY 21–22 (2008) (“The reason for adopting a rigorous distinction between \textit{jus ad bellum} and \textit{jus in bello} is the need for a bright-line cleavage . . . . Soldiers . . . know that, regardless of who started the conflict, certain means of warfare are clearly illegal.”).

\textsuperscript{65} See WALZER, \textit{supra} note 40, at 127–28 (stating that soldiers “are most likely to believe that their wars are just,” and that the perspective of all soldiers as morally equal allows for rules governing wartime conduct).

\textsuperscript{66} See MCMAHAN, \textit{supra} note 60, at 191 (“[Soldiers] might reason, for example, that if they will be punished in any case if they are defeated, . . . each might have nothing to lose . . . from the commission of war crimes or atrocities that would increase their chance of victory and thus of immunity to punishment.”).
impose some duty to resist participation upon combatants who know that they are fighting an unjust war.67

As a historical matter, IHL did not emerge from abstract deliberations concerning when killing might be morally permissible. The modern *jus in bello* rules, developed through international agreement in the nineteenth and twentieth centuries, reflect instead an evolving response to the established reality of warfare. For instance, the first international agreement to prohibit the use of a weapon of war—the 1868 Saint Petersburg Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight68—speaks of the “progress of civilization . . . alleviating as much as possible the calamities of war” and of “conciliat[ing] the necessities of war with the laws of humanity.”69 The literature on IHL is replete with similar indications of a compromised morality, in which the interests of humanity must compete against those of military necessity.70

As Jeremy Waldron has written, the development of IHL in this way “proceeded on the basis of moral sociology, discerning the possibility of a viable norm of restraint in this area,” one “that has emerged from centuries of ghastly conflict.”71 That IHL relaxes the rule against murder, in other words, does not imbue killing in war with a deep moral justification. Instead, IHL

---

67. See id. at 6 (arguing that “it is morally wrong to fight in a war that is unjust because it lacks a just cause”). The author presents a systematic argument rejecting the separation between *jus ad bellum* and *jus in bello*. See id. (arguing “against the view that unjust combatants act permissibly when they fight within the constraints of the traditional rules of *jus in bello*”).


69. Id. at 474–75.

70. See, e.g., Int’l Comm. of the Red Cross, Commentary to Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 392–93 (“The law of armed conflict is a compromise based on a balance between military necessity, on the one hand, and the requirements of humanity, on the other.”); id. at 399 (“Warfare entails a complete upheaval of values.”).

principally seeks to impose at least some “regulative line that can be defended (just!) in the midst of an activity that is otherwise comprehensively murderous.”

This feature of IHL has at least two important implications for the debate over targeted killing. First, as Waldron argues, one should exercise great caution about expanding IHL’s permissive approach to killing. In particular, one should avoid the automatic assumption that IHL reflects a principled approach to killing that is readily susceptible to expansion by analogy. That, as Waldron argues, is precisely “how a norm against murder unravels.” This caution is warranted both with respect to expanding the IHL rules beyond their present scope and to interpreting the requirements of IHL as they apply within the acknowledged scope of the law.

At the same time, the uncertain moral status of IHL also makes it difficult to identify absolute limits on when states may develop more permissive approaches to killing. In this respect, the development of IHL stands as an important precedent for the law’s adjusting the norm against murder to accommodate the inevitability of certain state practices. Notwithstanding the extreme caution that is warranted by this exercise, the very existence of IHL also makes it difficult to preclude the development of other analogous accommodations. If, for example, technological advances facilitate the effective use of targeted killing by states to combat terrorist organizations in ways that fall outside the traditional boundaries of IHL, it becomes difficult to say when precisely the law must hold the line and forbid the practice or when, by contrast, the law may once again surrender and draw a different line, one that tolerates the general practice while seeking, as IHL does, to align it with some basic, more minimal, principles of humanity.

72. Id.
73. See id. (“Understanding the background just outlined helps us understand the great caution that must be brought to any attempt to change the laws of war.”).
74. See id. at 128 (arguing that, in this context, the common analogies are “all reckless ways to proceed”).
75. Id. at 131.
It is instructive here to draw a contrast between the issue of targeted killing and another debate that has figured prominently in recent years, namely, the debate over U.S. interrogation policy with respect to detainees at Guantanamo Bay and elsewhere. Those who condemn the waterboarding of detainees, as well as other “enhanced interrogation techniques,” may persuasively invoke an absolute legal ban on any practices that qualify as torture as defined under international law. This prohibition appears most prominently in the United Nations Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, an international treaty that has received near-universal ratification by states. In addition to prohibiting both torture and cruel, inhuman, or degrading treatment, the Convention defines torture as a criminal offense subject to extradition obligations and specifies that states may not torture individuals for purposes of information gathering or


79. See U.N. Convention Against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment, supra note 77, art. 1 (defining torture); id. art. 4 (mandating that “acts of torture” be punishable criminal offences); id. art. 8 (mandating that these acts be “included as extraditable offences”); id. art. 16 (requiring states to prevent public officials’ “other acts of cruel, inhuman or degrading treatment or punishment which do not amount to torture as defined in article 1”).
any other reason.\textsuperscript{80} It also states that “[n]o exceptional circumstances whatsoever, whether a state of war or a threat of war, internal political instability or any other public emergency, may be invoked as a justification of torture.”\textsuperscript{81} Thus, opponents of torture may invoke the Convention as evidence that the international community has given the question of torture explicit attention and has judged that an absolute ban is required, one that allows no derogation.

The problem of killing, by contrast, is accompanied by no such absolute ban. And the ways in which IHL already permits killing are dependent on precisely the type of instrumental reasoning that international law rejects in the torture context.\textsuperscript{82} Again, this is not to say that one should proceed lightly in justifying expansive permissions for government-sponsored killing. The norm against murder, as Waldron notes, is not just any norm, and proposals to further dilute it must be met with great skepticism.\textsuperscript{83} But the structure of IHL also cautions against absolute proclamations.\textsuperscript{84}

In Parts III and IV, we will explore further the possibility of justifying some targeted killing outside of both IHL and traditional law-enforcement paradigms. In the remainder of this Part, we explore reasons why, when IHL does govern a targeted

\textsuperscript{80} See \textit{id.} arts. 1–2 (defining torture and requiring measures to prevent torture).

\textsuperscript{81} \textit{Id.} art. 2, para. 2. The Convention also obligates each state to “undertake to prevent in any territory under its jurisdiction other acts of cruel, inhuman or degrading treatment or punishment which do not amount to torture.” \textit{Id.} art. 16, para. 1. The Convention does not explicitly define these practices or regulate them with the same detail it employs for torture, but neither does the Convention acknowledge any exceptions to this prohibition. See \textit{id.} (prohibiting cruel, inhuman, or degrading treatment without listing any exceptions).

\textsuperscript{82} See \textit{id.} arts. 1–2 (prohibiting torture for any reason, including state of war or other public emergencies).

\textsuperscript{83} See Waldron, \textit{supra} note 71, at 131 (“What is objectionable is the inherently abusive character of the attitude [behind] reasoning that says: ‘We are allowed to kill some people by principles we already have; surely, by the same reasoning, in our present circumstances of insecurity, there must be other people we are also allowed to murder.’”).

\textsuperscript{84} See, e.g., U.N. Convention Against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment, \textit{supra} note 77, arts. 1–16 (refusing to proscribe killing explicitly and absolutely).
killing, this body of law is not as determinative as is often supposed.85

B. IHL and Targeted Killing

1. The Permission to Kill by Name

A threshold question is whether a practice of individualized, “named” killing is ever compatible with IHL. Some scholars have questioned the practice of targeted killing on the grounds that it personalizes the conduct of war in a manner incompatible with the underlying moral vocabulary of war.86 Michael Gross, for example, traces a prohibition on named killing to the seminal Lieber Code of 1863,87 which admonishes:

The law of war does not allow proclaiming either an individual belonging to the hostile army, or a citizen, or a subject of the hostile government, an outlaw, who may be slain without trial by any captor, any more than the modern law of peace allows such international outlawry; on the contrary, it abhors such outrage.88

As Gross elaborates, the permission to kill in wartime presupposes the moral belligerents’ innocence, and both the combatants’ permission to kill and their vulnerability to being killed derives from their role as impersonal agents of a collective entity, the state.89 The problem with targeting individuals on this

85. See infra notes 86–148 (discussing the application of IHL to targeted killing).

86. See, e.g., Gross, supra note 17, at 327–29 (questioning the morality of targeted killing in wartime); Statman, supra note 17, at 190 (arguing that “the problem with targeted killing” is that it undermines “the very justification for killing in war,” which is that we ignore an enemy’s personal merits or demerits).


88. Id. art. 148.

89. See Gross, supra note 17, at 326 ("No one is suggesting that soldiers do not represent material threats . . . . [A]ny uniformed soldier is vulnerable. . . . Soldiers may kill in the service of their state and are therefore innocent of any wrongdoing, a sweeping authorization that international law and all nations endorse.").
account is that “[o]nce we name soldiers for killing, . . . we upset this innocence with precisely the argument that Lieber presents. Naming names assigns guilt and, as Lieber suggests, proclaims soldiers outlaws. In doing so, named killing places war itself beyond convention.”

We do not share Gross’s concern about named killing. More precisely, we do not believe that it is the practice of naming targets itself that triggers Gross’s concern. Indeed, even the Lieber Code passage just quoted addresses a concern distinct from the practice of named killing per se. The Code focuses on a broad class of potential victims, including not only those belonging to “the hostile army,” but also those who are merely “citizen[s]” or “subject[s],” and thus outside the class of those permissibly targetable in war. It is, moreover, concerned with the proclaiming of individuals as “outlaw[s]” such that they “may be slain without trial by any captor.” The focus, therefore, is on the punitive killing of captured persons. The Code does not address whether belligerents may conduct named killings for other reasons and in other contexts.

There are indeed contexts in which named killing may operate as a routine battlefield practice without violating the moral innocence of targeted combatants. Fernando Tesón supplies the helpful example of a soldier who is targeted on the battlefield because he is an especially skilled machine gunner, who has proven adept at cutting down opposing soldiers seeking to advance. He maintains, correctly in our view, that it is permissible to target the machine gunner as a battlefield strategy. This sort of individualized targeting of an especially

---

90. Id.
91. Lieber Code, supra note 87, art. 148.
92. Id.
93. See Fernando R. Tesón, Targeted Killing in War and Peace: A Philosophical Analysis, in Targeted Killings: Law and Morality in an Asymmetrical World 403, 411 (Claire Finkelstein, Jens David Ohlin & Andrew Altman eds., 2012) (giving the example of soldiers on a battlefield who target a particularly effective enemy machine gunner).
94. See id. (“He is an unjust enemy combatant and as such may be permissibly killed, named or unnamed.”).
dangerous opponent reflects a straightforward pursuit of military advantage.95

Suppose now, as Tesón posits, that the opposing army happens to have learned the name of its target. Suppose, for example, that the belligerents facing the skilled machine gunner know that he is the infamous “Private Gunner.”96 The permissibility of an order to “take out Private Gunner before he cuts down any more of us” does not become illegal, or even immoral, merely because the target has now been named. What makes this an easy case is that the naming is incidental. Although Gunner is individually targeted, he is not targeted in his individual capacity. In this case it is wrong to say that “[n]aming names assigns guilt and . . . proclaims soldiers outlaws.”97 The targeting of Gunner preserves his moral innocence; it recognizes the danger that he poses as an especially effective combatant.

As Gross himself recognizes,98 such naming also assumes importance in the context of conflicts against terrorist groups who do not observe the formalities of war, including the core requirements that combatants carry arms openly, have a “fixed distinctive sign recognizable at a distance,” and “conduct[] their

95. One need not draw on hypothetical examples to justify such targeting. Consider the example of World War II German panzer ace Michael Wittmann who, at the battle of Villers-Bocage, commanded a tank that destroyed, within a space of fifteen minutes, at least eleven Allied tanks, nine half-track vehicles, four troop carriers, and two anti-tank guns, singlehandedly prompting Allied forces to abandon the recently captured village before a single German reinforcement had arrived. See George Forty, Villers Bocage 57–86 (Simon Trew ed., 2004) (describing Michael Wittmann’s fighting at Villers-Bocage and stating that he personally destroyed “seven cruiser/medium tanks (including one Firefly), three Stuart light tanks, one Sherman OP, nine half-tracks, four carriers and two anti-tank guns”); Daniel Taylor, Villers-Bocage: Through the Lens of the German War Photographer 33 (1999) (describing Wittmann’s role in the battle and stating that he destroyed “seven gun tanks, one of which may have been a Firefly, three Stuarts, one Sherman OP, nine half-tracks, four carriers and two anti-tank guns”). To the extent Allied forces were subsequently capable of identifying Wittmann on the battlefield, surely they would be justified in taking special effort to destroy his tank.

96. In Tesón’s original hypothetical, he is named “Colonel Sanders.” Tesón, supra note 93, at 411.

97. Gross, supra note 17, at 326.

98. See id. at 329 (assessing the appropriateness of named killing against terrorists).
operations in accordance with the laws and customs of war.”

Those who fail to qualify as privileged combatants may nevertheless be attacked “for such time as they take a direct part in hostilities.” The reach of this provision is currently a matter of controversy. A civilian targeted in the act of firing a machine gun on the battlefield is an easy case, and the legal advisor to the International Committee of the Red Cross (ICRC) has further opined that civilians who have assumed a “continuous combat function” may be treated as regular members of organized armed groups.

More divisive is how to apply IHL to terrorism suspects who may spend months or even years planning attacks in hiding. Assuming we do treat such persons as belligerents, a practice of named killing may in fact be essential because, as Gross observes, there are no means to identify such persons as belligerents absent individualized assessment. Accepting this logic does not, however, require us to adopt a special terrorism-based exception to a prohibition against named killing, as Gross would seem to suggest. Instead, it merely reinforces the broader point that there is no inherent tension between the

99. POW Convention, supra note 48, art. 4(A)(2).
100. Additional Protocol I, supra note 39, art. 51(3).
101. See Gross, supra note 17, at 326 (noting the lack of consensus “about the status of those who belong to an organization that does not meet the minimal standards set by [Additional] Protocol I”).
103. See, e.g., Jens David Ohlin, Targeting Co-Belligerents, in TARGETED KILLINGS: LAW AND MORALITY IN AN ASYMMETRICAL WORLD 60, 63 (Claire Finkelstein, Jens David Ohlin & Andrew Altman eds., 2012) (discussing the threats terrorists pose, individually and collectively, and arguing for a membership-based approach to identifying belligerents).
104. See Gross, supra note 17, at 330 (discussing the problem of identifying as combatants “guerrillas, militants, terrorists and others without uniforms” and stating that naming can be useful or even necessary in such cases).
105. See id. at 331 (“Perhaps targeted killings are an appropriate response to terrorism precisely because terrorists deserve to suffer harm in a way that just combatants do not.”).
concept of named killing itself and the paradigm of international humanitarian law.

2. Targeted Killing in a Global Battlefield

None of this is to say, however, that there is nothing in the practice of naming targets that challenges the structure of IHL. Returning to Tesón's machine-gunner example, suppose that military officials deem Gunner so dangerous that, so long as the war continues and he remains in military service, he is to be tracked and killed wherever he may be found. Intelligence officials trace Gunner to an island resort where he is enjoying an extended vacation with his family. A special commando unit is dispatched to kill him with instructions to follow the usual war rules of engagement. The commandos locate Gunner and shoot him while he is sunbathing on the beach. We suspect that many readers will be more troubled by this scenario than by our previous hypothetical, although it can be difficult to explain why.

a. The Feasibility of Capture

One troubling aspect of this hypothetical scenario is that the commandos have elected to neutralize Gunner's machine-gun skills by killing him instead of capturing him and holding him as a prisoner of war until the termination of hostilities. Note that under Koh's justification of bin Laden's killing, this aspect of the operation is permissible so long as Gunner has not clearly communicated his surrender. If we embrace Koh's statement as reflecting an inflexible rule applying to all belligerents in war, then it makes no difference to the law that Gunner poses no risk at all to the commandos, or that capture would be easy under the circumstances.

106. See supra notes 14–16 and accompanying text (describing Koh's justification of the bin Laden killing).

107. Note that U.S. government officials have indicated that the feasibility of capture may be a factor in the lawfulness of a targeted killing when the killing is directed against a U.S. citizen, as was the case with Anwar al-Aulaqi. See Holder, supra note 16 (stating that targeted killing of a U.S. citizen terrorist is lawful at least when there is an imminent threat of violent attack, "capture is
This, however, cannot be an accurate statement of the *jus in bello* rules, at least not if they are to be interpreted and applied in a way that is consistent with a plausible account of morality. The question of whether and when international law imposes a duty to capture rather than kill lawful targets of war has provoked debate in recent years. The Supreme Court of Israel’s 2005 ruling on Israel’s targeted killing policy stated that “a civilian taking a direct part in hostilities cannot be attacked at such time as he is doing so, if a less harmful means can be employed,” and the ICRC’s subsequent non-binding guidance on the notion of direct participation in hostilities argued more broadly that “it would defy basic notions of humanity to kill an adversary or to refrain from giving him or her an opportunity to surrender where there manifestly is no necessity for the use of lethal force." But these pronouncements have not attracted widespread acceptance; the ICRC’s conclusion, in particular, faced significant resistance from many of the experts whom the ICRC had invited to participate in its study.

The doctrinal argument favoring attempted capture in these circumstances appeals to a requirement of military necessity, the idea that the law permits “only that degree and kind of force, not otherwise prohibited by the law of armed conflict, that is required in order to achieve the legitimate purpose of the conflict, namely not feasible,” and the operation would be conducted according to law of war principles). This distinction between citizens and non-citizens does not derive from IHL, and official statements regarding the killing of bin Laden and other foreign terrorists have not included the general feasibility of capture as a legally relevant factor, except in the narrowly defined circumstances Koh identified, which require “a genuine offer of surrender that is clearly communicated” and that may be safely received. Koh, supra note 14.


109. ICRC Guidance, supra note 102, at 82.

110. See id. at 82 n.221 (“During the expert meetings, it was generally recognized that the approach [to choose capture over killing] is unlikely to be operable in classic battlefield situations . . . and that armed forces . . . may not always have the means or opportunity to capture rather than kill.”); W. Hays Parks, *Part IX of the ICRC “Direct Participation in Hostilities” Study: No Mandate, No Expertise, and Legally Incorrect*, 42 N.Y.U. J. INT’L L. & POL. 769, 799–801 (2010) (summarizing experts’ objections to Part IX of the ICRC Guidance).
the complete or partial submission of the enemy at the earliest possible moment with the minimum expenditure of life and resources.”111 The problem with this appeal is that IHL has traditionally declined to translate the necessity principle into rules that protect the lives of combatants themselves.112

This asymmetry finds expression in the Civil-War era Lieber Code, which provides that “[m]ilitary necessity admits of all direct destruction of life or limb of ‘armed’ enemies.”113 A similar selectivity appears in the requirements of Additional Protocol I, which gives effect to the restraint of military necessity. The employment of “weapons, projectiles and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering” is forbidden.114 Objects are not proper military targets unless their destruction “offers a definite military advantage.”115 Attacks on military targets are forbidden if expected to result in incidental loss to civilians or civilian objects that is excessive compared to the “concrete and direct military advantage anticipated.”116 Yet no parallel provision imposes a necessity-based limitation on the use of lethal force against combatants themselves.117

How does one reconcile this omission with the general principle that necessity justifies the permissive rules of IHL? The most plausible explanation appeals to military necessity itself.

111. ICRC Guidance, supra note 102, at 79 (quoting United Kingdom Ministry of Defence, The Manual of the Law of Armed Conflict, ¶ 2.2). This language closely mirrors that set forth in the Lieber Code. See Lieber Code, supra note 87, art. 14 (“Military necessity, as understood by modern civilized nations, consists in the necessity of those measures which are indispensable for securing the ends of the war, and which are lawful according to the modern law and usages of war.”).
112. See infra note 117 and accompanying text (pointing out the inconsistent application of the necessity principle).
113. Lieber Code, supra note 87, art. 15. Only with respect to “other persons” does the Code require that destruction be “incidentally unavoidable.” Id.
114. Additional Protocol I, supra note 39, art 35.
115. Id. art. 52(2).
116. Id. art. 51(5)(b).
117. See Parks, supra note 110, at 804 (“There is no ‘military necessity’ determination requirement for an individual soldier to engage an enemy combatant or a civilian determined to be taking a direct part in hostilities, any more than there is for a soldier to attack an enemy tank.”).
The law generally permits the killing of all enemy combatants, without imposing any obligation to assess the military value of individual targets or to consider less lethal means, because the modalities of war generally do not permit those pursuing military advantage to make such individualized assessments. The nature of armed conflict through history has been to pursue victory precisely through the destruction of the opposing forces. In the context of the traditional battlefield, belligerents are generally justified in treating all opposing forces as sources of danger and as targets whose destruction has military advantage. A rule to the contrary, one that requires combatants to make individualized threat assessments for each targeted enemy, would burden the waging of war in ways not acceptable to states, including by exposing combatants to increased risk. On this

---

118 Such practical considerations inform W. Hays Parks’s opposition to Part IX of the ICRC’s Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law. See Parks, supra note 110, at 809–10 (arguing against a necessity-based limit on the use of lethal force against combatants). Parks reasoned:

> [O]ther than the law of war prohibitions on perfidy and denial of quarter, governments and courts have seen the prudence in declining to draw such a line owing to the many vagaries that exist not only in domestic law enforcement situations but also, and in particular, on the battlefield. This is the case in combat in recognition of the obligation imposed by many nations on their military forces not to surrender and, indeed, to resist surrender either by force or through escape and evasion.

Id. On this point, Parks is especially concerned by the prospect that the ICRC Guidance would require individual soldiers to apply a “use-of-force continuum.” See id. at 796 (arguing that the ICRC Guidance wrongly “resurrects and offers Pictet’s unaccepted use-of-force continuum theory as if it were an internationally accepted, binding legal formula”). Pictet’s theory is best summarized by his statement that

> [i]f we can put a soldier out of action by capturing him, we should not wound him; if we can obtain the same result by wounding him, we must not kill him. If there are two means to achieve the same military advantage, we must choose the one which causes the lesser evil.

Jean Pictet, Development and Principles of International Humanitarian Law 75–76 (1985). The ICRC Guidance, however, interprets this statement to support the more basic point that “it would defy basic notions of humanity to kill an adversary or to refrain from giving him or her an opportunity to surrender where there manifestly is no necessity for the use of lethal force.” ICRC Guidance, supra note 102, at 82. The ICRC Guidance also acknowledges that
account, the general rule permitting the targeting of all combatants who are not *hors de combat* is not so much a violation of the necessity principle as an expression of it.

This account loses all plausibility, however, when we imagine circumstances, such as the killing of Gunner on vacation, that bear no resemblance to the battlefield modalities that have historically animated IHL, and that present no functional reason why a more permissive rule to killing should prevail. When Gunner is found isolated from active combat or other military presence and is defenseless in a bathing suit, there is no account of military necessity with which one can make sense of a rule that permits killing short of surrender, or indeed of any rule that would be more tolerant of killing than that applicable to law-enforcement officials seeking to apprehend Gunner under like circumstances.

One could argue that Gunner's utter defenselessness has already placed him “in the power of an adverse Party” and thus rendered him *hors de combat*. The ICRC Commentary to Additional Protocol I supports this view. If that is correct, then the permission to kill Gunner under IHL is already weaker than suggested by Koh's account of IHL in the context of the bin Laden killing. More broadly, however, a rule permitting the killing of Gunner under these circumstances runs up against a deep structural constraint that is foundational to IHL: the idea that IHL's expanded permission to kill is, at some level, linked to and...

---

Pictet's proposed approach “is unlikely to be operable in classic battlefield situations involving large-scale confrontations and that armed forces operating in situations of armed conflict, even if equipped with sophisticated weaponry and means of observation, may not always have the means or the opportunity to capture rather than kill.” *Id.* at 82 n.221 (citations omitted).


120. See Int'l Comm. of the Red Cross, *supra* note 70, at 484 (maintaining that Protocol I reflected the determination of states to embrace the protections of *hors de combat* status in cases in which land forces “have the adversary at their mercy by means of overwhelmingly superior firing power to the point where they can force the adversary to cease combat”). This claim has support in the change from the Third Geneva Convention’s reference to those who have “fallen into the power of the enemy” to Protocol I’s parallel reference, “in the power of an adverse Party.” *Id.* “A defenceless adversary,” the Commentary continues, “is 'hors de combat' whether or not he has laid down arms.” *Id.*
justified by considerations of military necessity. The problem is not merely that Gunner’s specific killing lacks military necessity given the feasibility of capture but also that a general rule permitting him to be killed under these circumstances is itself irreconcilable with an account of military necessity, or indeed with the pursuit of any military advantage.

Accordingly, to identify a capture duty in these circumstances does not, as W. Hays Parks has feared, implement a fundamental alteration in IHL by requiring individual soldiers to always consider the feasibility of capture before deploying lethal force. Instead, our conclusion rests on the more modest insight that IHL’s generally broad permission to kill reflects some basic assumptions about battlefield modalities. When IHL travels to contexts that, as a functional matter, fail the most expansive definition of a battlefield, the overarching requirement of necessity must be appraised anew.

Some scholars have reached a similar conclusion on the ground that targeted killing occupies a hybrid position between the IHL paradigm and the law-enforcement paradigm. This, for instance, is how Gabriella Blum and Philip Heymann understand the Israeli Supreme Court’s ruling that IHL properly applied to Israel’s policy of targeted killings, but that the legality of those killings hinged on the absence of a reasonable alternative for capturing the targeted terrorists. Or, as Nils Melzer would see it, the case we have just presented supports the view that necessity itself must dictate whether the legal paradigm is IHL or law-enforcement. In other words, the law-enforcement

121. See supra note 111 and accompanying text (discussing the importance of military necessity in state-sanctioned killing).

122. Given the lack of necessity, moreover, any resultant risk to civilians posed by employing lethal force against Gunner would necessarily be impermissible under the proportionality test. See Additional Protocol I, supra note 39, art. 51(5)(b) (protecting civilians from harm that “would be excessive in relation to the concrete and direct military advantage anticipated”).

123. See supra note 118 (discussing Parks’s concern).

124. See Blum & Heymann, supra note 17, at 158 (arguing that “the court’s requirement to try to apprehend the terrorist is far more easily situated within a law-enforcement model of regular policing operations and signifies the uneasiness that the court felt about the war paradigm”).

125. See ICRC Guidance, supra note 102, at 82 (“[I]t would defy the basic notions of humanity to kill an adversary or to refrain from giving him or her an
paradigm applies to the Gunner killing because there is no necessity supporting the invocation of the IHL paradigm. Whether we phrase the argument that way or not is largely a matter of semantics in this instance, however, because we arrive at the same conclusion under the IHL paradigm itself. Indeed, in our example, Gunner is a lawful combatant participating in a traditional armed conflict, and we have made no assumptions that his targeting is informed by law-enforcement considerations of any sort. Either paradigm yields the same result under the facts presented.

Critically, the difficulties we have just described are not a direct consequence of named targeting per se. It is not the act of naming Gunner for killing that complicates the decision to kill rather than capture. At the same time, this problem is one that is distinctly associated with a practice of individualized targeting. The impulse to follow Gunner across the world is one that derives from assigning unique danger to him as an individual, rather than as an anonymous and substitutable member of the broader collective force that remains engaged on the battlefield. And it is this impulse that puts pressure on IHL by asking us how its requirements apply under nonparadigmatic conditions that do not resemble those of a conventional battlefield.

b. The Harm of a Diffuse Battlefield

Killing Gunner on vacation also raises other problems that are distinct from the question of capture. We may bring these into focus by assuming that capturing Gunner is infeasible. Perhaps there is only a short window of time to neutralize him before he becomes untraceable, and the only feasible option is a missile strike. The operation now entails killing Gunner and unavoidable risk to the civilians in Gunner’s proximity.

This version of the hypothetical more directly raises the issue of proportionality: is the risk to civilians justifiable in light of the military advantage to be gained by killing Gunner?\textsuperscript{126} But the

\textsuperscript{126} Additional Protocol I, \textit{supra} note 39, art. 51(5)(b).
problem is also broader than the direct risk to specific civilians. There is something unsettling in the spread of war tactics beyond the conventional battlefield, in the idea that war may follow individual belligerents wherever they go. It is disturbing in particular to reflect on the fate of the persons and property in the area surrounding Gunner, that may now be victimized by *jus in bello* rules that treat them as potential collateral damage whose incidental destruction may be weighed against the utility of targeting Gunner. This fear that war may strike anywhere at any time has figured prominently in recent debates over U.S. targeted-killing policy. Critics have asked whether, with the inevitable spread of technology, we are willing to commit to a reading of the law that would allow, for example, Iran to launch missile strikes on U.S. cities against discrete individuals whom it considers to be enemy combatants.\(^\text{127}\)

Typically, this problem is treated as one of *jus ad bellum*. International law takes account of the dangers of spreading war by limiting the conditions under which states may permissibly resort to armed force,\(^\text{128}\) and the targeting of Gunner may constitute an act of aggression against the state where he is vacationing. The conditions under which a state may resort to armed force on the territory of a neutral third state is currently a matter of debate,\(^\text{129}\) but that debate is not the end of the matter. The *jus ad bellum* issue may be overcome: perhaps Gunner is vacationing in a state that has given its consent to the raid,\(^\text{130}\) or that is already party to the conflict. We will further assume, for

\(^{127}\) See, e.g., Stephen L. Carter, *The Violence of Peace: America’s Wars in the Age of Obama* 60–61 (2010) (acknowledging that a similar strike against the United States would be justified if our targeted killings are justified against those we deem terrorists).

\(^{128}\) See supra II.A (discussing IHL’s limits on use of force).

\(^{129}\) See supra note 54 (providing several scholars’ discussions about targeting terrorist suspects on third-party territory).

\(^{130}\) Indeed, in many cases of targeted killing, including in northwest Pakistan and in Yemen, the United States has benefitted from persuasive claims to state consent. See, e.g., Van Schaack, *supra* note 2, at 267 (noting that targeting killing operations in northwest Pakistan “likely enjoy at least some tacit diplomatic acquiescence, even though Pakistani officials occasionally publicly criticize them for domestic political consumption”); id. at 266 (observing that the operation against Anwar al-Aulaqi “appears to have had the benefit of Yemen’s consent and perhaps its involvement”).
purposes of this discussion, that IHL applies away from the traditional battlefield and follows Gunner to his vacation spot.\textsuperscript{131}

There is some debate in the literature concerning whether a combatant is ever targetable when located away from the battlefield.\textsuperscript{132} Yet even assuming that IHL governs this scenario, the attack remains troubling pursuant to IHL’s requirement of proportionality. In real life, even an ace machine gunner like Gunner is unlikely to be perceived, simply on account of his combat skills, as sufficiently dangerous to justify the type of dedicated operation we have imagined. We may therefore also question whether it is permissible, consistent with the demands of proportionality, to risk civilian lives in this way merely to neutralize a single combat soldier.

Arguments of this nature are complicated by the indeterminacy of the legal standard. To compare—as the proportionality formula requires—the “concrete and direct military advantage” against the expected “incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof”\textsuperscript{133} requires the weighing of incommensurables.\textsuperscript{134} The law also fails to specify the extent to which combatants must put their own forces at risk in order to reduce risks to civilians.\textsuperscript{135} As a consequence, application of the

\textsuperscript{131} On problems associated with the geography of IHL, see Laurie R. Blank, \textit{Defining the Battlefield in Contemporary Conflict and Counterterrorism}, 39 GA. J. INT’L & COMP. L. 1, 2–6 (2010).

\textsuperscript{132} Compare Tesón, supra note 93, at 412–15 (arguing that a combatant on vacation is an impermissible target), with Statman, supra note 17, at 196 (rejecting the distinction between a combatant in military headquarters and a combatant on vacation).

\textsuperscript{133} Additional Protocol I, supra note 39, art. 51(5)(b).

\textsuperscript{134} See Final Rep. to the Prosecutor by the Comm. Established to Review the NATO Bombing Campaign Against the Fed. Republic of Yugoslavia, ¶ 48 (June 13, 2000), available at www.icty.org/sid/10052 (“Unfortunately, most applications of the principle of proportionality are not quite so clear cut. It is much easier to formulate the principle of proportionality in general terms . . . . One cannot easily assess the value of innocent human lives as opposed to capturing a particular military objective.”).

\textsuperscript{135} Additional Protocol I requires combatants to take “all feasible precautions” to avoid loss of civilian life but does not specify what is meant by “feasible.” Additional Protocol I, supra note 39, art. 57(2)(a)(ii). On the problem of force protection versus enemy civilian protection, see David J. Luban, \textit{Risk Taking and Force Protection} 38 (Georgetown Pub. Law Research, Paper No. 11-
proportionality formula inevitably extends great deference to military judgment.

Does this already-difficult calculus necessarily change because Gunner is on vacation? We think it does. Although it is impossible to demark the precise point at which claims of military advantage fail to outweigh incidental risk to civilians, we believe that many will intuitively and justifiably conclude that any version of our Gunner-on-vacation hypothetical involving risk to civilians presents a case in which an attack on a military target cannot be justified in light of the danger posed to civilians.

With respect to the military advantage side of the equation, one difference between the battlefield and vacation settings is that the killing of Gunner on the battlefield more readily contributes to an immediate, more general military goal. If, for example, Gunner is killed during a battle to seize a strategic village, advance the front line, or defend a position, then the value of his individual death is subsumed by the military advantage associated with that broader objective. Even though he is targeted individually, the targeting occurs in the context of an operation that necessitates the killing of enemy combatants, and, in that broader context, there are sound reasons to give special attention to the threat posed by Gunner. When Gunner is specifically targeted on vacation, by contrast, one is forced to focus on him in isolation and ask how much military advantage is anticipated from the neutralization of Gunner alone, outside of any immediate battlefield context. It is true that Gunner is now unable to return to the battlefield and thus is prevented from making future military contributions. But this causal link is more speculative and attenuated, and it is harder to demonstrate why Gunner merits such individualized attention.

The dangers to civilians likewise require more focused attention. When a rocket attack on the vacationing Gunner kills bystanders, it is clear that the lives lost are casualties of the attack on Gunner alone. Individual incidents on the battlefield may also have this but-for quality, but there is also a generalized risk that exists in areas of continuous combat. Consider the Allied advance in Normandy during World War II, which, by

some estimates, claimed close to 20,000 French civilian lives.\textsuperscript{136} At a broad level of generality, one can say that these large-scale civilian deaths were a tragic, but predictable, feature of the campaign to liberate France. Although that broad point did not relieve individual combatants of a duty to protect civilian life in particular engagements, it was easier to justify incidental civilian losses in the context of a broader campaign that necessarily imposed a generalized risk on all of Normandy’s residents. When Gunner is targeted on vacation, by contrast, it becomes clearer that the attack has put a class of civilians in danger, a class that faced no special risks otherwise. The risk to civilians, in other words, is a risk associated most directly with the decision to target Gunner. This feature of the attack only serves to exert further pressure on the planners of the attack to justify the special attention owed to Gunner.

One may, in addition, identify harm to Gunner himself. Although Gunner is targetable as a combatant in the theater of war, his liability to be killed derives not from personal culpability, but from his commitment—voluntary or not—to render himself liable through service as a combatant fighting for a broader collective. Although the point is arguable, one can reasonably maintain that there is a limit to this commitment, and that Gunner has a justified expectation that his liability to be killed in battle does not entail a liability to be singled out and chased across the world in this manner.

For reasons already addressed, it is impossible to demark the precise point at which claims of military advantage fail to outweigh incidental risk to civilians, but we believe that many will intuitively conclude that our Gunner-on-vacation hypothetical presents a case in which an attack on a military target cannot be justified if the attack poses any danger to civilians. These considerations all urge special caution with respect to targeted-killing operations that invoke the permissive IHL rules in contexts outside the traditional battlefield.

We do not argue, however, that IHL can or should be interpreted to impose a blanket ban on targeted killings. Our

\textsuperscript{136} See Antony Beevor, D-Day: The Battle for Normandy 519 (2009) (stating that 19,890 French civilians were killed during the liberation of Normandy).
hypothetical case of Gunner on vacation is admittedly far-fetched, and the cases of targeted killing that have provoked controversy in recent years have benefited from more compelling appeals to necessity than what we have imagined here. Our point, rather, is that when targeted killing moves away from the traditional battlefield, its permissibility will generally require special justification in at least two interrelated ways, both of which flow naturally from the application of IHL and do not borrow from the law-enforcement paradigm.

First, the targeted killing of an individual in these circumstances generally cannot be justified merely by the target’s status as a combatant. The targeting instead requires special justification regarding the threat that the individual poses, either in isolation or in combination with others who are similarly targeted. For example, such a justification may present itself when the target poses an imminent or especially high threat; the target wields significant authority within an enemy organization; or there is sound reason to believe that a policy of targeting like-situated persons will, in the aggregate, prove sufficiently feasible to have a significant military advantage. Second, when military operations target isolated individuals, the acceptable risk of civilian casualties must generally be lower than in more conventional military operations.

These considerations are especially salient to the difficult interpretive question of how to determine which nontraditional combatants may be targeted, pursuant to IHL, as civilians directly participating in hostilities. For example, even if one maintains (as we do not) that this category should be read broadly to encompass all members of terrorist organizations engaged in hostilities, that interpretation will be unlikely to satisfy the requirements of necessity and proportionality in concrete cases of targeted killing. Civilian lives should not be risked, for example, in order to kill a member of al Qaeda whose

137. See supra notes 93–97 and accompanying text (describing circumstances in which an individual may be targeted because of the especially dangerous threat that he poses).
138. See supra notes 43–44 and accompanying text (stating that civilians may be targeted if they participate in hostilities).
139. See, e.g., Ohlin, supra note 103, at 84–85 (suggesting that all members of terrorist organizations might be engaged in continuous combat function).
contribution to the group is known to consist only of cooking meals or sweeping floors. This remains the case whether or not one labels the target a combatant under IHL.

Notwithstanding the difficulties inherent in reducing these intuitions to concrete operational guidance, evolving U.S. policies affirm the special sensitivities that accompany targeted killing as a strategy of war. Gregory McNeal has documented that U.S. policy for preplanned aerial drone strikes in Afghanistan have generally required a collateral damage estimation process that is intended to ensure that there will be a less than 10% probability of serious or lethal wounds to non-combatants. He reports that “less than [1%] of pre-planned operations that followed the collateral damage estimation process resulted in collateral damage.” Moreover, when even one civilian casualty is expected, the President of the United States or the Secretary of Defense must personally approve the strike. Even considering that official U.S. casualty claims reportedly rely on a contested criterion that generally “counts all military-age males in a strike zone as combatants,” these policies reflect a concern for limiting incidental civilian casualties that goes far beyond the indeterminate guidance provided by Additional Protocol I.

140. The more difficult cases concern financiers and others who occupy significant positions in the group but whose functions are not obviously analogous to traditional combat functions. See, e.g., HCJ 769/02 Pub. Comm. Against Torture in Israel v. Gov’t of Israel ¶ 35 [2005] (Isr.), available at http://elyon1.court.gov.il/Files_ENG/02/690/007/a34/02007690.a34.pdf (noting that “a person who aids the unlawful combatants by general strategic analysis, and grants them logistical, general support, including monetary aid” does not directly participate in hostilities, but that there “is a debate surrounding the following case: a person driving a truck carrying ammunition”).


142. Id. at 331.

143. Id. According to a recent New York Times article, the President’s involvement is even greater than mandated by this policy. See Becker & Shane, supra note 23 (“[N]ominations go to the White House, where by his own insistence and guided by Mr. Brennan, Mr. Obama must approve any name. He signs off on every strike in Yemen and Somalia and also on the more complex and risky strikes in Pakistan—about a third of the total.”).

144. Becker & Shane, supra note 23.
3. The Problem of Guilt

Our final twist on the Gunner scenario deals with questions of culpability and mixed motives. Suppose that the officer who orders Gunner’s killing on the battlefield is motivated by reasons other than Gunner’s machine-gun skills. It turns out that Gunner is a villain who, before the war, murdered the officer’s family. The officer therefore invokes the rules of war opportunistically, as a cloak to disguise what is in fact a desire to exact revenge.

One might argue that Gunner, as a combatant in armed conflict, remains liable to be killed under the IHL rules, and thus has no right to complain of his killing. So long as hostilities persist, opposing combatants may kill Gunner at any time. From one perspective, that the officer has a grudge against Gunner should not matter any more than that the officer might be fighting only for reasons of financial self-interest rather than patriotism. The individual motives of combatants have, at best, limited relevance when combatants act in conformity with the broader policy of the state or collective entity they serve. And here it is the state’s policy that combatants such as Gunner should be attacked even when they are not villains.

 Nonetheless, we suspect that most readers will agree that the targeting of Gunner under the circumstances described is unsustainable under the IHL paradigm. The problem with killing Gunner on account of his past crimes is that his targeting is not informed in any way by military considerations. Thus, the problem is similar to the one presented by our Gunner-on-the-beach hypothetical. This culpability-driven targeting is precisely the sort of killing that justifiably triggers Gross’s objection, which he associates more broadly (and incorrectly in our view) with the very concept of targeted killing.145 Here, the killing assigns guilt to Gunner in his individual capacity, thereby abusing the permission to kill that derives solely from Gunner’s impersonal status as an agent of the state.146

---

145. See Gross supra note 17, at 328 (suggesting that targeted killings based on an assumption of guilt place these killings outside the scope of just-war conventions that allow killing combatants without due process of the law).

146. We do not attempt here to resolve more difficult permutations of our hypothetical in which the officer’s motives are mixed. For example, suppose
In Parts III and IV, we explore the possibility that individual culpability may, in special circumstances, factor into justifications for targeted killing that do not draw from the IHL paradigm.\(^{147}\) Note, however, that even within the structure of IHL, individual guilt may inform targeting decisions, provided it does so for reasons that speak to considerations of military necessity.\(^{148}\) If we suppose that Gunner is a notorious war criminal, his targeting serves to prevent further abuses against civilians and enemy combatants. Similarly, if Gunner is deeply involved in the perpetration of an ongoing genocide, whose prevention is the very cause of war, then targeting Gunner on culpability-based criteria serves a preventive function that directly advances the *jus ad bellum*. The critical distinction in both cases, however, is that the targeting remains rooted in preventive, rather than punitive, considerations.

The question of improper or mixed motives has obvious salience to the context of terrorism, in which government officials invoke IHL to justify the targeted killing of alleged combatants who are simultaneously accused of serious crimes. It is partly on account of this duality that the practice of targeted killing has elicited so much controversy over whether IHL or law enforcement is the correct legal paradigm. Our hypothetical brings to light a limitation of the IHL paradigm that receives no explicit attention in the law and is often overlooked in this debate: an attack that is objectively justifiable under IHL may nevertheless be impermissible because it is pursued for the wrong reasons. At the same time, this limitation is complicated by the pressing military considerations dictate that a machine gun be destroyed at some point during a particular twenty-four-hour period. The machine gun is continuously manned in shifts so that destroying the gun will also entail killing a machine gunner, and personal grudges lead the gun to be destroyed during the precise moment when Gunner is manning it. This example is more difficult because we may assume that the officer would never have ordered Gunner's killing absent pressing military necessity. Nevertheless, improper motives resolved the arbitrary question of precisely when to attack the machine gun.

\(^{147}\) See infra Parts III–IV (discussing the scope of permissible killing in peacetime and discussing possible revisions to contemporary theory).

\(^{148}\) See, e.g., Gross, supra note 17, at 327 (addressing the difficulties of placing targeted killings within the framework of either the IHL or law enforcement).
ambiguity that culpability-based targeting is sometimes justified by the logic of IHL when rooted in preventive rather than retributive considerations.

C. The War Paradigm and the bin Laden Killing

Although the hypothetical scenarios we have outlined do not focus on terror cases, they highlight many of the issues that have provoked controversy over targeted killing in recent years, whether in the context of U.S. drone strikes in northwest Pakistan and Yemen or the more recent raid on bin Laden’s compound in Abbottabad. These cases demonstrate a trend toward targeted killing away from the conventional battlefield of named individuals who, in many cases, are also wanted for serious crimes.

The bin Laden killing is, in some respects, an easier case than it might have been. In authorizing the operation, President Obama reportedly rejected Defense Secretary Robert Gates’s recommendation to deploy a missile strike to obliterate the suburban compound.149 Although doing so would have minimized the risks to U.S. military personnel, the President was reportedly uncomfortable with the risk to civilian lives and property that would have resulted from deploying bombs of the required intensity into the suburban area.150 Reports have also noted uncertainty regarding whether bin Laden was, in fact, residing in the compound, pointing out that a missile strike would not have allowed the same positive identification of the target in the way the Navy SEALs’ raid on the compound did.151

These considerations supply powerful moral and political support to the President’s choice. Whether they reveal a legal obligation to privilege a special-forces raid over less discriminate

149. See Schmidle, supra note 4, at 38 (providing a detailed narrative of the planning of and raid on Osama bin Laden’s Abbottabad compound).

150. See id. (noting that the amount of explosives necessary to penetrate bin Laden’s compound would have created “the equivalent of an earthquake” and “[t]he prospect of flattening a Pakistani city made [President] Obama pause”).

151. See id. at 40 (reporting that the confidence among CIA analysts that bin Laden was in the compound ranged from forty percent to ninety-five percent).
means under the circumstances raises a more difficult question about the limits of IHL. Unlike our hypothetical machine gunner, Osama bin Laden was a leadership figure of central importance, whose targeting was a priority of U.S. military efforts for the last decade. 152 Although we have argued that cases of targeted killing will frequently justify a stricter application of the proportionality test, the possibility of incidental civilian losses in the pursuit of important military objectives remains a core feature of IHL’s permissive approach to killing. 153 The general requirement to discriminate between military and civilian objects, moreover, has resisted the development of determinate standards dictating how much personal risk combatants must assume in order to minimize risk to civilians. 154 Although we are inclined to agree with those who argue that military planners should not assign greater value to their own combatants’ lives than to those of civilians, this is not an issue that is firmly resolved by existing legal standards. 155

It is tempting to argue that if the U.S. military would have been justified in conducting a missile attack on the bin Laden compound, then any less destructive method of killing the al Qaeda leader would have also been permissible. We should be glad, as some have argued, that the military did not simply obliterate the Abbottabad compound but instead subjected U.S. servicemen to substantial risk to protect civilian lives. 156 We should therefore avoid imposing additional legal requirements that might deter similar humanitarian gestures. There is some merit to this argument, but this line of thinking only goes so far.

152. See id. at 36 (stating that, after several previous failed attempts, the Abbottabad raid was the team’s “first serious attempt since late 2001 at killing [bin Laden]”).

153. See supra notes 126–33 and accompanying text (discussing the proportionality requirement with respect to risk to civilians).

154. See supra note 135 and accompanying text (pointing out vagueness in the proportionality requirement).

155. See Luban, supra note 135, at 12 (discussing the viability of the idea that “soldiers must place higher value on their own civilians than on themselves, but higher value on themselves than on enemy civilians”).

156. See Schmidle, supra note 4, at 38 (noting that an airstrike would have avoided the risk of “having American boots on the ground in Pakistan” and lessened the risk to American soldiers).
It is one thing to employ a means of attack, such as a remote missile strike, which, if permissible, permits neither perfect discrimination between combatants and civilians nor capture in lieu of killing. But the justification for employing such means, if legitimate, must be rooted in considerations of military advantage, including some degree of force protection, and not in a simple desire to kill rather than capture. The decision to neutralize bin Laden by means of a raid complicates the picture because, by electing to undertake a more discriminating and risky operation, the military opened up the possibility of capture in lieu of killing. The challenge, then, is to explain why the commandos should remain free of any duty whatsoever to capture short of bin Laden’s affirmative surrender.

This challenge is particularly acute in this case because the mechanics of the operation—a raid on a residential compound in an (at least nominally) allied state, and in an area that was not otherwise the site of ongoing hostilities or of a hostile military presence—resemble peacetime law enforcement so closely that it is difficult to explain why considerations of military necessity impose functional requirements distinct from those we associate with law enforcement.

This is not to argue that the operation was illegal. Clearly, the Navy SEALs who raided the Abbottabad compound faced great risks in pursuing bin Laden, a fact that distinguishes their situation from our hypothetical combatants who pursue Gunner on vacation. As we discuss below, it is quite possible, depending on the facts, that the use of lethal force was justified even if we assume that traditional law-enforcement principles applied.157 We do not possess sufficient knowledge of the rules of engagement or the facts on the ground to answer this question, but surely even law-enforcement officials undertaking a dangerous raid on a notorious terrorist’s well-guarded residential compound will generally be justified in acting under the assumption that they are under a constant threat, and the decision to use lethal force must be understood with this in mind.

These considerations support arguments that the killing of bin Laden was justified under either standard law-enforcement

157. *Infra* Part VI.
rules or under what we will describe below as a TK 2 case of self-defense or defense of others. They do not explain, however, why one should treat the operation as a TK 1 case, in which IHL supplies more permissive rules of engagement than the rules that would otherwise apply if one imagines, for example, that Pakistan's own police forces—acting on a tip from U.S. intelligence officials—conducted the operation themselves. In particular, these considerations do not support the blanket rule, suggested by Koh's public statements, that only affirmative surrender would have triggered an obligation to take bin Laden into custody.\footnote{See Koh, \textit{supra} note 14 (“[C]onsistent with the laws of armed conflict and U.S. military doctrine, the U.S. forces were prepared to capture bin Laden if he had surrendered in a way that they could safely accept. . . . But where that is not the case, those laws authorize use of lethal force . . . .”).}

Like our hypothetical of Gunner on vacation, it is hard to see how considerations of military necessity could have supported the killing of bin Laden if capture would have been feasible, if he was in fact manifestly defenseless, or if the circumstances were such that one would demand that law-enforcement officials attempt capture.

Additionally, although we hesitate to speculate on the accuracy or completeness of the picture provided by unofficial, often anonymously sourced accounts, some media reports provided a narrative that, if true, raises questions about whether the rules of engagement were in fact consistent with our reading of IHL. An August 2011 \textit{New Yorker} article quoted an anonymous special-operations officer, reportedly involved in the operation, as stating, “[t]here was never any question of detaining or capturing him—it wasn't a split-second decision. No one wanted detainees.”\footnote{Schmidle, \textit{supra} note 4, at 43.} The same article also described bin Laden's killing as having taken place after Navy SEALs had already killed every armed man in the compound, and immediately after one serviceman undertook great personal risk to clear a path to the al Qaeda leader by bear-hugging a woman feared to be armed with explosives.\footnote{\textit{Id.} at 42–43.} The killing was then succeeded by approximately thirty minutes of intelligence-gathering, during which time surviving residents of the
compound were bound and guarded.\textsuperscript{161} The special forces then departed, taking bin Laden's body with them.\textsuperscript{162} Although none of these facts are legally conclusive in our view (assuming that they provide an accurate picture), they do suggest that the decision to kill bin Laden is more difficult to explain by reference to potential sources of military advantage, such as a desire to avoid personal risk or an inability to take detainees for lack of resources or time.\textsuperscript{163}

Finally, we maintain that avoiding detention is not itself a justifiable reason to kill rather than capture under IHL. Although one can credibly speak of advantages of avoiding the burdens of detention and trial, and from precluding a charismatic terrorist leader from exploiting the publicity of legal process as a recruiting tool, these are not the sorts of benefits embraced by the concept of military advantage, which focuses on actions aimed at weakening the military forces of the enemy, and not on the realization of broader political goals.\textsuperscript{164} A policy to avoid trial on these grounds would also present irreconcilable tensions with the due process guarantees afforded by IHL to detainees accused of criminal offenses.\textsuperscript{165} These guarantees affirm that the avoidance of due process is not the type of military advantage that IHL generally privileges combatants to pursue.

\begin{enumerate}
\item \textsuperscript{161} See id. at 43 (describing the cuffing of surviving residents and the scouring of the compound).
\item \textsuperscript{162} See id. at 44 (describing how bin Laden's body was given Muslim rites and then heaved into the sea).
\item \textsuperscript{163} See id. at 43 (stating that bin Laden was found unarmed and could have been taken alive if he had immediately surrendered).
\item \textsuperscript{165} See POW Convention, \textit{supra} note 48, art. 3(d) (prohibiting “the passing of sentences and the carrying out of executions without previous judgment pronounced by a regularly constituted court affording all the judicial guarantees which are recognized as indispensable by civilized peoples”).
\end{enumerate}
III. The Lawfulness of Targeted Killings Beyond War

A. A Typology of Cases

Thus far, we have focused on so-called TK 1 cases, in which IHL is invoked to justify targeted killing. We now turn to cases that do not rely on IHL. One case—call it TK 2—involves a governmental actor instructed to kill an evildoer who is believed to be planning an attack that would endanger the lives of many innocent people. The government agent catches up to the evildoer seconds before he detonates a bomb that will destroy a residential building. The agent thwarts the attack by killing the evildoer. In a sense, this case involves a targeted killing, for the government named an individual beforehand as a target for killing. Nevertheless, the case is uncontroversial because the killing is clearly lawful under the domestic laws pertaining to defense of self and others.166 Because killing the evildoer was necessary to prevent an imminent attack against innocent people, the act is obviously justified.167 For this reason, cases like TK 2 do not trigger the sort of concerns that make the practice of targeted killing morally and legally problematic.

A more interesting case—call it TK 3—arises when the agent kills the evildoer before the attack commences. In some cases it could be argued that such a killing is necessary to thwart a future attack.168 However, the timing makes it unclear whether the act is justified under the conventional understanding of defense of self and others. Cases like TK 3 raise the problem of preemptive or anticipatory self-defense.169 In these cases, the killing is

166. See, e.g., Model Penal Code § 3.04(1) (1985) (stating that the use of force is justifiable if “immediately necessary” to protect against an attack).
167. Killings that thwart an imminent wrongful attack are clearly justified. It is unclear, however, whether killings that thwart a non-imminent future attack are justified. See Kimberly Kessler Ferzan, Defending Imminence: From Battered Women to Iraq, 46 Ariz. L. Rev. 213, 215 (2004) (discussing the challenges to stringent imminence requirements in the law).
168. See id. at 224 (discussing the “complex” issues raised by anticipatory self-defense in the international context).
169. See Amos N. Guiora, Anticipatory Self-Defence and International Law: A Re-Evaluation, 13 J. Conflict & Sec. L. 3, 5 (2008) (proposing that international law needs to be revised to allow for preemptive action against a non-state actor, provided that sufficient intelligence is present).
deemed necessary to prevent an attack that is neither imminent nor ongoing but is certain to take place in the future. 170 Even though these cases are difficult to justify under the traditional approach to the law of self-defense, 171 more modern formulations of the rules governing the use of force in protection of the person seem to allow enough leeway to justify such acts, at least in some circumstances. 172 As a result, TK 3 cases are more controversial than TK 2 cases, but the legality of such conduct can be justified fairly easily by effecting some modifications in the traditional law of self-defense. 173

The more difficult and controversial case—call it TK 4—arises when a named target, one who is believed to be dangerous but who is not a member of enemy armed forces, is killed while she is not engaging in an attack and when there is no knowledge of a specific future attack that the target is planning, which can only be prevented by use of lethal force. Arguably, this is the case most reflective of the bin Laden killing. 174 TK 4 cases are

170. See id. at 4 n.3 (positing that preemptive self-defense “allows for reaction when a serious threat to national security exists,” but in doing so it “expands the notion of imminence”).

171. These cases are difficult to justify under the traditional law of self-defense because the deadly force is not used to prevent an imminent attack. This is also the problem with providing a justification defense to battered women who kill their abusers in nonconfrontational settings. See Joshua Dressler, Battered Women and Sleeping Abusers: Some Reflections, 3 OHIO ST. J. CRIM. LAW 457, 461 (2006) (introducing the problem of the self-defense justification when the domestic abuse is not imminent).

172. The Model Penal Code’s self-defense provision authorizes killings that are “immediately necessary” to thwart a future attack even if the attack is not yet imminent. MODEL PENAL CODE § 3.04(1) (1985); see also Russell Christopher, Imminence in Justified Targeted Killing, in TARGETED KILLINGS: LAW AND MORALITY IN AN ASYMMETRICAL WORLD 253, 256 (Claire Finkelstein, Jens David Ohlin & Andrew Altman eds., 2012) (“Some argue that targeted killings of terrorists [representing continuous and ongoing threats of unlawful aggression] do satisfy the imminence requirement.”).

173. See Guiora, supra note 169, at 5 (proposing a strict-scrutiny test that would allow states to act in self-defense to prevent an attack).

174. As a result of the raid on bin Laden’s compound, U.S. officials reportedly gained evidence of the al Qaeda leader’s involvement in multiple terror plots, but reports do not indicate that officials viewed the raid as necessary to prevent a particular, known attack. See, e.g., Mark Mazzetti & Scott Shane, Data From Raid Shows Bin Laden Plotted Attacks, N.Y. TIMES, May 6, 2011, at A1 (discussing documents taken from the Abbottabad compound).
problematic because the target does not meet the status-based requirements of IHL, and, therefore, the killing cannot be justified by appealing to the permissive rules governing the targeting of combatants in wartime. These cases are also difficult to justify under conventionally accepted accounts of domestic criminal law defenses, given that it cannot be shown that killing the person was necessary to prevent an ongoing or future attack. As a result, TK 4 cases raise some of the most challenging questions about the justifiability of targeted killings.

In the remainder of this Part, we consider the legality of targeted killings under current understandings of the domestic criminal law defenses of self-defense, lesser evils, and law-enforcement authority. Our discussion focuses on U.S. law and precedents for illustrative purposes because each of these defenses is consistent in definition and scope across most common and civil law jurisdictions and finds support in international human rights jurisprudence. Accordingly, we do not suggest targeted killings taking place abroad are themselves subject to U.S. criminal law. Moreover, while we examine the legality of TK 1, TK 2, and TK 3 scenarios, our primary emphasis is on the potential justification of killing in TK 4 cases.

B. Targeted Killing as Traditional Defense of Self or Others

One way to argue for the legality of targeted killings is to invoke the traditional law of self-defense or defense of others. It is worth clarifying that we refer to the domestic law of self-defense rather than to the international rule that authorizes countries to resort to force in the exercise of so-called national self-defense. Pursuant to Article 51 of the United Nations Charter, states have

175. See supra Part II (discussing the treatment of targeted killings under the conventions of jus ad bellum).
176. TK 4 cases are difficult to justify under both traditional and expansive formulations of the law of defense of self and others. See Christopher, supra note 172, at 255 (stating that, because terrorists are not combatants, targeting them for killing cannot be justified under the existing self-defense paradigms).
177. See, e.g., Melzer, supra note 11, at 232 (summarizing the circumstances, based on a review of international legal sources, in which lethal force may be used).
a right to use force in order to thwart an armed attack against the nation.\textsuperscript{178} As we have already described, this right is the source of much debate in ways critical to the legality of a targeted-killing policy that crosses international borders. That debate, however, is largely outside the scope of this Article because it is principally concerned with breaches of state sovereignty, rather than the rules pertaining to killing itself. Not all cases, moreover, will implicate problems of state sovereignty. A state, for example, may avoid sovereignty concerns by inviting a foreign government onto its territory to carry out targeted killings, but that consent does not make the killings themselves legal.\textsuperscript{179}

Domestic self-defense can sometimes succeed when the international law of self-defense fails. It is plausible to imagine a targeted killing that can be justified under the domestic law of self-defense but cannot be justified under the international rules governing the use of force. Imagine, for example, an intelligence officer who is dispatched overseas to monitor a suspected terrorist. Suppose that the officer kills the terrorist right before the terrorist is to detonate a bomb inside a restaurant that will kill several innocent people. Putting aside whether the officer is otherwise legally present in the state, his actions do not implicate the international rules governing the use of force. The killing of a single individual in this manner does not rise to the level of an “armed attack” under Article 51.\textsuperscript{180} At most, he might be guilty of murder, but he is not in this case because the killing is easily

\textsuperscript{178}. See U.N. Charter art. 51 (“Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations . . . .”).

\textsuperscript{179}. See, e.g., Van Schaack, supra note 2, at 268 (“While Yemen can consent to another state entering its territory, . . . it cannot consent to that state violating IHL or human rights law while there. Thus, some lawful justification for the use of deadly force must still be identified.”).

justified under the domestic law of individual self-defense or defense of others, which allows a person to kill an aggressor in defense of self or others if he reasonably believes that killing the aggressor is necessary in order to repel an imminent and unlawful threat of death or serious bodily injury.\textsuperscript{181} The killing of the terrorist thus amounts to a justifiable act of individual defense of self or others as long as it was reasonable to believe that the act was necessary to save the lives of third parties, which it surely was.

The law of domestic self-defense is therefore well equipped to deal with cases such as TK 2, in which an evildoer is killed just before he engages in conduct that would endanger the lives of others. Nevertheless, as we have already described, the domestic law of self-defense cannot be invoked to justify targeted killings in TK 3 cases, in which an evildoer is killed in order to prevent a future but non-imminent attack—at least not as this body of law has been understood traditionally. The conventionally accepted account of the domestic law of self-defense authorizes killings only when this course of action is believed to be necessary to repel an unlawful and \textit{imminent} threat.\textsuperscript{182} Therefore, assuming that the evildoer is poised to engage in an attack that will take place sometime in the next few days, weeks, or months, rather than shortly after he is killed, his killing cannot be justified under the traditional law of domestic self-defense. Such cases fall within what criminal theorists call “preventive” self-defense, which, although hotly debated, is almost universally recognized to fall outside of the core of traditional justifiable self-defensive action.\textsuperscript{183}

\textsuperscript{181} \textit{See, e.g.}, N.Y. PENAL LAW § 35.15 (McKinney 2004) (stating that a person may “use physical force upon another person when and to the extent he or she reasonably believes such to be necessary to defend himself”); People \textit{v.} Goetz, 497 N.E.2d 41, 47 (N.Y. 1986) (recognizing that self-defense permits the use of deadly force when the actor reasonably believes use of such force is necessary to repel an imminent attack).

\textsuperscript{182} \textit{See} FLETCHER & OHLIN, supra note 64, at 90–91 (discussing how an imminence requirement helps to distinguish self-defense from unlawful preemptive attacks and punitive reprisals).

\textsuperscript{183} \textit{Id.} at 162 (noting that preventative self-defense is analogous to describing war as “politics by other means”).
If domestic self-defense cannot legally justify killing in TK 3 cases, it *a fortiori* cannot justify killing in TK 4 cases, in which an evildoer who does not pose an imminent threat is killed despite a lack of specific knowledge of planned future attacks that would be thwarted by the killing. The nonexistence of an imminent threat is, once again, fatal to any purported claim of justification under the domestic law of self-defense in these types of cases. As a result, the laws of individual self-defense are useful only in justifying targeted killings in easy cases, in which no one but the most radical pacifist would oppose the use of force. The laws of individual self-defense fail to justify targeted killings in the more complicated TK 3 and TK 4 scenarios.

**C. Targeted Killing as Expanded Defense of Self or Others**

If the most controversial aspects of the practice of targeted killings cannot be legally justified by appealing to the traditional formulation of the law of self-defense, perhaps a more modern and expanded version of self-defense can better justify these acts. Common reformulations start by arguing that the concurrence of an “imminent attack” should not be considered amongst the conditions that trigger the justifiable use of deadly force against another.184

Why do some scholars argue in favor of abandoning the imminence requirement? One reason is that there are situations in which it seems the only way of repelling a future deadly attack is to use force in a nonconfrontational situation before the attack becomes imminent.185 These cases present a conflict between necessity and imminence.186 For Jane to save her own life, it might be necessary to kill Bill when he is not attacking her. If she

---

184. *See, e.g.,* Christopher, *supra* note 172, at 255 (asserting that the imminence requirement, which “bars targeted killing of terrorists, should be rejected”).

185. *See, e.g.,* Fletcher & Ohlin, *supra* note 64, at 163 (“Suppose a terrorist threatens to implant an undetectable nuclear device that is set to explode in a year. He can be stopped now, but once the device is implanted, it will be too late.”).

waits for an attack, she stands no chance. If the necessity of using force is deemed to be more important than the concurrence of an imminent attack, then Jane's preemptive strike ought to be justified. If, on the contrary, imminence trumps necessity, then Jane's anticipatory use of force should not be justified.

The debate is of significant practical importance in battered women cases, in which wives defend the killing of their husbands in nonconfrontational situations, such as when their husbands are asleep, on the ground that waiting until the attack became imminent would have prevented the wives from successfully curbing the future attack that the husbands would have surely launched. Normally, of course, one would demand that the victim call the authorities to prevent a non-imminent attack, but in some cases the battered spouse has reached out to authorities to no avail, indicating that authorities are unwilling or unable to help her.

There are many who argue that in such cases the traditional law of self-defense should give way, the imminence requirement should be relaxed or abandoned, and the emphasis should be placed on whether or not the use of force is necessary to prevent the future attack. There are also many staunch supporters of the imminence requirement who oppose such an expansion of self-defense law. Regardless of which of these views is more

187. This is the issue presented in the well-known case of State v. Norman. See State v. Norman, 378 S.E.2d 8, 9 (N.C. 1989) (concluding that a woman who had been physically and mentally abused by her husband over period of several years, and who had been diagnosed as suffering from battered woman's syndrome, was not acting in self-defense when, in order to prevent future abuse, she shot her husband while he slept). Much ink has been spilled analyzing the issues raised by Norman. See, e.g., Dressler, supra note 171, at 458 (suggesting that expanding self-defense to include nonconfrontational homicide would be a mistake); Whitley R.P. Kaufman, Self-Defense, Imminence, and the Battered Woman, 10 NEW CRIM. L. REV. 342, 343 (2007) (arguing that the imminence rule is an essential element of the law of self-defense and should not be modified).

188. See Norman, 378 S.E.2d at 19 (stating that a police officer told the wife that “he could do nothing for her unless she took out a warrant on her husband” and left, even after the wife said that “if she [took out a warrant], her husband would kill her”).

189. See, e.g., Rosen, supra note 186, at 380 (“If action is really necessary to avert a threatened harm, society should allow the action, or at least not punish it, even if the harm is not imminent.”).

190. See, e.g., Ferzan, supra note 167, at 217 (positing that “imminence serves as the actus reus for aggression, separating those threats that we may
normatively appealing, it is worth noting that the more modern and expansive account of self-defense is slowly but surely gaining supporters. The Model Penal Code’s formulation of the defense rejects the imminent-attack limitation on the use of force in defense of self or others, and instead includes the more lenient and necessity-based requirement that the force be immediately necessary to repel an unlawful aggression.\textsuperscript{191} Assuming, for the sake of argument, that this position is correct, can it serve to justify targeted killings in the more interesting and complicated cases discussed here?

For starters, an expanded version of self-defense that does away with the imminence requirement fares better in justifying TK 3 cases than the traditional imminence-based account of self-defense. In such scenarios, much like in certain battered-women cases,\textsuperscript{192} deadly force is used to prevent a near certain future attack, and waiting until the future aggression becomes imminent will significantly reduce or even eliminate the defender’s chance of successfully warding off the aggression. As long as the use of such force is the only way of thwarting the future attack, the absence of imminence does not preclude a finding of legal justification under the more expansive necessity-based account of self-defense.\textsuperscript{193} This is precisely the argument recently put forth by Russell Christopher as a possible justification for targeted killings.\textsuperscript{194} Building primarily on what

\textsuperscript{191.} See \textit{Model Penal Code} § 3.04(2)(b) (1985) (“The use of deadly force is not justifiable under this Section unless the actor believes that such force is necessary to protect himself against death, serious bodily injury, kidnapping or sexual intercourse compelled by force or threat . . . .”).

\textsuperscript{192.} See, \textit{e.g.}, \textit{Norman}, 378 S.E.2d at 9 (concluding that a woman who suffered from battered woman’s syndrome was not acting in self-defense when she shot her husband while he slept in order to prevent future abuse); see also \textit{Dressler}, supra note 171, at 459–61 (detailing the constant verbal and physical abuse that Judy Norman endured from her husband and discussing whether Mrs. Norman was justified in killing her husband despite the lack of an imminent attack).

\textsuperscript{193.} See \textit{Rosen}, supra note 186, at 404–07 (discussing the viability of a self-defense statute without an imminence requirement).

\textsuperscript{194.} See supra note 172 and accompanying text (discussing how the imminence requirement is becoming easier to satisfy).
others before him have argued in the context of battered-women cases, Christopher argues that the absence of an imminent attack need not be fatal to the claim that targeted killings ought to be justified under the rubric of self-defense.\textsuperscript{195}

What Christopher does not do, however, is explain whether this expanded version of self-defense can account for the legality of TK 4 cases. Unfortunately for those who believe that targeted killing can be justified even in some TK 4 cases, the paradigm of expanded self-defense is inadequate. The problem is that TK 4 cases involve killing an individual who is believed to be dangerous, without any degree of certainty about whether doing so is necessary to prevent a future attack. Although in such cases the agent has knowledge that the evildoer is dangerous and that, in light of his past acts, he will likely attack again in the future, there is no concrete evidence of specific future attacks that can only be prevented by killing him. An expanded version of self-defense would thus fail to justify such killings because the essential element of the defense—that the use of force be necessary to prevent the death or serious bodily injury of third parties—is missing.

\textit{D. Targeted Killing as Law-Enforcement Action}

Claire Finkelstein has suggested that some targeted killings might be legally justified by appealing to the rules governing the use of deadly force pursuant to law-enforcement authority.\textsuperscript{196} In particular, Finkelstein contends that law-enforcement authority can justify targeted killings that are not justified by the law of defense of self or others.\textsuperscript{197} Can law-enforcement authority justify

\textsuperscript{195} See Christopher, supra note 172, at 284 (concluding that “[t]he imminence requirement for justified self-defense is problematic and should be abandoned”).

\textsuperscript{196} See Claire Finkelstein, \textit{Targeted Killing as Preemptive Action, in} TARGETED KILLINGS: LAW AND MORALITY IN AN ASYMMETRICAL WORLD 156, 178–79 (Claire Finkelstein, Jens David Ohlin & Andrew Altman eds., 2012) (arguing that “there is justification for the use of force . . . as applied to a targeted killing situation” and that the justification is “most clearly demonstrated by certain domestic law enforcement circumstances”).

\textsuperscript{197} See id. at 179 (considering the reach of law-enforcement authority in cases in which the standard justifications are inapplicable).
these more controversial cases? It depends on the scope of the rules governing law-enforcement authority, an issue that Finkelstein unfortunately does not address in much detail.

U.S. legal history demonstrates the complexity of the issue. At early common law, deadly force could be used by law-enforcement agents to arrest a fleeing felon. The right to use such force was quite expansive because its use was not conditioned on whether the fleeing felon posed a danger to the officer or to third parties. Deadly force could also be used to prevent the commission of a felony. This expansive view of law-enforcement authority subsequently eroded in many states, either by statute or case law. Presently, many, if not most, states limit the use of deadly force by government officials to the prevention of violent felonies. The Model Penal Code adopted a similar limit on the use of deadly force by law-enforcement officers, which is authorized pursuant to Section 3.07 if the crime for which the arrest is made involved the use or threatened use of deadly force.

Although the common law rules are fairly clear and easy to apply, the U.S. Supreme Court complicated matters when it held that the Fourth Amendment right against unreasonable seizures placed constitutional limits on the amount of force that a police officer may use to make an arrest. In *Tennessee v.*

---


199. See id. (noting that the common law crime-prevention defense “is remarkably broad in that it authorizes use of necessary deadly force to prevent nonviolent felonies”).

200. See id. (explaining that deadly force at common law is justified if the user of force reasonably believes such other person is committing a felony).

201. Id.; see also, e.g., People v. Ceballos, 526 P.2d 241, 245 (Cal. 1974) (adopting the common law rule that the use of deadly force to prevent a felony is justifiable only if the offense is a “forcible and atrocious crime”).

202. Dressler, supra note 198.

203. See Model Penal Code § 3.07(2)(b) (1981) (explaining that deadly force is not justified unless the actor believes “the crime for which the arrest is made involve[s] conduct including the use or threatened use of deadly force”). There are additional requirements that must be met under § 3.07 before deadly force by law enforcement is authorized. See id. § 3.07 (listing the requisite elements that justify use of deadly force).

204. See Tennessee v. Garner, 471 U.S. 1, 7–8 (1985) (rejecting the argument that, as long as there is probable cause to warrant an arrest, the
Garner, the Court held that the killing of a fleeing unarmed youth suspected of committing a nonviolent felony amounted to an unreasonable seizure that ran afoul of the Fourth Amendment. Subsequently, the Court held in Scott v. Harris that the police could use deadly force by ramming a fleeing driver's car from behind because the driver's reckless conduct could have jeopardized the life or limb of other motorists. It is unclear whether these cases impose additional limits on the use of deadly force pursuant to law-enforcement authority, beyond the ones statutorily in place in most states today.

Regardless of whether the U.S. Constitution imposes additional limitations on the use of deadly force by governmental agents, it is fairly obvious that police officers may sometimes use more force than private citizens would be allowed to employ pursuant to the law of self-defense and defense of others. The facts of Harris constitute a case in point. The police officers in Harris were allowed to use deadly force against the fleeing motorist even though the motorist was not endangering the life or limb of third parties at the time force was used. While the driver's conduct was reckless and dangerous in the abstract (it could have endangered someone had someone been there), there was no evidence that the life or bodily integrity of another motorist was in imminent danger at the moment the police

Fourth Amendment has no role in determining how an arrest is made).

205. See id. at 22 (holding that an officer's use of lethal force without adequate physical threat of harm was unconstitutional).

206. See id. at 20–21 (stating that the officer did not have probable cause to believe the suspect posed any physical danger).

207. See Scott v. Harris, 550 U.S. 372, 386 (2007) (“A police officer's attempt to terminate a dangerous high-speed car chase that threatens the lives of innocent bystanders does not violate the Fourth Amendment, even when it places the fleeing motorist at risk of serious injury or death.”).

208. See id. at 383–84 (balancing the extent of the intrusion on the defendant's Fourth Amendment interests against the government's interests and finding that the latter prevails because of the number of innocent lives at risk).


210. See id. at 384–86 (finding that the Fourth Amendment did not prevent the police officers from using deadly force against a fleeing motorist who had not definitively placed any lives in imminent danger).
rammed the driver’s car.211 The reasonable possibility of future harm that might be caused by the motorist seemed to be sufficient to justify using deadly force. Such reasonable probability of future harm would not be enough to trigger the imminence requirement in self-defense.

Applied to targeted killings, the rules governing the use of deadly force by law-enforcement officers could authorize using deadly force against the evildoer if he is fleeing. In many targeted-killing cases, the person targeted is suspected of criminal responsibility for grave atrocities, which would satisfy the violent-felony requirement, assuming the accusations can be substantiated. The problem, however, is that law-enforcement authority justifies the use of deadly force only to prevent a felon from escaping or to thwart the imminent commission of a violent felony.212 In TK 4 cases (as well as in many TK 3 cases), however, the person targeted is not fleeing and might not even be aware that the agent is targeting him.213 Also, the targeted individual in these cases is not about to commit a crime when he is targeted. Therefore, the version of law-enforcement authority that has prevailed in domestic law cannot justify using force in the more complicated and controversial TK 4 cases unless the targeted suspect is fleeing or about to commit a crime.

E. Targeted Killings as Necessity or Lesser Evils

Perhaps some targeted killings can be justified as cases of necessity or lesser evils. Under the Model Penal Code, an actor is entitled to a lesser-evils or necessity justification if she inflicts an evil in order to prevent an even greater evil.214 Thus, the person

211. See id. at 384 (acknowledging that there was a less than certain probability that the defendant’s actions would cause death).
212. See DRESSLER, supra note 198, § 21.03 (explaining the circumstances in which deadly force is permissible under the modern majority rule).
213. This would almost always be the case when the targeted killing is carried out by way of an aerial strike. For obvious reasons, most targeted individuals will not be aware of the fact that they are about to be killed by a missile or a drone strike.
214. See MODEL PENAL CODE § 3.02(1)(a) (1985) (providing that conduct one believes to be necessary to avoid harm to oneself or another is justifiable if “the harm or evil sought to be avoided by such conduct is greater than that sought to
who breaks a car window without the owner’s consent in order to save a child who is suffocating inside the car acts justifiably pursuant to the lesser-evil defense. Although her conduct nominally satisfies the elements of the offense of criminal mischief, it is not considered wrongful because the evil prevented by the act (the death of a child) is greater than the evil inflicted (damage to the property). Similarly, it could be argued that killing the targeted individual may sometimes be justified because doing so prevents the occurrence of some greater harm. Perhaps, for example, by killing the target, one can foil the target’s plot to bomb a residential building. Doing so would save dozens or hundreds of people by killing one. Although it is controversial whether one can ever kill a person in order to save other people in circumstances other than self-defense or law-enforcement authority, many have argued that such killings ought to be justified if the amount of people saved by the conduct is considerable.

Fernando Tesón has recently defended some targeted killings by appealing to a logic similar to the one undergirding the lesser-evils justification. More specifically, he argues that it might sometimes be justified to engage in the targeted killing of an evil
ruler or a terrorist if doing so prevents him from killing hundreds or thousands of people. It is important to note that Tesón’s proposal justifies the targeted killing only if engaging in such tactics is the only way to prevent the en masse killing of human beings. According to Tesón, this means that, in the sui generis context of terrorism, “the state reasonably knows that acting now may be its only chance to avert a terrorist strike.”

Tesón’s proposal illustrates how and when the lesser-evils defense can be invoked to justify targeted killings. Although Tesón’s argument is plausible, it has at least two limitations for purposes of our analysis. First, and taking into account that Tesón does not impose an imminence requirement, it is unclear whether his justification for targeted killings is any different from the expanded defense of self or others justification that we have already described. If killing the target now is the only way to thwart a future threat to innocent lives, it would seem that engaging in such conduct would also be justified under the expanded concept of self-defense. Perhaps the most important contribution of this proposal, therefore, is its novel approach to the international-relations dimension of targeted killing: the ability to save a substantial number of lives may explain why a particular state may violate the territorial integrity of another state to carry out the killing.

Second, and more importantly, Tesón’s proposal cannot justify killings in TK 4 scenarios, for it is the nature of such cases that it is not known with any certainty whether the targeted individual was about to launch an attack. Although it is reasonable to believe that the individuals targeted in cases like TK 4 could help plan or launch future attacks, there is no concrete evidence about specific plans that can only be thwarted by killing the person. This is why it is difficult, if not impossible, to invoke traditional lesser-evils principles to justify targeted killings in TK 4 scenarios. Given that the government official in

218. See id. at 423 (proposing that targeted killing of terrorists in a peacetime setting is permissible “only when necessary to prevent the deaths of a substantial number of innocent persons”).
219. See id. (explaining that it must be impossible or prohibitive to capture the target for a targeted killing to be justified).
220. Id. at 427.
such cases is unaware of any specific future (and reasonably imminent) attacks that would be foiled by killing the target, it cannot be said with any reasonable degree of certainty that killing the individual is indeed the only way to prevent the killing of innocent human beings in the future.

IV. Another Paradigm for Targeted Killings? The Killing of Pablo Escobar

Pablo Escobar was one of the most infamous criminals of the late twentieth century. Escobar was the head of the “Cartel de Medellín,” Colombia’s largest and most feared criminal organization. At its zenith, the Cartel de Medellín was one of the most successful and sprawling criminal organizations of all time, smuggling tons of cocaine into the United States on a weekly basis.

Escobar kept his grip on power by employing a double strategy. First, he would create a loyal following amongst many people living in dire poverty, especially in his hometown of Medellín and, more generally, in the province of Antioquia. He did this by building housing complexes for the poor, helping the needy, and performing other charitable acts, such as building soccer fields with lighting so that workers could play at night. Although Escobar tried to get these sectors of the population to love him, he simultaneously attempted to get governmental support.

221. See Mark Bowden, Killing Pablo: The Hunt for the World’s Greatest Outlaw 59 (2001) (noting that, by 1989, Escobar was “one of the richest men in the world, and perhaps its most infamous criminal”). At least ten books have been written about Pablo Escobar, including books written by his accountant, some of his lovers, and some of his coconspirators. See, e.g., Roberto Escobar & David Fisher, The Accountant’s Story: Inside the Violent World of the Medellín Cartel, at inside cover (2009) (stating that Pablo Escobar’s brother, Roberto Escobar, is the author of the book and was the top accountant for the Medellín Cartel).

222. At one time, Escobar’s Cartel “dominated the cocaine traffic through Colombia to the United States.” Death of a Cocaine King, BALTIMORE SUN, Dec. 4, 1993, at 8A.

223. Escobar & Fisher, supra note 221, at viii.

224. See Bowden, supra note 221, at 28–29 (noting that Escobar was one of Medellín’s largest employers and spent millions of dollars improving the city).

225. Id. at 29.
actors to fear him. He did this by implementing, with ruthless efficiency, his infamous policy colloquially known as “plata o plomo.” Translated literally, the phrase means “silver or lead,” which meant that one would either have to accept bribes to allow Escobar to continue his business or face the bullet fire that would ensue if one refused to do so. The policy was a huge success for Escobar, as he was able to bribe most government officials who could have gotten in his way. Those who could not be bribed were often killed. There is even evidence that he managed to kill a popular presidential candidate who had publicly vowed to go after Escobar with all of the government’s might. When such strategies failed to work, it is believed that Escobar would resort to even more sinister means, including terroristic strategies such as suicide and commercial-aircraft bombings. Some have even suggested that Escobar was involved in the murder of nearly half of the justices of the Colombian Supreme Court, carried out by guerillas from the 19th of April Movement.

226. See id. at 24–29 (describing citizens’ love for Escobar and government officials’ fear of him).
227. Id. at 24.
228. See id. (providing the English translation of the phrase).
229. See id. (“One either accepted [Escobar’s] plata (silver) or his plomo (lead).”).
230. See, e.g., id. at 52 (detailing Escobar’s bribes of the Columbia Attorney General and judiciary).
231. At least 228 judges and court officers were killed during the 1980s. Steven Gutkin, Colombia Trying to Repair Judicial System Battered by Organized Crime, TIMES-NEWS (Hendersonville, N.C.), Jan. 10, 1991, at 25.
233. See Bowden, supra note 221, at 80–81 (describing an airline bombing plot ordered by Escobar). Escobar was implicated in the downing of an Avianca airliner in an unsuccessful effort to kill then-presidential-hopeful César Gaviria. Id. at 59.
234. See id. at 52–53 (discussing Escobar’s targeting of Colombia’s judicial system in the 1980s). The 19th of April Movement—known in Colombia as the M-19—was a Colombian guerrilla group that operated in the 1970s and 1980s. STEPHEN E. ATKINS, ENCYCLOPEDIA OF MODERN WORLDWIDE EXTREMISTS AND EXTREMIST GROUPS 185 (2004).
Eventually, Escobar decided to strike a deal with Colombian authorities in order to avoid extradition to the United States, where he was wanted for various drug-related offenses.235 The terms of the deal were quite favorable to Escobar. He would surrender only if the government agreed to give him a lenient sentence and to not extradite him.236 Escobar also demanded that he serve his prison time in a facility that he himself would build for the purposes of serving his sentence.237 He was eventually sentenced to five years of imprisonment and was allowed to serve his sentence in the facility he built, which came to be known as “La Catedral” (meaning “The Cathedral”).238 More Ritz Carlton than San Quentin, La Catedral was essentially a mansion equipped with a discotheque, gym, “dirt bike track,” and several “chalets” for entertaining his female friends.239 His pseudo-imprisonment did not stop his drug dealings, as he was able to surround himself with “prison guards” who were loyal to him and allowed him to meet with clients and continue managing his criminal enterprise.240

A little over a year after his sentence, Escobar murdered a pair of subordinates whom he believed to be disloyal to him.241 This was the impetus for then-President of Colombia César Gaviria to order the transfer of Escobar to a real correctional facility.242 Unfortunately for the government, Escobar was able to

236. See BOWDEN, supra note 221, at 96–97 (describing Escobar’s deal with the Colombian government to avoid extradition to the United States).
237. See Prison Prepares to Welcome Drug Billionaire, TORONTO STAR, June 16, 1991, at H3 (noting that Escobar “could equip the room as he wished,” which would be “no problem” considering his “personal fortune of $3 billion”).
238. See id. (referring to the location from which the article was reported as “La Catedral, Columbia”).
239. DOMINIC STREATFEILD, COCAINE: AN UNAUTHORIZED BIOGRAPHY 444 (2001).
240. See BOWDEN, supra note 221, at 110 (“The prison guards were no more than Pablo’s employees . . . .”)
241. See id. at 118 (describing how Escobar had two of his “powerful lieutenants” killed because he was “[s]uspicious . . . of their loyalty”)
242. See id. at 120 (describing President Gaviria’s decision, after the killings, to move Escobar “to a real prison”).
learn of the plan beforehand.\textsuperscript{243} Although the prison was surrounded by “an entire brigade of the Colombian army, roughly four hundred men,” Escobar managed to escape unharmed.\textsuperscript{244}

\textit{A. Killing Escobar}

Escobar’s escape from La Catedral was the last straw for the Colombian government. With full knowledge that capturing Escobar would require more training and resources than those that Colombia could offer at the time, the government turned to the United States for help.\textsuperscript{245} The same year that Escobar escaped from his prison hotel, agents from the United States’ elite combat team, Delta Force, started training and advising the Colombian police and military.\textsuperscript{246} This led to the creation of the “Bloque de Búsqueda” (Search Bloc),\textsuperscript{247} a Colombian police force trained and assisted by Delta Force that was tasked with finding and capturing Escobar.\textsuperscript{248} Commanded by Colonel Hugo Martínez, members of the Search Bloc were meticulously screened to ferret out any prospective agent that might be susceptible to corruption, or who might already be working as a spy for Escobar.\textsuperscript{249} The Search Bloc was also assisted by a U.S. Army special unit known as Centra Spike\textsuperscript{250} and American planes, which provided vital

\textsuperscript{243}. See \textit{id.} at 123 (“\textit{B}ecause of radio and TV reports, Escobar . . . knew that armed forces were massing around his prison. Any hope of surprise was gone.”).

\textsuperscript{244}. \textit{Id.} at 134.

\textsuperscript{245}. See \textit{id.} at 140 (noting that, in seeking help from the United States, Gavaria had “opened the door to \textit{anything}”).

\textsuperscript{246}. See \textit{id.} at 141–42 (describing U.S. officials’ deliberations, and ultimate decision, to use Delta Force to track down Escobar in Colombia).

\textsuperscript{247}. \textit{Id.} at 66.


\textsuperscript{249}. See \textit{Bowden, supra} note 221, at 67 (noting that officers on the Search Bloc could not be “native Antioquian[s]” from Escobar’s home town and that the men chosen were “considered elite and incorruptible”).

\textsuperscript{250}. See \textit{id.} at 73 (describing Centra Spike and its purpose); Mark Bowden, \textit{Quietly, Search Bloc Pins Escobar Down}, \textit{PHILA. INQUIRER}, Dec. 14, 2000, at A2 (noting that “American surveillance experts at Centra Spike” worked with the Search Bloc to locate Escobar).
intelligence that proved decisive in tracking down Escobar. The United States agreed to provide all of this help pursuant to President Reagan’s 1986 order that declared drug trafficking to be a threat to the national security of the country, and thus authorized the use of military assets to assist police forces in neutralizing the threat.

While the Search Bloc was hunting Escobar down, a vigilante group—financed both by Escobar’s enemies in the drug world and by right-wing paramilitaries—began a bloody campaign designed to weaken Escobar by killing his closest associates and relatives and to exact vengeance for what he had done to them and the country. The group called themselves “Los Pepes,” which stood for “Los Perseguidos por Pablo Escobar” (People Persecuted by Pablo Escobar). Los Pepes were almost as ruthless as Escobar, often employing the same tactics that Escobar used to terrorize his enemies. The group would achieve its objective by any means necessary, including planting bombs, setting fire to Escobar’s property, and murdering many of his lawyers, bankers, and extended family. There is also much evidence suggesting that there was some collaboration between the Search Bloc and Los Pepes and, indirectly, between the American military and Los Pepes. While there is no direct evidence linking the Search

---

251. See Bowden, supra note 221, at 154 (noting the volume of U.S. military aircraft monitoring Escobar’s activity over Medellín).

252. See id. at 54 (noting that National Security Decision Directive 221, signed in 1986, “opened the door to direct [U.S.] military involvement in the war on drugs”).


254. Bowden, supra note 221, at 176.

255. See id. (“If [Escobar] stood atop an organizational mountain that consisted of family, bankers, sicarios, and lawyers, then perhaps the only way to get him was to take down the mountain.”).

256. See id. at 191–95 (listing Escobar’s associates that were targets in the “bloodbath”).

Bloc and the American military to the perpetration of the crimes committed by Los Pepes, some have suggested that both the United States and the Colombian government tolerated the actions carried out by Los Pepes, albeit never expressly approving of them.258

Slowly but surely, the combined pressure exerted by the Search Bloc and Los Pepes wore Escobar down. By 1993, he was constantly on the run and feared not only that the Search Bloc would catch up to him, but also that Los Pepes would kill his wife and children.259 On December 2, 1993, the Search Bloc located Escobar using radio triangulation technology provided by the United States.260 After the Search Bloc stormed the Medellín house in which Escobar was hiding, Escobar opened fire.261 Realizing that he was outnumbered and outgunned, Escobar jumped out of a window and started running across the roofs of surrounding houses in a desperate attempt to escape.262 Members of the Search Bloc shot and killed Escobar on the spot.263 He was shot three times: once in the leg, once in the torso, and a fatal shot in his ear.264 Although the exact circumstances of his death are unknown, it is widely believed that Escobar was executed by the police, as there is evidence tending to demonstrate that the fatal shot was fired after Escobar was already neutralized by the Search Bloc.265

---

258. See id. (noting that “[i]f Los Pepes were working with the Search Bloc, that would explain their apparent access to fresh U.S. intelligence” but that any evidence of “American linkage with Los Pepes remained circumstantial” at best).
259. See Escobar on Run from Vigilante Attacks, TOLEDO BLADE, Feb. 28, 1993, at D5 (noting that Escobar was “on the run and getting what he dishes out” from Los Pepes, who recently bombed the homes of his “family members”).
260. See BOWDEN, supra note 221, at 239–43 (describing the details of locating Escobar with radio triangulation).
262. BOWDEN, supra note 221, at 248.
263. Id.
264. Id. at 253.
265. See id. (opining that “[t]he shots to his leg and back most likely would have knocked him down, but probably would not have killed him,” thus making it more likely that Escobar was “shot in the head after [he] fell”).
B. Domestic and International Reaction to Escobar’s Killing

Most Colombians received the news of Escobar’s death with jubilation.266 His demise marked an important milestone for the country. The government was finally able to triumph over the “biggest and baddest” drug lord the world had ever seen. Stifling the Medellín Cartel was a first win in a series of victories over the country’s cartels and thugs. The government focused on dismantling Escobar’s rivals, especially the Cali Cartel, and Colombia embarked on its slow but steady journey to regain control over the country.267 First would come the drug cartels and then the infamous FARC (Fuerzas Armadas Revolucionarias Colombianas) guerrillas.268 Although we do not wish to exaggerate the significance of this single event, Escobar’s death played an important role in Colombia’s resolution of its security problems. Today, the country is characterized by optimism in the fight against drug-smuggling and guerrillas. There is also cautious optimism about the economic growth that goes hand-in-hand with the political stability that such victories bring. In recent years, Colombia has been doing well economically and politically by Latin-American standards.269 Drug- and guerilla-related violence is down dramatically from 1990 levels.270 All in all, Colombia is

266. See id. at 261 (“[T]his day was for celebration.”); see also, e.g., Al Fin Cayó Escobar!, El Tiempo, Dec. 31, 1993, http://www.eltiempo.com/archivo/documento/MAM-282510 (last visited Sept. 24, 2012) (breaking news that Escobar was killed) (on file with the Washington and Lee Law Review).

267. See Bowden, supra note 221, at 272 (discussing the unraveling of the Cali cartel).


270. See Enrique S. Pumar, Colombian Immigrants, in Multicultural America: An Encyclopedia of the Newest Americans 353, 371–72 (Ronald H. Bayor ed., 2011) (finding that the Colombian government has become regarded
now a relatively safe country. Escobar’s death was an important component of this recovery.

This is not to say, of course, that Colombia has solved all of its problems. The guerrillas retain control over a small portion of the country, and drug violence is still a problem. Nonetheless, Colombia is undeniably a safer, less violent country than it was twenty years ago. This is due, at least in part, to the government’s dismantling of the Medellín cartel, including its killing of Escobar.

Perhaps the most important consequence of Escobar’s death was that it proved to the government and to Colombians that they were capable of defeating serious threats to their national security. The Search Bloc was so effective in tracking down Escobar that it was used as a model for other operations carried out by the government. Subsequent Search Blocs were successful in dismantling most of the remaining drug cartels in the country. The victory over Escobar also proved to be determinative of Colombian–U.S. relationships. The strong ties that developed between the two countries continue to thrive to this day. The United States has also learned from Colombia’s struggle against terrorism and organized crime. After all—as

---

271. See id. at 372 (describing the reduced levels of violence in Colombia).


others have pointed out—the Colombian experience helped the United States formulate its post-9/11 counterterrorism policies.275

In spite of the widespread jubilation with which the news of Escobar’s death was received in the country, some Colombians were saddened when they learned about his killing.276 For some, Escobar was a modern day Robin Hood who took from the government and the corrupt rich people and gave back to the poor. This was especially the case in some barrios277 of Escobar’s hometown of Medellín.278 These were the same people who knew of his whereabouts and never said a word to local authorities: the silent accomplices of Escobar’s effort to hide from the Search Bloc.

Most, however, were ready to turn the page and explore what a world without Escobar would mean for the country. Of special significance is the fact that Escobar’s death failed to attract vocal criticism from defenders of civil liberties and individual rights, either in Colombia or abroad. This is a telling fact. After all, at that time, it was widely rumored that Escobar was killed execution-style after he had been incapacitated by shots to the leg and torso.279 If so, the fatal shot would have clearly been unlawful under domestic and international law. Law-enforcement agents can only use deadly force when it is the only way to prevent a fleeing felon from escaping.280 But if—as many believe—Escobar was already shot down before the fatal shot was fired, the third shot was not necessary to prevent his escape.281 The alleged

275. See, e.g., Michael Kenney, From Pablo to Osama: Counter-Terrorism Lessons from the War on Drugs, 45 SURVIVAL 187, 187 (2003) (“Indeed, members of the US intelligence community acknowledge that drug enforcement raids in Colombia during the 1990s serve as models for today’s counter-terror operations in Afghanistan . . . .”).

276. See BOWDEN, supra note 221, at 266 (describing Escobar’s funeral as “an occasion for grief” for many Colombians).

277. See WEBSTER’S THIRD NEW INT’L DICTIONARY 179 (1993) (defining “barrio” as “a ward, quarter, or district of a city or town in Spanish-speaking countries”).

278. See BOWDEN, supra note 221, at 266 (noting that those who lived in Medellin were especially aggrieved over Escobar’s death and directed “promises of revenge” toward the government).

279. See id. at 253–54 (describing the circumstances surrounding Escobar’s death and the possibility that he was executed during the Search Bloc’s raid).

280. See supra note 198 and accompanying text (stating the common law rule that lethal force may be used against a fleeing felon).

281. This, in fact, is what journalist Mark Bowden concluded after
firefight that erupted between Escobar and the authorities when the Search Bloc stormed Escobar’s hiding place does not justify his killing either, at least if the aforementioned accounts of the incident are believed. Escobar was killed while attempting to escape from the police. Assuming that the police agents shot him twice and, by doing so, successfully prevented him from escaping, the previous firefight did not give them lawful authority to use deadly force against him again if, at the time, Escobar was already neutralized and no longer posing a threat.

C. Was Escobar’s Killing Justified?

Although the facts of Escobar’s killing suggest that Colombian authorities may have violated the law, the death did not generate outrage amongst the Colombian people, the academy, or the international community. At an abstract level, this is rather startling. How can the possibility of an extrajudicial execution carried out by governmental authorities not raise serious concerns? At a more basic level, however, there is nothing puzzling about the lack of attention paid to the legality of Escobar’s killing. Everyone knew that Escobar was a ruthless killer. There was no doubt that he was responsible for some of the worst crimes perpetrated in the latter portion of the twentieth century. He was still a very dangerous man, and it seemed impossible to bring him to justice. Members of the judiciary were either in cahoots with him or feared for their lives. There seemed to be no way out. Escobar had been imprisoned, and he had escaped. Capturing him alive was an option fraught with peril.

The extrajudicial killing of any human being is a matter of grave concern, and there remain powerful arguments against the killing of Escobar. We are also mindful that Escobar’s killing did not take place in a vacuum, and that the struggle against and between Colombia’s drug lords included other extrajudicial killings under different, less sympathetic circumstances, including many by paramilitary groups suspected of government

examining the evidence. According to Bowden, the evidence suggests that the shot that likely killed Escobar “is consistent with a shooter administering a coup de grâce while standing over a downed man.” Bowden, supra note 221, at 254.
ties. Our aim is not to endorse Colombian government policy as a whole, or even to mount a full-blown defense of the Escobar killing.

Our aim, instead, is more modest: to consider the Escobar case in isolation and to identify features that favor its justification. Escobar’s case presents, in many ways, the best case that could be made for ordering the killing of someone outside of war, in circumstances not necessary to prevent the person’s escape, or to neutralize an imminent threat to the life or limb of government officials or third parties. In other words, Escobar’s killing presents us with the best case that can be made for justifying a TK 4 case.

At least three factors seem to favor a justification for Escobar’s killing. First, although not judicially established, Escobar’s guilt for killing innocent human beings on numerous occasions was never in serious doubt. This, of course, is not enough to justify killing him in circumstances other than those that allow for the use of deadly force pursuant to law-enforcement authority, or defense of self or others. Absent these circumstances, there are very powerful reasons militating in favor of capturing and trying an individual for his crimes, rather than allowing government officials to kill him without affording him due process of law. Nevertheless, the fact that Escobar’s guilt was clearly established helps explain—along with other factors—why an order to kill rather than capture him may have been justified in this particular case.

A second factor justifying Escobar’s killing was that he clearly remained a dangerous individual with the potential to cause massive amounts of harm in the future. Even if it is assumed that Escobar did not pose a threat to the life or limb of the police officers or third parties at the time he was killed, there were good reasons to believe that he would continue to engage in serious crimes in the future. Once again, this alone is not enough to justify killing Escobar. Many, if not most, people who are guilty of committing homicide are dangerous and could thus continue to engage in similar conduct in the future. Nonetheless, Escobar’s

282. See, e.g., supra notes 253–58 and accompanying text (describing actions of paramilitary groups).
dangerousness was different in kind from that of the typical violent felon. Escobar was widely considered during the 1980s and up to the time of his death to be the most dangerous drug trafficker in the world.\footnote{See id. at 138 (stating that DEA Chief Toft labeled Escobar “the most notorious and dangerous cocaine trafficker in history”).} His crimes were different in scale from the offenses committed by ordinary criminals. Before car bombs became associated with the Middle East, Escobar used them to kill innocent people and terrorize the citizenry.\footnote{See id. at 57 (describing Escobar’s attempt to kill a high-ranking Colombian general in charge of hunting him down by setting off a car bomb alongside his vehicle, wounding numerous civilians).}

Escobar was complicit in the killing of police officers and presidential candidates.\footnote{See supra notes 226–42, 248–49 and accompanying text (describing Escobar’s numerous egregious crimes).} He was widely believed to have played a role in the killing of half of the members of the Colombian Supreme Court.\footnote{See supra note 234 and accompanying text (noting that Escobar was accused of being responsible for the deaths of multiple Colombian Supreme Court Justices).} There was overwhelming evidence suggesting that he ordered the downing of a commercial jetliner.\footnote{See supra note 233 and accompanying text (referencing accusations that Escobar downed a commercial airliner as a means to assassinate a potential presidential candidate).} Escobar was no ordinary criminal. And his dangerousness was not ordinary either.

Furthermore, Escobar showed no signs of slowing down, so the authorities had every reason to expect his criminal activities would continue and that, as a result, so would the body count. His death put an end to the bloodbath. Coupled with the fact that there was no doubt that Escobar was guilty of past serious offenses, the case in favor of neutralizing him by whatever means necessary was strong.

In spite of this, Escobar’s dangerousness could not, in and of itself, justify killing him extrajudicially, even though his guilt for committing serious offenses was not in doubt. These factors clearly provide enough reasons to justify using force—deadly if necessary—to capture an individual. They do not, however, justify killing a person in a nonconfrontational scenario if it is assumed that he could have been neutralized before he was
An additional element was present here: the unavailability of the traditional justice system as a secure means to punish and incapacitate him.

Trying Escobar was virtually impossible. He had managed to render the judiciary ineffective, at least with regard to him, by either killing judges who were willing to take a stand against him, or by bribing those who would accept the money. Furthermore, even if trying him would have been feasible, imprisoning him would have been nearly impossible. As his escape from La Catedral demonstrated, Escobar was able to get to almost anyone in order to have his way, including prison guards and government officials. At the time, there appeared no way of successfully trying, and subsequently imprisoning, Escobar in Colombia. Finally, even if it is assumed, for the sake of argument, that Escobar could have been successfully tried and imprisoned, there is little reason to doubt that doing so would have endangered the lives of those who played a role in the process, including the officers who would escort him to and from court, the judges who would preside over his trial, and the guards who would be in charge of his custody. Additionally, Escobar’s trial and imprisonment could very well have endangered the lives of civilians who did not play a role in his capture and detention. Escobar had shown that he was willing to do anything to get what he wanted, and he had no trouble killing innocent people. He would not have hesitated to do what he needed to prevent trial and imprisonment, even if it meant killing innocents. Given this reality, Colombia had strong reasons to avoid paying the price involved in capturing and trying Escobar. These factors help explain the absence of complaints in Colombia or elsewhere following Escobar’s death.

Despite these considerations, there remains at least one troubling aspect of Escobar’s demise, even under an account that gives weight to the factors outlined above. This is the reported fact that Escobar was already cornered and debilitated at the time of his shooting. The problem here is similar to the one we

---

288. See supra note 281 and accompanying text (noting that Escobar was probably neutralized before the fatal shot was fired).

289. See Schmidle, supra note 4, at 43 (stating that bin Laden was found unarmed).
briefly considered with respect to hors de combat status as defined in IHL.290 Under the war paradigm, the permissive approach to killing ceases upon the target’s incapacitation, namely upon his having been rendered unconscious “or otherwise incapacitated by wounds or sickness” and being, therefore, “incapable of defending himself.”291 This distinction reflects the logic of risk—combatants are presumptively dangerous and are targetable only so long as that presumption retains plausibility—but not entirely. The wounded soldier maintains her immunity from attack even as she is evacuated to a friendly hospital where she may recover to rejoin the fight. Manifest, here, is also a sense of chivalry, a moral revulsion against killing the helpless irrespective of their potential future danger.

This applies to the Escobar case, as well. The considerations favoring his killing, as we have outlined them, apply whether or not he was at large at the time of his targeting. But even if one accepts a more permissive approach to lethal force in cases like Escobar’s, the killing of a suspect already arrested presents a different moral calculus, one in which the use of lethal force becomes indefensible, even in the extraordinary circumstances we have described. One might also draw the line at the moment of Escobar’s incapacitation, when he effectively fell into the power of government authorities.

This qualification, however, is not fatal to the broader claim that a more permissive approach to killing was appropriate in Escobar’s case. For example, one could object to the specific circumstances of his killing while nevertheless maintaining that he, like a combatant in armed conflict, was appropriately subject to being targeted in a broad set of nonconfrontational scenarios, so long as he remained at large.

V. Pablo Escobar and the Morality of Targeted Killings

The suggestion that Escobar might have been a justifiable TK 4 target presents a prima facie conflict with the limits that

290. See supra note 42 and accompanying text (discussing hors de combat status).
291. Additional Protocol I, supra note 39, art. 41(c).
deeply ingrained principles of criminal law impose on official action. In this Part, we discuss how a narrowly drawn TK 4 justification might be integrated into conventional criminal law thinking. In particular, we consider the relationship of this justification to limits on preemptive action, punishment philosophy, and due process values.

In addition, we consider whether this justification is reducible to a judicially administrable test. We acknowledge that there is indeterminacy in the framework we outline, and we observe that this indeterminacy is not qualitatively different from that already inherent in administering established justifications for the use of lethal force. That said, we are mindful that special caution is warranted when considering proposals to relax the norm against killing, including concerns about official abuse of an expanded rule and the risk of creating a slippery slope. If these concerns preclude the acceptance of a TK 4 justification within the law, our model may nevertheless have value in indicating when targeted killings can be tolerated as a matter of social morality, even if not technically legal.

Finally, we qualify our argument by highlighting several important questions that arise in the context of targeted killing but remain outside the scope of our analysis.

**A. Targeted Killing Partially Justified as Preemptive Action**

Most justifications afford individuals a defense to criminal liability when they harm another in order to prevent a future harm to self or others. Self-defense is a paradigmatic example. The person who employs defensive force takes preemptive action in order to avoid harm to his person. Although there is much debate about the reasons that justify acting in self-defense in borderline cases,292 there is wide agreement about the fact that the harm avoided by the defensive action is relevant to explaining the justifiable nature of such conduct.293


293. See id. at 365 (noting that “it is hard to see either the justice or efficacy
The use of deadly force pursuant to law-enforcement authority is also morally justified because it prevents a future harm. By using such force, police officers prevent harm to themselves, harm to others, and the harm inherent in allowing a dangerous felon to escape. 294 Admittedly, it is debatable whether the harm inherent in escaping from police custody is sufficiently grave to warrant using deadly force to prevent it. 295 Nevertheless, once it is accepted that the harm of escaping capture is significant enough to justify employing deadly force, it becomes clear that the justification for using such force is prospective rather than retrospective. The individual is not harmed to exact retribution for what he is suspected to have done, or to retaliate in response for past wrongdoing, but rather to prevent the occurrence of a future harm.

The use of force pursuant to the lesser-evils defense can be morally justified in a similar manner. The reason why it is acceptable to cause harm to an innocent person pursuant to this defense is that, by doing so, one prevents an even greater harm from taking place. 296 Once again, the moral justification for engaging in this harmful conduct is preemptive because the use of force in these circumstances is a way of averting an untoward state of affairs that will take place in the future, rather than a method for punishing the harmed individual for past behavior.

Can TK 4 killings be morally justified by appealing to such preemptive rationales? There is certainly a preemptive dimension to such targeted killings. A salutary consequence of killing a manifestly dangerous individual such as Escobar is that doing so...
eliminates the possibility that he will be involved in future attacks.297 If the possibility of future attacks is substantial, then the prospect of preventing them by killing the actor certainly counts in favor of this course of action.

Nevertheless, this factor alone should not be enough to justify the killing of an individual in TK 4 cases. Domestic criminal law circumscribes the authorization to use deadly force to cases in which the force is necessary to prevent an imminent future attack298 or a non-imminent, known future aggression that can only be averted by the preemptive use of deadly force.299 There are good reasons for imposing these strict limitations. When the attack is not sufficiently imminent, or there is uncertainty about when and if a future aggression will take place, it is likely that the conflict may be resolved without using deadly force.300 Perhaps the individual will not carry out the attack or the police may thwart it without resorting to deadly force. It is a generally accepted principle that force calculated to take life should only be used when all else fails.301 It is unclear whether

---

297. It is worth noting that killing the dangerous individual might also generate violence. For example, the associates of the killed individual might harm innocent people in order to avenge the death.

298. See Ferzan, supra note 167, at 222–23 (noting that the use of deadly force in “[d]omestic self-defense is wholly preventative,” but a threat of unlawful force must be imminent to “trigger the right to self-defense”).

299. See MODEL PENAL CODE § 3.04(1), (2)(b) (stipulating that the use of deadly force is justified if “immediately necessary for the purpose of protection against "death, serious bodily injury, kidnapping or sexual intercourse" on the "present occasion").

300. See Dressler, supra note 171, at 467 (noting the importance of “imminence” in self-defense justifications and that, because there is no imminence when a battered woman murders her sleeping spouse, “we will never know for sure . . . whether some other, less extreme, remedy would have been sufficient”).

301. This is why deadly force pursuant to self-defense, law-enforcement authority, and the lesser-evils defense is only justified when it is necessary to prevent a harm. See Graham v. Connor, 490 U.S. 386, 396 (1989) (authorizing the use of deadly force if the law-enforcement officer reasonably believes it is necessary to defend himself or others); see also MODEL PENAL CODE § 3.02 (1985) (stating that, under certain circumstances, “[c]onduct that the actor believes to be necessary to avoid a harm or evil to himself or to another is justifiable”); Dressler, supra note 171, at 466 (“Stemming from the common law, a core feature of self-defense law is that the life of every person, even that of an aggressor, should not be terminated if there is a less extreme way to resolve the
using lethal force is necessary in TK 4 cases because the agents who order the killing simply do not know with sufficient certainty whether the targeted individual is actually planning a future attack, one which can only be thwarted by killing the target. Although they might justifiably expect that future attacks will take place, that expectation is not enough to authorize the killing of a human being.

This does not mean that the possibility of thwarting a future attack is irrelevant to justifying the killing of the targeted individual. It surely is. More needs to be demonstrated, however, in order to flout the basic principle that a person should not be killed unless it is the only way of preventing future harm to the lives of others. The likelihood of averting a future attack is, in sum, a necessary, but insufficient, condition for justifying a TK 4 killing.

B. Targeted Killing Partially Justified as Retaliatory (Punitive) Action

It is sometimes morally justified for the government to inflict harm to a non-threatening individual deliberately. Sometimes it may even be morally justified for the government to kill a non-threatening person deliberately. The paradigmatic example is state-sanctioned punishment.

Many theories have been advanced to justify the imposition of punishment. Consequentialist theories of punishment—such as those rooted in deterrence, incapacitation, and rehabilitation—justify punishment primarily by reference to the good consequences that follow from its imposition. Deontological theories of punishment—such as retribution—justify punishment primarily by reference to the fact that punishing a person who

302. In cases of group crime, for example, killing a group’s leader might fail to prevent others from carrying out a planned attack.

303. This statement, however, assumes that the death penalty is morally justified, which is an issue that is beyond the scope of this Article.

304. See R. A. Duff, PUNISHMENT, COMMUNICATION, AND COMMUNITY 3–4 (2001) (discussing consequentialism, which “insists that the justification of any human practice depends on its actual or expected consequences”).
has done something worthy of condemnation is intrinsically good even if no additional good consequences follow from doing so.\(^{305}\) In spite of their differences, consequentialist and deontological approaches to punishment have something important in common. Both theories conceive of the imposition of punishment as a response to a wrongful act committed by the person to be punished.\(^{306}\) Thus, regardless of what is believed to be the aim of punishment, it is imposed \(\text{for an act of wrongdoing that took place in the past.}\)\(^{307}\) Otherwise, the imposition of sanctions would seem random and arbitrary because punishment is, by definition, something that happens only after a determination that the person to be punished has committed an offense.\(^{308}\) In other words, there is an essential feature of punishment that is backward-looking even if its imposition is justified by appealing to forward-looking consequentialist criteria.\(^{309}\) Punishment is imposed \(\text{because someone has done something wrong, even if the aim of imposing it is to deter others from engaging in similar acts in the future or to prevent the offender from recidivating}.\)\(^{310}\)

Can targeted killings in TK 4 cases sometimes be morally justified by appealing to reasons that are similar to those that

\(^{305}\) See Michael Moore, \textit{Placing Blame} 87 (1997) (stating that in retributivism “the good that punishment achieves is that someone who deserves it gets it”). Furthermore, “[p]unishment of the guilty is thus for the retributivist an \textit{intrinsic} good.” Id.


\(^{307}\) See \textit{id.} at 113 (“[N]o punishment can be coherently imposed when there is no wrongdoing \textit{no matter what theory of punishment one espouses}.”).

\(^{308}\) See \textit{id.} at 106 (stating that “punishment would simply not make sense without wrongdoing”). Furthermore, “imposing punishment without the commission of an offense would be akin to the state’s production of random and arbitrary violence.” Id. at 110.

\(^{309}\) See \textit{id.} at 109–10 (explaining that the connection between punishment and offense “highlights the centrality of the concept of ‘wrongdoing’ in explaining the true nature of punishment insofar as it entails inflicting pain upon a person \textit{for having committed an offense and not just for the sake of social protection}”).

\(^{310}\) See \textit{id.} at 106–07 (explaining that scholars have agreed over the years that punishment is a response to wrongdoing even though the aim of punishment varies with different theories).
justify the practice of state-sanctioned punishment? Perhaps a state-ordered targeted killing is nothing more than an extrajudicial way of imposing punishment. This is especially the case if it is assumed that the targeted person has committed grave crimes in the past.

The problem with justifying targeted killings in this manner is that it runs afoul of the basic principle that the state is not allowed to harm an individual for punitive purposes without first establishing his guilt in a judicial proceeding. There are many sound reasons for that principle. First, we trust the judicial process more than any other government process, both in terms of providing unbiased outcomes and in terms of providing the best forum for discovering the truth. Second, we are skeptical of consolidating in one branch of government the power to investigate wrongdoing, adjudicate guilt, and carry out sentences. We simply do not trust the executive to be judge, jury, and executioner. Although it can be argued that judicial proceedings are a formality when there is no doubt about the offender’s guilt, it is generally best to adhere to such formalities. Making exceptions to this rule would rapidly take us down a dangerous, slippery slope, which makes it very difficult to tell when a judicial proceeding is necessary.

This does not mean that targeted killings in TK 4 cases do not have a punitive dimension or that this feature of the practice is irrelevant to explaining its moral justifiability. Certainty about past wrongful acts committed by the targeted person is relevant to the legitimacy of authorizing his killing. If there is no certainty about such matters, it is not legitimate to order the killing. Such certainty, however, cannot by itself establish the moral legitimacy of orders to kill an individual in a nonconfrontational setting.

C. Targeted Killing Partially Justified Because of Difficulty or Dangerousness of Trying the Perpetrator

Targeted killings in TK 4 cases serve a preemptive function. The government protects its citizens when it kills a manifestly dangerous individual who is nearly certain to engage in future acts of wrongdoing. These killings also serve a punitive function.
There are good desert-based and consequentialist-based reasons for punishing a person who has engaged in unspeakable crimes. Although one would prefer that judicial means be used to adjudicate guilt and mete out punishment in such cases, achieving justice is not always so easy. Ordering the killings of individuals may sometimes be the only available means of doing some kind of justice. Nevertheless, morally justifying this practice requires more than demonstrating that it serves preemptive and punitive functions. In the case of preemptive strikes, deadly force is usually authorized only if its use is necessary to thwart an imminent attack.\(^{311}\) If there is no imminent attack, the use of deadly force is usually not necessary and therefore not morally justifiable.\(^{312}\) In the case of punitive acts, deadly force is only authorized pursuant to a judicial determination of guilt. The rule of law requires no less.

There are a handful of cases, however, in which ordering the use of deadly force in a nonconfrontational setting and without a judicial adjudication of guilt becomes more defensible as a reasonable and morally acceptable course of action. The cases that come to mind involve TK 4 scenarios in which deadly force is used in a nonconfrontational setting against a dangerous individual who has engaged in serious crimes in the past. The use of deadly force appears to be morally justified when trying the individual will be either extraordinarily difficult or unacceptably dangerous. The difficulty or dangerousness of capturing and trying the targeted person is morally relevant to ordering her killing because it makes a judicial determination of guilt impossible or too risky.

The difficulties involved in trying the individual can be insurmountable, depending on the circumstances. Perhaps the individual cannot be apprehended without exposing law enforcement or bystanders to excessive risks. The jurisdiction in which the individual is going to be tried might not have a properly functioning legal system, if it is in a war-torn area, for instance. In other cases, despite the existence of a legal system

---

311. See supra notes 169–72 and accompanying text (discussing preemptive strikes and the requirement of an imminent attack).

312. See id. (explaining that deadly force is not necessary without an imminent attack during a preemptive strike).
that works acceptably in most cases, the trial of a particular person or group of persons is not possible for whatever reason.

This is what happened in the case of Pablo Escobar. Although Colombia’s legal system was not entirely dysfunctional, it was not properly equipped to deal with the likes of Escobar and other top-level drug traffickers.313 As a result of Escobar’s *plata o plomo* policy, there appeared to be no judge in Colombia who could preside over his trial.314 Those courageous enough to defy him were killed.315 Those who were not as courageous were bought off by Escobar.316 By the late 1980s, it was obvious that the Colombian judiciary did not have the tools necessary to try a man as ruthless and dangerous as Escobar.317

In addition, trying the individual could sometimes be unwarranted because, although logistically feasible, doing so would prove unacceptably dangerous. It is important to note the risks inherent in detaining and trying the suspect would need to be extraordinarily high before contemplating an extrajudicial killing.

It is not enough, in our view, to claim that trying the targeted individual might jeopardize the lives of innocent people. Speculative harm, even probable harm, is not enough to justify dispensing with due process. One would have to know to a substantial certainty that trying the targeted individual would cause significant harm to innocent parties. It would also be necessary to know to a substantial certainty that the harms inherent in trying the individual would significantly outweigh the harms inherent in killing him.

313. See Bowden, *supra* note 221, at 51 (explaining that Pablo Escobar, through his power and popularity, was able to buy off and threaten the Colombian court system).

314. See id. (stating Escobar’s *plata o plomo* strategy became so effective by 1984 that Escobar became untouchable to the courts, and Colombia’s democracy was undermined).

315. See id. at 53 (stating that by the end of 1986 there were few judges alive who defied Escobar’s *plata o plomo* strategy).

316. See id. at 52 (stating that Escobar was responsible for the deaths of at least thirty judges).

317. See id. at 53 (discussing that by the late 1980s the drug cartel had taken over and “Colombia had been corrupted and terrorized to its core”).
This last qualification is essential, given that every course of action generates certain kinds of benefits and costs. Although killing the individual could have the salutary consequence of stopping some violence, it could also generate significant societal costs inherent in not observing due process. It might also generate retaliatory bloodshed. On the other hand, trying the individual could have the positive effect of enhancing the perceived legitimacy of imposing punishment on the individual. It might, however, generate negative consequences, such as kidnappings, suicide bombings, and other terrorist attacks orchestrated by the detainee’s supporters intent on securing his release.

In sum, regardless of the specific costs and benefits that attach to either killing or trying the individual, killing in a TK4 case should not be permissible unless it is known with substantial certainty that the risks inherent in capturing and trying the individual are extraordinarily high and significantly greater than the risks inherent in killing without trial. Because the right to due process of law is crucial to maintaining a legitimate and just system of criminal justice, this right must always be observed, except in the most extraordinary of circumstances, when killing without capture and trial is necessary to avoid certain and significant harm to innocent human beings.

Escobar’s case is again a case on point. Even if logistically feasible, trying Escobar was unacceptably dangerous. The government was acutely aware of this because an organization known as “Los Extraditables” publicly vowed to kill any judge who dared try Escobar for murder.318 It also threatened to kill the families of those who wanted to indict Escobar in the local courts.319 Escobar’s supporters were also known for kidnapping innocent people whom he would then use as bargaining chips to obtain what he wanted from the government.320 There was,

318. See id. at 55 (revealing that, when a judge had attempted to indict Pablo Escobar for murder, Los Extraditables threatened to kill the judge and his family).

319. See id. (“We are capable of executing you at any place on this planet . . . in the meantime, you will see the fall, one by one, of all the members of your family.”).

320. See id. (detailing how supporters of Pablo Escobar kidnapped “the
therefore, reason to expect with some certainty that kidnappings would have increased had Escobar been caught and detained for trial. Escobar had also successfully escaped from prison the one time he had been detained. He would surely have attempted to do so again, and he had demonstrated an ability to kill those who might stand in his way.

Asking for a judicial determination of guilt in cases like Escobar’s is quixotic. Sometimes a legal system is not equipped to make the necessary determinations, and setting the judicial wheels in motion is likely to trigger a series of events that endanger so many innocent people that it is better not to proceed. Circumstances such as these do not occur often. But when they do, ordering the extrajudicial killing of the individual might not only be sensible but also morally justifiable.

D. Justifiable Targeted Killings in TK 4 Cases—In Search of a Test

The TK 4 justification we have outlined centers on three criteria, all present in the Escobar case. First, an extrajudicial killing in a TK 4 case should only be authorized if the targeted individual is likely to carry out, or to substantially help carry out, atrocities in the future. There need not be proof of an imminent attack that is being planned by the individual, but there must be an expectation, grounded on specific and articulable facts, that would lead a reasonable person to conclude that a future attack on innocent human beings is likely, even if its occurrence is not absolutely certain. Furthermore, the future attacks must be of a particularly grave nature that transcends the commission of a discrete offence. As a general rule, the attack must be part of some widespread or systematic campaign. Future large-scale terrorist attacks surely satisfy this standard. A typical robbery or murder will not.

---

journalist son of a former president and Conservative candidate for mayor of Bogotá”.

321. See id. at 107–55 (describing the details of Escobar’s escape from prison).
Second, there must be no reasonable doubt about the targeted individual’s responsibility for past atrocities. This requirement is often satisfied because the targeted individual has publicly taken responsibility for such atrocities.

Third, the capture and trial of the individual must be logistically impossible or extraordinarily dangerous. As we discussed in the previous subpart, this standard is limited to exceptional circumstances and will seldom be satisfied because there is a very strong presumption in favor of affording the targeted individual the due process of law. The legal system must be unavailable to try the individual because it is not functioning, because it is impossible to try the specific individual for reasons such as the killing and intimidation of those judges willing to try the targeted individual, or because capture and trial would involve a substantial certainty of harm to innocent parties—a harm that significantly outweighs the harms inherent in killing the individual.

Even if one accepts, in principle, that certain killings meeting the above criteria can be morally justified, distinct concerns center on the desirability of reducing these factors to a legal test. It may be objected, for example, that the factors we have outlined fail to provide sufficient guidance in concrete cases and are resistant to judicial application (assuming that judicial review is available), thus exposing the law to a slippery slope. There is also the potential for governmental abuse, in which the TK justification is invoked to legitimize killings that do not rightfully fall within the narrow exception we have outlined.322

As a general matter, it is difficult—if not impossible—to craft a bright-line test for determining when the killing of a human being should be justified. Even in easy cases, such as self-defense, the general parameters of the justification are crafted in relatively vague terms that allow for some leeway in the application of the rule. Thus, we say that using deadly force in self-defense is justified if such force was reasonably believed to be necessary to avert an imminent and unlawful aggression, and only if such force was proportional to the threatened harm.323

---

322. See Waldron, supra note 71, at 5–9 (discussing bad faith examples of targeted killing).
323. See Dressler, supra note 171, at 461 (stating that the traditional rule
precise scope of the italicized term's is rather fuzzy, especially in borderline cases. Is the reasonable belief to be judged from an impartial perspective that does not take into account the experience and physical and mental attributes of the actor, or should it be judged according to what would appear reasonable in light of the specific traits and experience of the individual? Is the force used by the actor necessary only if it is the sole available means to avert the threat, or is it necessary as long as the actor is unaware of other means that may also defuse the threat? Does a threat to use force within the next five minutes count as an “imminent” threat, or does the defender need to wait several minutes until the force is about to be employed? Is a threat unlawful if it violates any law or regulation regardless of whether it is civil, administrative, or criminal law, or is it unlawful only if it violates some particular body of law (criminal or tort law, for example)? Finally, is force lawful only if it is strictly proportional to the harm threatened, or is some degree of disproportionality allowed?324

Determining whether deadly force is lawful pursuant to other justification defenses is equally problematic. Consider the case of deadly force used pursuant to law-enforcement authority. As a general rule, such force is only lawful if it is necessary to prevent the escape of a fleeing felon that poses a threat to the lives of the officer or third parties.325 Furthermore, as the U.S. Supreme Court pointed out in Scott v. Harris, deadly force used by law-enforcement agents comports with Fourth Amendment standards only if the use of such force was reasonable given the dangerous

---

324. This is an under-studied question in American criminal law. It would seem that self-defense allows for the use of disproportionate force as long as it is not grossly disproportionate. For example, in every jurisdiction one may use deadly force to avert serious bodily injury, although the force used causes more harm (death) than the threatened harm (grave bodily injury). Similarly, it would seem that one may use deadly force to thwart rape.

325. See supra notes 198–203 and accompanying text (discussing what force is necessary to use for law enforcement in the case of a fleeing felon).
circumstances. Although perhaps necessary, the vagueness of this approach is evident.

Given that the parameters for using deadly force are fuzzy even in the case of well-established justification defenses, such as self-defense and law-enforcement authority, it should come as no surprise that we have not provided a bright-line test for authorizing targeted killings. It is simply impossible, in this context, to come up with something other than general guidelines that frame the relevant issues that ought to be considered when assessing the justifiability of targeted killings in TK 4 cases.

Although admittedly fuzzy, the framework proposed here does not strike us as being significantly vaguer or more problematic than the general frameworks that are currently in place to assess the justified nature of force used pursuant to self-defense and law-enforcement authority.

We further acknowledge the possibility—indeed probability—that some governments will abuse a legal doctrine establishing a justification along the lines we have set forth. Once again, however, this problem is ever-present and not unique to the particular context of our analysis. A government wishing to abuse its authority under the cloak of the law already has ample room to do so within existing accepted legal doctrines, for example, by manufacturing claims of self-defense to justify what is in fact an impermissible extrajudicial killing. Indeed, we suspect this would typically be the easier path, considering the strictness of the factors we have offered. The question, therefore, is not whether governments would seek to abuse a new legal doctrine justifying a limited number of TK 4 cases, but whether the establishment of such a doctrine would be uniquely susceptible to abuse. We doubt that it is.

Nevertheless, we are mindful of the sensitivity and caution that is warranted whenever contemplating any expansion of the legal permission to kill. We therefore have our doubts about the advisability of adopting our framework as law, but it may nevertheless have value as a measure of social morality. In other words, even if the law does not itself justify any TK 4 killings,

326. See Scott v. Harris, 550 U.S. 372, 383–84 (2007) (stating that the officer's actions were reasonable considering the high likelihood, although not certainty, of danger from the fleeing motorist's driving).
those that meet the criteria we have outlined will receive a de facto public justification. In this sense, our framework has descriptive value: it identifies circumstances in which TK 4 killings, whether or not legally permitted by codified law, will receive broad public acceptance and prove resistant to official scrutiny.

E. Some Qualifications

The criteria we have identified provide, in our view, the best justification for using lethal force in a TK 4 case. It is important to acknowledge, however, several points that further qualify and limit our framework.

First, we do not address the threats to international security that can result from targeted killings that cross international borders. In suggesting that Colombian authorities may have been justified in using lethal force to neutralize Escobar in a nonconfrontational setting, 327 we do not suggest that the United States, for example, could have invoked the same justification to deploy agents into Colombian territory to target Escobar without the consent of Colombian authorities. Whether and when the interests justifying a TK 4 targeting might likewise justify a cross-border intervention is an important question that remains outside the scope of this Article.

Second, we do not maintain that the interests justifying a TK 4 targeting might likewise justify the expected incidental loss of innocent life, as is permitted, for example, by the IHL proportionality test that applies to the conduct of armed conflict. 328 Our framework instead presumes the continued force of the general criminal law rule dictating that governmental agents may not knowingly take innocent life even if doing so is the only way of killing or capturing a fleeing felon or an otherwise dangerous individual. 329 Whether or not extraordinary

327. See supra Part IV (discussing the Escobar case).
328. See supra note 126 and accompanying text (describing the proportionality requirement).
329. See supra Part III.D (providing the law-enforcement defense).
circumstances might justify relaxing this rule is a matter we do not consider.\footnote{330 For an analysis of this question as applied to the defense of self and others, see Chiesa, \textit{supra} note 306.}

Finally, we emphasize that the three-part test we have identified requires an individualized assessment of the targeted individual. For example, an individual targeted based solely on her membership in a group that meets our criteria only on a collective basis would fall outside the framework. For the criteria to apply, they must be met on an individualized basis.

\textit{VI. The Killing of Osama bin Laden as Self-Defense or Law Enforcement}

With this framework in mind, we now return to the killing of Osama bin Laden and consider whether justifications other than that provided by IHL might apply.

\textit{A. Was bin Laden's Killing Justified Pursuant to Self-Defense?}

Assuming, for present purposes, that IHL is inapplicable, might bin Laden's killing nevertheless have been justified as an act of self-defense? Of course, the answer depends on what exactly transpired the day of his killing. If he was in fact armed, or if he fired at members of the Navy SEAL team that raided his compound, his killing would amount to a justifiable act of defense of self or others.

Matters become more complicated, however, if one assumes—as some reports indicate\footnote{331 See, e.g., Schmidle, \textit{supra} note 4, at 43 (stating that bin Laden was unarmed).}—that bin Laden was unarmed and did not threaten physical violence. In that event, killing is not as easily justified as an act of self-defense. The reason for this is simple. The use of force in self-defense is triggered by the use or threat of unlawful force.\footnote{332 See \textit{Fletcher & Ohlin, supra} note 64, at 89 (explaining that it would be “hard to find a national statute on self-defense that failed to require that the attack be unlawful”). The unlawful aggression must also be
imminent. If bin Laden was unarmed and not about to attack the SEAL team, there would have been no imminent wrongful aggression to trigger the right to use defensive force.

It should be noted, however, that the absence of an imminent and wrongful aggression is not in and of itself fatal to a claim of self-defense. Bin Laden's killing would also be justified in self-defense if the shooter acted upon a reasonable belief that bin Laden was about to attack them. The reasonable belief would justify the killing even if the belief happened to be mistaken. This is, in fact, one of the arguments advanced by U.S. officials in defense of the bin Laden killing.

But what if one assumes that it was unreasonable for the SEAL team to believe that bin Laden was about to launch an attack? Of course, we will likely never know whether such a belief was reasonable or not for the SEALs to hold. For that matter, we may never know whether individual members of the team subjectively believed that bin Laden was a threat. For all we know, the question may have played no role in the operation because the team members were acting under rules of engagement derived from IHL. Thus, it is worth asking whether the killing of bin Laden would have been justified even if the SEALs did not believe that their lives were in danger at the time.

333. See id. at 90–91 (explaining that most jurisdictions require an unlawful attack to be imminent to justify self-defense).

334. See, e.g., supra note 181 and accompanying text (describing one state's self-defense law, which requires reasonable belief); People v. Goetz, 497 N.E.2d 41, 49 (N.Y. 1986) (applying the test that a sufficient basis to use deadly force is justified if "the situation justified the defendant as a reasonable man in believing that he was about to be murderously attacked").

335. Goetz, 497 N.E.2d at 48 (emphasizing that if a reasonable belief is established, then "deadly force could be justified . . . even if the actor's beliefs as to the intentions of another turned out to be wrong").

Self-defense is off the table if this assumption is made. A person’s lack of an objectively reasonable belief that his life or the life of another is in danger is fatal to a claim of self-defense.337

The same would be true under an expanded version of self-defense. Even if we assume that the U.S. government possessed credible and specific information that bin Laden was going to launch an attack against American interests in the near future, the killing could not be justified as an instance of preemptive self-defense unless killing was the only way to prevent the attack.338 This requirement could not be met under the circumstances if bin Laden were readily susceptible to capture. Moreover, absent the type of specific information we have just hypothesized, mere speculation about possible attacks that bin Laden might launch in the future would not be enough to justify killing him pursuant to the standard arguments favoring preemptive self-defense.339

B. Was bin Laden’s Killing Justified Pursuant to Law-Enforcement Authority or the Lesser-Evils Defense?

Governmental agents may on occasion use more force in furtherance of law-enforcement authority than they may use pursuant to self-defense or defense of others. Although deadly force in defense of self or others may only be employed in order to deflect an imminent aggression, police officers may sometimes use lethal force in order to prevent a fleeing felon from escaping even if, at the time the force is used, the felon is not threatening imminent harm.340 This defense, as Finkelstein has persuasively argued, may thus justify certain killings in circumstances in which traditional self-defense is inapplicable.341 Can it be invoked to justify bin Laden’s killing?

337. See Goetz, 497 N.E.2d at 48 (stating that for self-defense to be justified, there must “be a reasonable basis, viewed objectively, for the beliefs”).
338. See supra Part III.C (explaining the elements of an expanded version of self-defense).
339. See supra Part III.C (explaining the elements of an expanded version of self-defense).
340. See, e.g., supra Part III.D (describing when law-enforcement agents may use deadly force against felons).
341. See supra notes 196–97 (discussing Claire Finkelstein’s theory that
Again, the answer to this question depends on the circumstances surrounding the killing. If bin Laden was attempting to flee, then the SEAL team members might have been justified in using deadly force to prevent his escape. On the other hand, if the SEALs fired at bin Laden without first attempting to arrest him, or if bin Laden did not attempt escape, conventional law-enforcement authority could not supply a justification for the use of deadly force.

Assuming that bin Laden was not armed when he was killed and that he did not threaten harm, the choice-of-evils or necessity defense would likewise fail: necessity may only be invoked to justify a use of force that is necessary to prevent an imminent harm from taking place. Once again, bin Laden’s dangerous character does not by itself justify using deadly force against him. The use of force pursuant to necessity, like self-defense, is triggered by the threat of suffering imminent harm, not by the possibility that a dangerous individual will try to cause harm in the future.

C. Was bin Laden’s Killing Otherwise Justifiable?

Assuming that bin Laden’s killing was not supported by a reasonable belief that he threatened to harm the SEAL team members, it becomes difficult to justify by appealing to the conventionally accepted justification defenses recognized under domestic law. Might bin Laden’s killing be an instance of a TK 4 case, in which ordering the killing of an unarmed and non-threatening individual is nevertheless justifiable?

The case is similar in some aspects to Escobar’s case. First, bin Laden had publicly taken responsibility for engaging in unspeakable crimes. There was thus no doubt about his targeted killings might be legally justified pursuant to the law-enforcement authority to use deadly force).

342. See supra Part III.E (explaining the elements of the necessity defense for use of force).

responsibility for the death of thousands of innocent human beings. As a result, trying him would have been, in a sense, a formality. Although perhaps deemed necessary to uphold the rule of law, bin Laden’s trial would not really be necessary to establish his guilt beyond a reasonable doubt.

Second, bin Laden, like Escobar, was the head of a very dangerous organization that was still operating at the time of his killing. We had (and still have) good reasons to believe that al Qaeda will attempt to kill innocent people in the future.\textsuperscript{344} As a result, bin Laden’s killing could be viewed as serving both punitive and preventive functions. On the one hand, his killing could be construed as punishment for his past crimes. On the other, his killing could prevent crime by helping debilitate al Qaeda’s command structure.

In a typical case, these considerations are insufficient to justify killing a non-threatening individual. As a general rule, deadly force should only be authorized in order to defuse an imminent threat or pursuant to a sentence secured after affording the individual the due process of law.\textsuperscript{345} As Escobar’s case demonstrates, however, extrajudicially killing the individual may be a reasonable course of action, even assuming capture is feasible, if trying him is either impossible or unacceptably dangerous. This is where significant differences arise between Escobar’s case and bin Laden’s.

First, it is difficult to argue that trying bin Laden would have been logistically impossible. The United States certainly has the resources to orchestrate trials of dangerous individuals. Although setting up a fair trial for bin Laden might have been difficult and costly, there is no doubt that it would have been feasible.

Second, it is speculative to assume that the dangers associated with capturing and trying bin Laden would have been sufficiently high to justify dispensing with a trial. Of course, trying bin Laden would have posed certain risks. For instance,

\textsuperscript{344} See Obama, supra note 1 (“[The United States] quickly learned that the 9/11 attacks were carried out by al Qaeda—an organization headed by Osama bin Laden, which had openly declared war on the United States and was committed to killing innocents in our country and around the globe.”).

\textsuperscript{345} See supra notes 170–73 and accompanying text (explaining that deadly force is usually authorized in cases when the force is necessary to prevent an imminent future attack).
his capture and trial might have prompted retaliatory terrorist attacks or put American citizens at risk of being kidnapped and used as bargaining chips to gain his release. This is a risk that Colombian authorities had to ponder when deciding whether to capture and try Escobar. Although these concerns should not be trivialized, one can only speculate with regard to the likelihood that they would actually materialize. Also, it is quite possible that the risks inherent in killing bin Laden were as significant as the risks inherent in capturing and trying him. The risk of retaliatory attacks, for example, exists under either scenario.

We do not believe these risks suffice to justify dispensing with a trial in bin Laden’s case. The decision to order the extrajudicial killing of an individual should not be taken lightly, and speculative assessments of the dangerousness of trying the actor should not justify taking such momentous action.

At the same time, however, we believe that the reasons that undergird our theory of justifiable killings in TK 4 cases may explain public intuitions about the bin Laden killing better than the reasons associated with the other available justifications for targeted killings. That is, many people seem to believe that killing bin Laden was the right thing to do because he was a dangerous individual, he was responsible for mass atrocities, and trying him would be a complicated affair. This helps explain why President Obama asserted that “justice has been done” when bin Laden was killed. This assertion does not sound like the language of preventive self-defense or law-enforcement authority. It does not sound like the language of national self-defense either. The American ideal is that justice is meted out in the courtroom after observing due process and judicially establishing the defendant’s guilt. A killing in individual or national self-defense is not a way to do justice, but rather a way to defuse a threat. The language originally used by President Obama to describe the killing did not fit this preventive paradigm. Similarly, most Americans who celebrated bin Laden’s death described it as something that provided them with some “sense of closure.”

346. See supra note 320 and accompanying text (discussing that Escobar kidnapped innocent people to threaten legal authorities).
347. Obama, supra note 1.
348. See, e.g., David Jackson, Bush on Bin Laden: ‘A Sense of Closure,’ USA
But closure has nothing to do with self-defense. Closure is what one feels after justice has been done.

What the President tapped into, we believe, was a deeply felt and widely shared intuition that some extrajudicial killings that do not squarely fit within the self-defense or law-enforcement paradigms may nevertheless be justified for both preventive and retributive reasons. This is the same intuition underlying the widespread agreement regarding the justifiability of Pablo Escobar’s killing. Escobar’s killing appeared morally acceptable, even if not in self-defense or pursuant to law-enforcement authority, because he was dangerous and he was undoubtedly responsible for past atrocities of an incredible scale. This is why his death provided a sense of closure for Colombians. Escobar’s killing—like bin Laden’s—was a way of preventing possible future attacks, but it was also a way of doing some justice.

Accordingly, and unlike the war paradigm, this account provides a superior explanation of why one might defend a more tolerant approach to justifying killing—an approach that does not require the feasibility of capture—in cases that do not share the functional requirements generally associated with war. To the extent that governments turn to targeted killing in this context, there is value to considering such cases on their own terms, rather than stretching IHL to contexts far removed from the battlefield realities that led to IHL’s creation.

The problem with the Escobar analogy, as we have argued, is that it was both logistically impossible and extraordinarily dangerous for Colombians to try Escobar. Nevertheless, it appears that it was logistically possible and not unacceptably dangerous to try bin Laden. Therefore, although feelings of closure and justice are understandable responses to the killing of a dangerous individual and a mass killer, bin Laden’s extrajudicial killing is difficult to justify under the residual justification that explains the moral propriety of certain TK 4 killings. Retributive and preventive reasons are necessary, but not sufficient, conditions to authorize an extrajudicial killing. In

addition, it must be clear that trying the targeted individual is not feasible. If bin Laden could have been tried, due process should have been observed. Accordingly, although we have suggested that some targeted killings may be justifiable even if they fall outside the traditional defenses supplied by IHL and the criminal law, the bin Laden case does not appear to fall within this narrowly defined category.

VII. Conclusion

As we observed at the outset of this Article, the aim of our analysis is not to find a definitive answer to the legality of the bin Laden killing. Depending on circumstances that may never be known, the killing may have been readily justifiable even under the rules generally applicable to law-enforcement operations. If, for example, the killing resulted from a reasonable belief that bin Laden posed an imminent threat that could be defused only through the immediate use of lethal force, then the killing was justified as a classic case of self-defense or defense of others.

Our interest in the bin Laden case focuses instead on the legal landscape of targeted killing and asks whether the killing might be justified under a more permissive legal regime, one that relaxes the restrictions generally imposed by criminal law. Public statements of government officials have identified such a regime in the rules applicable to killing in wartime. We have raised questions about that account. Although there is much about IHL’s requirements that is indeterminate and debatable, we have argued that the best reading of IHL (assuming this law applies to non-battlefield scenarios, such as the bin Laden killing) is inconsistent with the view that bin Laden remained a proper target even if capture was feasible under the circumstances.

We have further considered whether some targeted killings might be acceptable, even if not supported by IHL, by the conventional justification defenses generally afforded to law-enforcement officials, or by more expansive models of preemptive self-defense. We have identified, in the case of Pablo Escobar, a best-case scenario for a rule that does not hinge on the existence of an armed conflict and, unlike IHL, emphasizes culpability considerations alongside preventive considerations. If the
Escobar case can supply such a standard, it is a narrow one: in addition to demanding that such killings be justified by compelling evidence, both of the target’s undeniable guilt for grave atrocities and of compelling evidence that killing will thwart serious future crimes, we argue that this category of targeted killings should only be allowed when it is not feasible to detain or try the individual because doing so would be logistically impossible or extraordinarily dangerous.

We have also argued that the bin Laden killing, considering the apparent feasibility of trial, does not fall within this narrow category of cases. Nevertheless, the model we have outlined retains explanatory power. It provides, we suspect, the best account of the public understanding of bin Laden’s demise, one in which his killing was an act of justice, and not merely the neutralization of a wartime opponent. More broadly, we believe that the criteria we have outlined identify circumstances in which similar operations are likely to receive public legitimation even if they fail de jure requirements.
The CARD Act on Campus

Jim Hawkins*

Abstract

In February 2010, the Credit Card Accountability, Responsibility, and Disclosure (CARD) Act intervened in student credit card markets in a dramatic way, attempting to prevent student over-indebtedness, to end aggressive marketing to college students, and to reveal and change avaricious agreements between credit card issuers and colleges. Yet, two years after it became effective, we still have little measurement of whether the Act has accomplished these goals.

This Article offers the first empirical assessment of the rationales for the CARD Act and the Act's effects. Over the two years since the CARD Act went into effect, I conducted surveys of more than 500 students at two different colleges. I also examined 300 agreements between issuers and college-related organizations, which the CARD Act made publicly available for the first time.

Based on this survey and study, I found that many of the CARD Act's student and young consumer provisions have not affected credit markets in the ways the Act's proponents had hoped. Young consumers are still qualifying for credit cards without enough earned income to pay off the debt, and students are still reporting high levels of credit card marketing efforts aimed at the students. Most strikingly, the requirement that credit card companies disclose the secret agreements between issuers and colleges has caused virtually no change in the number of these agreements or their terms.

* Assistant Professor of Law, University of Houston Law Center. I am grateful to Ronald Mann, Julie Hill, Jeff Brown, and the participants of a workshop at the University of South Carolina School of Law and the AALS Section on Commercial and Related Consumer Law Roundtable on the CARD Act for help on earlier drafts of this paper. I also want to express appreciation for Shaun Cassin, Jennifer Chang, and Rebekah Reneau for research assistance.
# Table of Contents

I. Introduction ........................................................................... 1473

II. Empirical Strategy for Measuring the Effect of the CARD Act’s Young Consumer Provisions ......................... 1478
   A. College Student Survey Methodology .......................... 1478
   B. College–Card Issuer Agreement Study Methodology .......... 1485

III. Understanding the CARD Act’s Young Consumer Provisions and Assessing the Act’s Rationales ............. 1490
   A. Ability to Pay ................................................................ 1491
   B. Restrictions on Marketing to Young Consumers .......... 1494
      1. Prescreened Mail Offers ...................................... 1494
      2. Tangible Gifts on Campus .................................. 1496
   C. College–Issuer Marketing Agreements ..................... 1498
      1. The Provision and Its Rationale.............................. 1498
      2. Assessing the Rationale for the College–Issuer Disclosures .............................................. 1500
   D. Misplaced Reliance on Credit Card Usage and Debt Levels .......................................................... 1505

IV. Measuring the Effectiveness of the CARD Act’s Young Consumer Provisions ......................................... 1512
   A. The Effects of the Ability-to-Pay Provision ................. 1512
      1. Qualifying as an Individual ................................. 1513
      2. Cosigning ....................................................... 1515
      3. Access to Credit ................................................ 1517
   B. The Effects of Restrictions on Marketing to Young Consumers ........................................................... 1518
      1. Prescreened Mail Offers ...................................... 1519
      2. Marketing on and off Campus and Marketing Using Gifts ......................................................... 1520
   C. The Effects of the College–Issuer Marketing Agreements Provisions ............................................. 1525

V. Conclusion: Lessons from the CARD Act’s Young Consumer Provisions ..................................................... 1529

Appendix A: The CARD Act Student Survey ...................... 1532
I. Introduction

The Credit Card Accountability, Responsibility, and Disclosure Act of 2009 (CARD Act)1 is the most important credit card legislation of our generation.2 Among the many important provisions of this ground breaking Act, the Act’s sponsors highlighted its protections of young consumers and college students as some of the most significant. Senator Christopher Dodd argued: “It is time to insist that credit card companies take into account a young person’s ability to repay before allowing them to take on what is all too often a lifetime worth of debt. Very little we do in our legislation will be more important than these provisions.”3

Senator Dodd was referring to several provisions that affect how credit card companies interact with students and young consumers. First, the CARD Act requires that credit card companies verify that people under twenty-one have the ability to repay their credit card debt.4 Second, it places restrictions on credit card issuers’ marketing activities aimed at young consumers, including prohibiting giving tangible gifts to students on college campuses and banning credit bureaus from giving out young consumers’ addresses.5 Finally, it obligates credit card companies and colleges to disclose their agreements about credit card marketing to students.6 The central goal of these provisions was to prevent young consumers from accumulating excessive credit card debt.7

---

2. See, e.g., Christopher L. Peterson, The CARD Act in Perspective: Ongoing Efforts to Find Balance in Credit Card Regulation, 2011 Utah L. Rev. 335, 336 (2011) (stating that “the country has struggled to strike a balance between the risks of consumer indebtedness and the convenience that credit cards provide” and calling the CARD Act “the most important effort to recast this balance in several generations”).
4. See infra Part III.A.
5. See infra Part III.B.
6. See infra Part III.C.
7. See infra Part III.D.
Student credit cards are a hotly contested issue, and the CARD Act’s young consumer provisions have similarly generated significant academic debate about their theoretical underpinnings and likely effects. But, two years after these protections became effective, we still have little empirical measurement of whether the Act’s goals have been achieved and whether either critics’ or supporters’ predictions about the Act have come true. While academics have conducted empirical studies on other aspects of the Act and have noted the efficacy


of many of its provisions, no one has measured the impact of the young consumer provisions. Members of Congress have been quick to congratulate the government for stopping “students from being sent credit card offers” without bothering to check if the CARD Act’s provisions have actually had that effect.

The Act’s consequences for college students and other young consumers should be the central concern of those studying these provisions. As Elizabeth Warren observed, while serving as Assistant to the President and Special Advisor to the Secretary of the Treasury on the Consumer Financial Protection Bureau,

[The Bureau] think[s] it is appropriate to ask whether [the Act] has had its intended effects and how the credit card marketplace has changed. Where there are clear causal links, we need to draw them out. And where the connections are more tenuous, we need to keep asking questions and analyzing data.


11. See Williams & Emley, supra note 9, at 1419 (“Early indications suggest that the CARD Act has been successful in eliminating some of the more controversial practices of card issuers.”); Frank, supra note 10, at 4 (arguing that the CARD Act has not made credit cards more expensive or less accessible).


This Article offers the first empirical measurement of the effects of the CARD Act’s young consumer provisions. To capture information about the Act’s effects, I conducted a series of surveys over a two-year period that asked more than 500 college students about their experiences with credit card companies. Also, I examined 300 agreements between credit card companies and colleges over two years, evaluating the terms of those agreements and any changes that had occurred since the CARD Act’s implementation.

The results are surprising. Contrary to the predictions of the Act’s sponsors, the Act’s restrictions on credit card companies’ activities have not substantially decreased the number of students reporting instances of credit card marketing. Similarly, provisions that require credit card issuers to evaluate young consumers’ ability to repay their debt have not prevented over-indebtedness among students. I offer data that demonstrate how students are using other forms of debt to qualify for their credit card debt. The starkest outcome of my research is the finding that requiring credit card issuers to disclose the terms of their agreements with colleges has had almost no effect on the number of agreements between issuers and colleges or on the terms of those agreements.

In addition to measuring the CARD Act’s effectiveness, information from the surveys and study calls into question some important rationales that academics and policymakers used for intervening in this market, while confirming other justifications for the Act. First, the low levels of student credit card indebtedness reported in the student surveys undermine the claim that the CARD Act was necessary to stop students from becoming overly indebted to credit card companies. Second, information from the agreements between issuers and colleges reveals that the claims that colleges are being incentivized to trap students in debt have been overstated. But, on the other hand, the agreements are primarily aimed at moving students into credit card accounts, a finding that confirms the suspicions of policymakers seeking disclosures.

By revealing the flaws in some of the justifications for the CARD Act and the ways that the Act has failed to live up to its potential, this Article hopes to guide policymakers as they consider amending the Act. In addition to informing potential
amendments to the CARD Act itself, this Article’s data may prove useful to the Consumer Financial Protection Bureau as it considers how to regulate student credit cards.\textsuperscript{14} The Bureau’s architect initially proposed restricting all marketing to college students,\textsuperscript{15} and this Article can inform the discussion of that suggestion. Finally, this Article contributes to the academic debate about student credit cards. Before this Article, the social science literature on student credit cards had only documented the effects of credit education as a means of affecting student credit card behavior.\textsuperscript{16} This Article adds to that literature by studying the effectiveness of the CARD Act as an example of legal intervention into the student credit card market.

Part II outlines my empirical approach, discussing how I conducted my two-year survey and my study of college–issuer agreements. I present information about the nature and limitations of my survey and study as well as some background information from the findings of these projects.

Part III uses the results of these efforts to assess the rationales that proponents of the CARD Act offered in its support. In describing the CARD Act’s young consumer provisions, I use existing empirical and theoretical research to explain why proponents of the Act believed its young consumer provisions were important. Then, using the data from my survey of students and study of college–issuer agreements, I evaluate those justifications, finding some of them sound and others, including the most important justification for the Act, deeply flawed.

In Part IV, I offer an empirical measurement of the effects of each of the CARD Act’s young consumer provisions. For the ability-to-repay provision, I describe how the loopholes Regulation Z created around the ability-to-pay requirement have engulfed the rule. The survey data reveal the extent to which


\textsuperscript{15} See Elizabeth Warren, \textit{Unsafe at Any Rate}, DEMOCRACY, Summer 2007, at 8, 18 (predicting the Bureau would discourage “marketing targeted at college students or people under age 21”).

\textsuperscript{16} See, e.g., Troy Adams & Monique Moore, \textit{High-Risk Health and Credit Behavior Among 18- to 25-Year-Old College Students}, 56 J. Am. C. HEALTH 101, 101 (2007) (discussing the factors that various social science studies have researched and linked to credit card usage among students).
students are using other forms of debt, such as student loans, to qualify for credit card debt. For the marketing provisions, I explain that the number of students reporting instances of credit card marketing remains high even after the Act’s effective date, but I use the data obtained over the two years I conducted the surveys to illustrate how the Act appears to be having its intended effect of decreasing marketing efforts aimed at students. Finally, the study of college–issuer agreements reveals that the Act’s disclosure requirements have had little effect on the relationships between issuers and college-related organizations.

In Part V, I conclude by suggesting lessons that the CARD Act offers to regulators who are crafting consumer credit regulation.


This Part describes the novel approaches I took to understand how the CARD Act is affecting students and the relationship between credit card issuers and colleges or college-related organizations. In addition to laying out the methodology I followed and the limitations of my approaches, I also describe some of the background findings that inform the remainder of the Article.

A. College Student Survey Methodology

To obtain information from students about their experiences with credit card marketing, I surveyed 527 students at two different universities over the course of two years. The bulk of the students were undergraduate students at the University of Houston. The University of Houston is a large, urban public school. In November 2010, I surveyed 338 students in three

17. All of the surveys were conducted under the approval of the University of Houston’s Institutional Review Board.

18. See Univ. of Hous., UH at a Glance, http://www.uh.edu/about/uh-glance (last visited Sept. 24, 2012) (“Founded in 1927, the University of Houston is the leading public research university in the vibrant international city of Houston. Each year, we educate more than 39,800 students in more than 300
different history classes. In November 2011, I changed the survey instrument slightly to reflect a new year and surveyed 79 students in another history class.\textsuperscript{19} In addition to these students at the University of Houston, in January 2012, I also surveyed students at Baylor University, a private, religiously affiliated university located in Waco, Texas.\textsuperscript{20} At Baylor, I surveyed 110 students in an introductory geology class.

In all of the classes, the response rate was very high with a large majority filling out the surveys. I calculated the exact response rate in two classes by comparing the number of people marked present in the class and the number of surveys I received back. In both of these classes, the response rate was close to 90%. I estimate a similar response rate in the other classes. While it may be ideal to have an exact response rate in every class, similar surveys of students relating to credit card use often do not report any response rates at all,\textsuperscript{21} so the reported rate of responses goes beyond the standard reflected in other studies.

\textsuperscript{19} The updated survey is presented in Appendix A.

\textsuperscript{20} See Baylor Univ., \textit{Get to Know Us}, http://www.baylor.edu/about (last visited Sept. 24, 2012) ("Baylor University in Waco, Texas, is a private Baptist university . . . \[\textit{w}ith more than 15,000 students working toward degrees in 151 areas of study . . .\") (on file with the Washington and Lee Law Review).

\textsuperscript{21} See, e.g., Sheri Lokken Worthy et al., \textit{Sensation-Seeking, Risk-Taking, and Problematic Financial Behaviors of College Students}, 31 J. FAM. ECON. ISSUES 161, 165 (2010) (discussing the survey questions, the survey process, and the respondents but not the response rate); Jill M. Norvilitis & Michael G. MacLean, \textit{The Role of Parents in College Students' Financial Behaviors and Attitudes}, 31 J. ECON. PSYCHOL. 55, 57 (2010) (discussing the survey method, including the composition of the body of respondents, but not mentioning the response rate); Jill M. Norvilitis et al., \textit{Personality Factors, Money Attitudes, Financial Knowledge, and Credit-Card Debt in College Students}, 36 J. APPLIED SOC. PSYCHOL. 1395, 1402 (2006) [hereinafter \textit{Personality Factors} (stating that the refusal rate is unknown and the response rate therefore cannot be reported confidently); Jill M. Norvilitis et al., \textit{Factors Influencing Levels of Credit-Card Debt in College Students}, 33 J. APPLIED SOC. PSYCHOL. 935, 938 (2003) [hereinafter \textit{Factors Influencing Debt Levels}] (explaining that it is impossible to know how many refused to take the survey form but reporting the response rate of those who actually took a survey form); James A. Roberts & Eli Jones, \textit{Money Attitudes, Credit Card Use, and Compulsive Buying Among American College Students}, 35 J. CONSUMER AFF. 213, 222 (2001) (discussing the study and the sample set but not the response rate).
The high level of responses should dispel any concerns about self-selection bias.\textsuperscript{22} Other studies like this one that stated a response rate reported much lower rates than the rate in this Article’s study.\textsuperscript{23}

When compared to other surveys studying students and credit cards, this study generally has a higher number of subjects.\textsuperscript{24} Additionally, many other similar studies only survey

\textsuperscript{22} See Michael E. Staten & John M. Barron, \textit{College Student Credit Card Usage} 11 (Georgetown Univ. McDonough Sch. of Bus. Credit Research Ctr., Working Paper No. 65, 2002), available at http://faculty.msb.edu/prog/CRC/pdf/WP65.pdf (criticizing two studies about student card use because they did not report response rates and, therefore, it was impossible to determine the level of self-selection bias).


\textsuperscript{24} See Emma Davies & Stephen E.G. Lea, \textit{Student Attitudes to Student Debt}, 16 J. ECON. PSYCHOL. 663, 667 (1995) (pulling data from a survey of 140 students); Hayhoe et al., \textit{Differences in Spending Habits and Credit Use of College Students}, supra note 23, at 118 (using a sample of 480 students); Hayhoe et al., \textit{Discriminating the Number of Credit Cards Held by Students Using Credit and Money Attitudes}, supra note 23, at 648–49 (reporting survey responses from 426 students and using 359 surveys for analysis); Jeff Joireman et al., \textit{Concern with Immediate Consequences Magnifies the Impact of Compulsive Buying Tendencies on College Students’ Credit Card Debt}, 44 J. CONSUMER AFF. 155, 162 (2010) (surveying 249 students); So-hyun Joo et al., \textit{Credit Card Attitudes and Behaviors of College Students}, 37 C. STUDENT J. 405, 406 (2003) (using data from 242 surveys); Ali Kara et al., \textit{Credit Card Development Strategies for the Youth Market: The Use of Cojoint Analysis}, 12 INTERNATIONAL J. OF BANK MARKETING 30 (1994) (using a sample of 229 surveys); Lokken Worthy et al, supra note 21, at 165 (using data from 450 students); Phylis M. Mansfield et al., \textit{Self-Control and Credit-Card Use Among College Students}, 92 PSYCHOL. REP. 1067, 1072 (2003) (analyzing 165 surveys); Markovich & DeVaney, supra note 23, at 62 (using a sample of 236 surveys); Norvilitis et al., \textit{Personality Factors}, supra note 21, at 1400 (surveying 448 students); Norvilitis & MacLean, supra note 21, at 57 (surveying 173 students); Palan et al., supra note 23, at 86 (analyzing a sample of 260 surveys); Roberts &
students at a single school, so including a public and a private school suggest the results of this survey have a greater potential to be more representative, although the results are geographically located within a single state.

All of the information from the surveys was entered into and analyzed using Stata software. The students in the sample ranged from freshmen to students who had been in college for more than four years. Table 1 provides details of the sample.

Table 1: Sample Demographic Information

<table>
<thead>
<tr>
<th>Years in School</th>
<th>Univ. of Houston</th>
<th>Baylor Univ.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49.16%</td>
<td>58.18%</td>
<td>51.04%</td>
</tr>
<tr>
<td>2</td>
<td>25.18%</td>
<td>25.45%</td>
<td>25.24%</td>
</tr>
<tr>
<td>3</td>
<td>13.43%</td>
<td>9.09%</td>
<td>12.52%</td>
</tr>
<tr>
<td>4</td>
<td>6.47%</td>
<td>6.36%</td>
<td>6.45%</td>
</tr>
<tr>
<td>More than 4</td>
<td>5.76%</td>
<td>.91%</td>
<td>4.74%</td>
</tr>
</tbody>
</table>

Jones, supra note 21, at 222 (using a sample of 406 college students).

25. See Davies & Lea, supra note 24, at 667 (University of Exeter students); Joo et al., supra note 24, at 406 (students from “the College of Human Sciences of one large university in a southwestern state”); Mansfield et al., supra note 24, at 1072 (students of “a public college in the northeastern United States”); Markovich & DeVaney, supra note 23, at 62 (Purdue University students); Norvilitis & MacLean, supra note 21, at 57 (students of “a medium-sized state university in the United States”); Palan et al., supra note 23, at 86 (students at “a major public [Midwestern university”); Robb & Sharpe, supra note 23, at 29 (students at “a large Midwestern university in the United States”); Roberts & Jones, supra note 21, at 222 (students of “a private university with an enrollment of 13,000 students in Texas”).


27. The percentages do not add up to 100% because the responses of 1.52% of students were either missing or impossible to interpret.
The distribution of men and women roughly approximates the ratios at the University of Houston, but women are overrepresented in the sample from Baylor University. Similarly, in the sample from Baylor, Non-Hispanic White students are slightly overrepresented, and Asians and Non-Hispanic African Americans/Black students are slightly underrepresented in the sample. In the University of Houston

28. See Univ. of Hous., Facts and Figures, http://www.uh.edu/about/uh-glance/facts-figures/index.php#distribution (last visited Sept. 24, 2012) (providing figures from which one can calculate that 50.17% of students at the University of Houston are men and 49.83% are women) (on file with the Washington and Lee Law Review).

29. See Baylor Univ., Profile of Undergraduate Students Fall 2010 and Fall 2011 2 (Sept. 9, 2011), http://www.baylor.edu/content/services/document.php/151566.pdf (reporting 42.1% of Baylor's fall 2011 students are men and 57.9% are women).

30. See id., at 3 (stating that 7.9% of Baylor’s fall 2011 students are African American, 7.9% are Asian, 13.6% are Latino, 65.6% are white, and 5% are other (including Alaskan Native/American Indian, Pacific Islander, Multiracial, and
sample, Non-Hispanic African Americans/Black students are slightly overrepresented, and Non-Hispanic White students are slightly underrepresented. These differences in racial background are likely of no significance because existing research indicates race does not affect rates of credit card ownership. Research has also found that attitudes toward credit are not affected by gender. The samples were purposefully skewed to include more freshmen and sophomores than the general university populations. Figure 1 compares the racial backgrounds of the sample and general student populations.

Figure 1: Comparison of Racial Backgrounds of Sample Groups and Actual Populations

As with any study, this type of survey-based study has several limitations. First, the data are all based on answers from

Not Specified/Unknown)).

31. See Univ. of Hous., supra note 28 (stating that 12.1% of the University of Houston’s 2011 students are African American, 19.3% are Asian American, 23.5% are Hispanic, 33.1% are White/Other, and 12% are Hawaiian/Pacific Islander, International, Multiracial, Native American, or Unknown).

32. See Robb & Sharpe, supra note 23, at 26 (citing evidence to support the assertion that “there is little difference in terms of credit card ownership based on college students’ ethnicity”).

33. See Joo et al., supra note 24, at 415 (finding that gender was not a significant factor affecting credit attitudes).
the students, and I could not and did not undertake any steps to
verify that the responses were true. Some researchers contend
that students may underreport levels of credit card use or debt
because of the social desirability bias;34 others note that self-
reports are not longitudinal so they do not account for credit card
debts that have been paid off by student loans;35 and others claim
students may simply misremember information about their
experiences with credit cards.36 Alternative approaches, however,
would be very expensive, so surveys offer a plausible method for
capturing information about students’ experiences with credit
card marketing.37 Almost every study of student credit cards
employs this strategy, so this limitation comports with
established standards.38

Second, the samples are not nationally representative, so I
can only make claims about the universities I studied. This
limitation is also present in virtually all studies of student
cards,39 and nothing indicates that these particular public and
private schools are atypical.

34. See Mansfield et al., supra note 24, at 1076–77 (acknowledging and
discussing the “potential social desirability bias associated with the balances
reported”); Wayne Jekot, Note, Over the Limit: The Case for Increased
Regulation of Credit Cards for College Students, 5 CONN. PUB. INT. L.J. 109, 112
(2005) (suggesting that self-reporting could produce inaccurate results “because
respondents may incorrectly report unflattering data” (citation omitted)).

35. See Robert D. Manning & Ray Kirshak, Credit Cards on Campus:
Academic Inquiry, Objective Empiricism, or Advocacy Research?, 35 J. STUDENT
FIN. AID 39, 45 (2005) (stating that “credit card debt statistics tend to be
underestimated by respondents and do not include past credit card debts that
were paid with student loans, family loans, or other bank consolidation loans”).

36. See U.S. GEN. ACCOUNTING OFFICE, GAO-01-773, CONSUMER FINANCE:
COLLEGE STUDENTS AND CREDIT CARDS 16 (2001) (noting that reliance on
memory is a limitation on the accuracy of reports that are based on student-
reported information).

37. See Michael E. Staten & John M. Barron, Usage of Credit Cards
Received Through College-Marketing Programs, 34 J. STUDENT FIN. AID, no. 3,
2004 at 7, 20–21 (criticizing data based on self-reporting but noting that “survey
responses are a unique source of information on such questions as how and
when students first receive their credit cards and their general attitudes toward
card usage”).

38. See supra notes 21–25 and accompanying text (providing numerous
examples of studies using self-reporting).

39. See supra note 25 and accompanying text (providing numerous
examples of studies that drew their entire sample from a single school).
Finally, the sample was a purposive sample, so it is non-random. This approach, however, was necessary to obtain a higher number of responses from freshmen, sophomores, and students under the age of twenty-one. The limitations of this purposive sample were mitigated by selecting classes that were part of the general degree requirements and by surveying at both public and private schools.40

B. College–Card Issuer Agreement Study Methodology

The study of college–issuer agreements made use of agreements that the CARD Act compelled issuers to disclose.41 The Federal Reserve Board has posted all of these agreements on the Internet and has published reports about some aspects of them.42 My goal in the study of college–issuer agreements was to code information about a representative sample of these agreements and to determine what changes occurred within agreements after the CARD Act went into effect.

In 2009, there were 1,044 agreements between credit card issuers and universities or related organizations.43 To obtain a representative sample of these agreements, I exceeded established precision levels where the confidence level is 95% and P = 0.5 by evaluating 300 agreements.44 To ensure that I sampled

40. See Palmer et al., supra note 26, at 111 (noting that a purposive sample, rather than a random sample, is a drawback, but that such a sample was necessary to ensure better response rates and less bias, and the selection of both public and private schools and students with different majors attempted to minimize the harm).
41. See infra Part III.C (discussing these agreements in greater detail).
44. See Glenn D. Israel, Sampling the Evidence of Extension Program Impact (2009), http://edis.ifas.ufl.edu/pd005 (last visited Sept. 24, 2012)
a random collection of the 1,044 agreements when I selected the 300 to review, I used a web-based True Random Number Generator to generate a list of numbers between 1 and 1,044 by using atmospheric noise to produce the results. Thus, the data discussed in this Article are representative of the entire universe of agreements between college-related entities and card issuers.

Three research assistants obtained and entered information about the college–issuer agreements. I developed a written protocol that they followed after receiving training. After the results were entered and the study was complete, I reviewed the data for anomalies.

For each agreement, we obtained thirty different data points. We pulled statistical data from the Federal Reserve’s compilation of information about the agreements, such as the annual payments by the issuer and the number of accounts opened. We then obtained information about whether the entities were part of public or private institutions and the precise types of association. The most significant coding work involved reading the agreements and recording information about (1) the obligations of the collegiate entities under the agreement, such as requirements to provide mailing lists, to exclusively promote the issuer, and to provide advertising help to the issuer; (2) the rights of collegiate entities, such as the right to approve advertisements and the right to royalties; (3) the terms of the credit cards issued pursuant to the agreement, including whether they had annual fees and how much interest is charged; and finally, (4) any changes in the agreement between 2009 and 2010, the period during which the CARD Act went into effect. All of this

THE CARD ACT ON CAMPUS

information was entered into a custom-designed Excel spreadsheet and imported into Stata for analysis.

Most of the agreements came from a single issuer, a characteristic that also dominates the aggregated data reported by the Federal Reserve. Table 2 breaks the sample down by the credit card issuer.

Table 2: Issuers in Sample

<table>
<thead>
<tr>
<th>Issuer</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIA Card Services, N.A.</td>
<td>86.67%</td>
<td>260</td>
</tr>
<tr>
<td>U.S. Bank National Association ND</td>
<td>5.33%</td>
<td>16</td>
</tr>
<tr>
<td>Chase Bank USA, N.A.</td>
<td>3.00%</td>
<td>9</td>
</tr>
<tr>
<td>Pennsylvania State Employees Credit Union</td>
<td>3.00%</td>
<td>3</td>
</tr>
<tr>
<td>UMB Bank, N.A.</td>
<td>3.00%</td>
<td>3</td>
</tr>
<tr>
<td>INTRUST Bank, N.A.</td>
<td>3.00%</td>
<td>3</td>
</tr>
<tr>
<td>GE Money Bank</td>
<td>0.33%</td>
<td>1</td>
</tr>
<tr>
<td>USAA Savings Bank</td>
<td>0.33%</td>
<td>1</td>
</tr>
<tr>
<td>First National Bank of Omaha</td>
<td>0.33%</td>
<td>1</td>
</tr>
<tr>
<td>Barclays Bank Delaware</td>
<td>0.33%</td>
<td>1</td>
</tr>
<tr>
<td>Capital One Bank (USA), N.A.</td>
<td>0.33%</td>
<td>1</td>
</tr>
<tr>
<td>Commerce Bank, N.A.</td>
<td>0.33%</td>
<td>1</td>
</tr>
</tbody>
</table>

In terms of the university-related organizations with whom the issuer contracted, Figure 2 depicts that 37.67% of the agreements were between credit card issuers and undergraduate colleges, 32.67% were with alumni associations, 7.33% were with foundations, 2.00% were with professional schools, and 1.67% were with alumni associations and universities together. Entities that did not fall within one of the other categories made up
18.67%. Of the agreements with undergraduate colleges, 17.70% of the institutions were public and 82.30% were private.

Figure 2: Types of Institutions with Issuer Agreements

The agreements describe the interest rates for a variety of types of credit accounts. For the basic credit card in each agreement, the rates ranged from 6.15% to 19.9%, with a median rate of 13.15%. The agreements also contain information about royalties provided to the college-related organization for different aspects of the credit card accounts, such as royalties for each account opened, royalties for each annual fee paid, and royalties for accounts remaining open at the end of the year. Account-opening royalties ranged from $0.80 to $50 for each account opened, with a median royalty of $1. Annual-fee royalties ranged from $1 to $20 for each annual fee paid, with a median royalty of $1. Remaining-open royalties ranged from $1 to $6.10 for each

---

46. This information is provided in the Federal Reserve’s spreadsheet aggregating data disclosed by issuers. See FRB 2009 Full Data Spreadsheet, supra note 43 (providing the information about each agreement).

47. We obtained this information by doing Internet searches about each of the undergraduate institutions.

48. To determine the interest rates for agreements with multiple possible rates for the basic card, I always picked the highest listed rate. Twenty-one percent (n=63) of the agreements did not include an interest rate in the agreement.
account remaining open at the end of the year, with a median royalty of $1.

The vast majority of the agreements, 97.67%, required the college to promote exclusively the issuer’s credit card. The agreement between MBNA America Bank, N.A. (MBNA America) and Alabama State University (ASU) provides a good example of such a provision:

ASU agrees that during the term of this Agreement it will endorse the Program exclusively and that neither ASU nor any ASU Affiliate shall, by itself or in conjunction with others, directly or indirectly: (i) sponsor, advertise, aid, develop, market, solicit proposals for programs offering, or discuss with any organization (other than MBNA America) the providing of, any Financial Service Products of any organization other than MBNA America; (ii) license or allow others to license the Trademarks in relation to or for promoting any Financial Service Products of any entity other than MBNA America; and (iii) sell, rent or otherwise make available or allow others to sell, rent or otherwise make available any of its mailing lists or information about any current or potential Members in relation to or for promoting any Financial Service Products of any entity other than MBNA America. Notwithstanding anything else in this Agreement to the contrary, ASU may accept print advertising from any financial institution provided that the advertisement does not contain an express or implied endorsement by ASU of said financial institution or the advertised Financial Service Product.

In addition to these terms, which serve as a backdrop for understanding the arrangement between issuers and colleges, the other terms of the agreements are discussed at length in Parts III and IV.

One limitation of my approach is that interrater reliability was not assessed. But, because the data we gathered was based on relatively objective criteria, interrater reliability should not be a significant factor in the validity of the study.

49. Six agreements lacked any term about exclusivity, and one agreement explicitly stated the agreement was not exclusive.

III. Understanding the CARD Act’s Young Consumer Provisions and Assessing the Act’s Rationales

This Part briefly introduces the changes the CARD Act made to the laws governing young consumer and student credit cards. For each change, I outline the most significant theoretical and empirical academic research that animated the changes, and I describe the arguments made in the U.S. House of Representatives and Senate in support of the law. In Part III.C, I use the findings of my study of college–issuer agreements to assess the rationales that proponents of the disclosure provision offered in support of it. Some of the concerns policymakers and academics had about these agreements were proven to be accurate by my study, such as concerns that these agreements were aimed at students specifically and that they require college-related organizations to provide private information to issuers and provide forums to market to students. Other rationales, however, including one of the most prominent, related to the extent to which these agreements engender high credit card utilization, appear to have been based on faulty, albeit understandable, predictions about what these previously secret agreements contained.

The Part concludes by applying the findings of my student survey to dispute the central rationale for the CARD Act—excessive student credit card indebtedness. I argue that a pivotal series of studies about student cards conducted by Sallie Mae over the last decade have been repeatedly misused by many academics offering student-card policy prescriptions and by members of Congress who promoted the young consumer provisions of the CARD Act. This argument is not intended to imply that the Act should not have been passed or should be repealed. Rather, it is meant as a call to policymakers and academics to establish an accurate view of student credit card usage and debt levels.

In its provisions on young consumers, the CARD Act makes four changes to the Truth in Lending Act (TILA)51 and one

change to the Fair Credit Reporting Act (FCRA).\textsuperscript{52} These statutory changes are clarified by administrative rules in Regulation Z.

\textbf{A. Ability to Pay}

The Act’s most substantial change to the TILA is its requirement that companies evaluate young consumers’ ability to repay debts incurred before extending credit to them.\textsuperscript{53} Consumers can demonstrate an ability to repay either by getting a cosigner who can repay or by showing an “independent means of repaying any obligation arising from the proposed extension of credit.”\textsuperscript{54}

Although this appears to establish a strict standard, Regulation Z’s implementation of the provision reveals otherwise. First, under Regulation Z, young consumers who are applying for themselves must only have the ability to repay the minimum balance due each month on the account, not the outstanding balance.\textsuperscript{55} Paying one’s minimum balance does little to extricate most people from their debt because the minimum balance is a small fraction of the overall debt owed.\textsuperscript{56} Second, the Federal Reserve has been clear that students can use any income or assets to show an ability to pay the minimum balance. In explaining why it rejected suggestions to limit the income a student can rely on to show earned income, the Federal Reserve stated that it believed a lower standard “will provide sufficient protection for consumers less than twenty-one years old without


\textsuperscript{53} See 15 U.S.C. § 1637(c)(8) (2012) (establishing requirements that a consumer under age twenty-one must satisfy before being given an open end consumer credit plan).

\textsuperscript{54} Id.

\textsuperscript{55} See 12 C.F.R. § 226.51(a) (2012) (stating that the issuer must consider the ability of the consumer “to make the required minimum periodic payments under the terms of the account based on the consumer’s income or assets and current obligations”).

\textsuperscript{56} See Julia Lane, Note, \textit{Will Credit Cardholders Default over Minimum Payment Hikes?}, 18 LOY. CONSUMER L. REV. 331, 346 (2006) (stating that the minimum monthly payment is generally 2\% or 3\% of the total balance on a credit card).
unnecessarily impinging on their ability to obtain credit and build a credit history."\(^{57}\)

Regulation Z also clarified the statute’s provision on cosigners. Cosigners can be either primarily liable on the account or serve as guarantors.\(^{58}\) It states that authorized users are not covered by the statute.\(^{59}\) Although several commentators have argued that students need credit cards for purchases such as airplane tickets, Regulation Z provides young users an easy way to reap many of the benefits of having a credit card by allowing students to be authorized users. Finally, Regulation Z explains that the cosigner’s liability can terminate at age twenty-one for all debt incurred after the young consumer turns twenty-one.\(^{60}\) In addition to these changes in Regulation Z, TILA itself was amended to require that cosigners agree in writing to any increases in a young consumers’ credit limit.\(^{61}\)

The central rationale for requiring that students demonstrate an ability to repay their debts was a concern that students were amassing substantial debt that had negative consequences for their own lives and for society. As Part III.D discusses in detail, members of Congress and academics repeatedly cited high debt loans and high degrees of credit card use as perverse outcomes from lax credit standards for young consumers.\(^{62}\) Several high profile, tragic instances of students committing suicide because of debt fueled alarm about mounting


\(^{58}\) See 12 C.F.R. § 226.51 (2012) (stating that cosigners can either be jointly liable or secondarily liable).

\(^{59}\) See id. (stating that the statute does not apply to individuals who are under twenty-one and who are added to the account of another).

\(^{60}\) See id. (allowing an issuer to provide that a cosigner will not be liable for debts incurred by the consumer incurred after the consumer reaches the age of twenty-one).

\(^{61}\) See 15 U.S.C. § 1637(p) (2012) (stating that a credit limit cannot be raised on the cosigned account of an individual under twenty-one, unless the cosigner approves the increase in writing and accepts the joint liability for the additional amount); 12 C.F.R. § 226.51(b)(2) (2012) (stating that, for an individual under twenty-one who has a cosigned credit card account, the credit limit cannot be raised before the individual reaches age twenty-one “unless the cosigner, guarantor, or joint accountholder who assumed liability at account opening agrees in writing to assume liability on the increase”).

\(^{62}\) See infra notes 113–39 and accompanying text.
debts. Academics posited that excessive debt prevented graduates from getting loans and sometimes jobs, and that it causes great stress and poor financial well-being. One school official stated that his school lost more students because of excessive indebtedness than any other reason. These concerns are particularly acute as other means for financing education provide the most aid to the richest students. Part III.D assesses this rationale.

In addition to concern about student debt loads, the Act’s provisions relating to cosigners may be a response to parents’ complaints about harassment from creditors even when the parents did not cosign for the debt. Members of Congress expressed concern that parents ended up being unofficially liable for their children’s debt when credit card companies allowed students to be overextended.

63. See Kimberly M. Gartner & Elizabeth R. Schiltz, What’s Your Score? Educating College Students About Credit Card Debt, 24 ST. LOUIS U. PUB. L. REV. 401, 401–02 (2005) (“Observers have expressed concern about burgeoning credit card debt loads which, when combined with already-high student loan burdens, can force students into quitting college, declaring bankruptcy, and even, in a few tragic cases, suicide.” (citations omitted)); Jekot, supra note 34, at 110–11 (giving examples of students who committed suicide because of credit card debt).

64. See MANN, supra note 8, at 158 (“News reports explain, for example, that high credit card debt by recent graduates often inhibits their ability to obtain credit (for car loans or the like) and in some instances even impairs their employability.” (citation omitted)).

65. See Norvilitis et al., Personality Factors, supra note 21, at 1396 (“High levels of debt are related to a decreased sense of ability to manage one’s money and lower self-esteem, as well as a decreased sense of financial well-being and higher levels of overall stress.” (citations omitted)).

66. See Small CLAIMS, 13 COM. L. BULL., Nov.–Dec. 1998, at 6, 7 (“The Chicago Tribune quoted Indiana University administrator John Simpson: ‘This is a terrible thing. We lose more students to credit card debt than academic failure.’”).


68. See MANNING, supra note 8, at 168 (“Second, and more disconcerting, were the harassment and even lawsuits against parents of students in default on their credit cards—even if they had not cosigned the loan agreement. Significantly, both of these practices persist and are major complaints of students and their parents.” (citations omitted)).

69. See, e.g., 155 CONG. REC. S5488 (daily ed. May 14, 2009) (statement of
B. Restrictions on Marketing to Young Consumers

In addition to general concerns about students being unable to pay their credit card debts, the CARD Act also responded to problems that members of Congress observed about how credit cards were being marketed to students. The Act contains provisions about sending credit card offers to students and handing out tangible gifts on college campuses.

1. Prescreened Mail Offers

Through an amendment to the FCRA, the CARD Act attempts to discourage credit card companies from mailing young consumers credit card offers by forbidding credit reporting agencies from providing issuers credit reports for people under twenty-one, unless the young consumer consents. The Act does not directly forbid sending credit card offers, but instead it attempts to stop the practice indirectly by choking off a source of information for credit card companies.

Before the CARD Act, academic research had established that college students frequently received credit card offers. Indeed, the “preferred marketing technique for potential customers was direct mail.” One study found that 69% of students surveyed reported receiving a credit card offer in the mail in the prior week; another claimed that students receive

---

Sen. Claire McCaskill (“They send these cards to kids because they know their parents, if they are in college, don’t want them to get into trouble and they will bail them out if they get in too deep.”).

70. See 15 U.S.C. § 1681b(c)(1)(B)(iv) (2012) stating that, for transactions not initiated by the consumer, a consumer reporting agency cannot furnish a consumer report for use in extending credit or insurance if the report shows the consumer is under twenty-one, unless the consumer consents).

71. See U.S. GEN. ACCOUNTING OFFICE, supra note 36, at 6.

72. See Norvilitis et al., Factors Influencing Debt Levels, supra note 21, at 941 (“In the week prior to the survey, 69% . . . of students received at least one credit-card offer.”). Other studies found lower levels, such as one study’s finding that only 37% of student cardholders received their applications in the mail. See Jacquelyn Warwick & Phylis Mansfield, Credit Card Consumers: College Students’ Knowledge and Attitude, 17 J. CONSUMER MARKETING 617, 621 (2000) (finding that 37% of respondents with credit cards received the application through direct mail).
twenty-five to fifty card solicitations a semester. High credit card utilization was directly caused, studies reported, by aggressive marketing: “The majority of college students who own credit cards do not actively seek them out, but are aggressively pursued through the mail and on-campus by credit card issuers.” Another study found:

Financially at-risk students are more likely than other students to acquire their credit card(s) through a mail application, at a retail store, and/or at a campus table. These findings suggest that aggressive marketing practices by credit card companies to target college students (i.e., mass mailings, retail store discounts, and credit card representatives on campus) have likely contributed to the recent rise in credit card debt on college campuses putting some students at more financial risk than others.

Credit card companies have a strong incentive to capture the student credit card market because students tend to continue using the account they opened in college, and academics have raised the concern that allowing students to have credit cards normalizes and routinizes paying with credit.

Members of Congress were outraged that young consumers received credit card offers in the mail. For instance, Senator Menendez pointed out that he knew a two-year-old child who had received an offer for a credit card and that his own children received an “incredible number of preapproved credit cards.”

---

73. See Wood, supra note 9, at 163 (“This heavy marketing is demonstrated by the twenty-five to fifty credit card solicitations students receive per semester.” (citation omitted)).

74. Warwick & Mansfield, supra note 72, at 623.

75. Lyons, supra note 23, at 73.

76. See U.S. GEN. ACCOUNTING OFFICE, supra note 36, at 4, 35 (stating that some issuers “marketed to college students because they viewed them as good customers who would continue using the issuers' credit cards in a responsible way” and that many college students will earn higher incomes and be profitable credit card customers after they graduate).

77. See MANN, supra note 8, at 45–49 (describing the psychology of payment with credit cards as compared with other payment methods).

78. See 155 CONG. REC. S5410 (daily ed. May 13, 2009) (statement of Sen. Robert Menendez) (recalling seeing his children, “when they were in college and studying but not working, get an incredible number of preapproved credit cards” and mentioning his “State director's 2-year-old who got a preapproved credit card”); see also 155 CONG. REC. S5548 (daily ed. May 18, 2009) (statement of Sen. Byron Dorgan) (criticizing companies for offering cards to very young children).
Similarly, a Representative reported his thirteen-year-old son had received credit card offers.79

2. Tangible Gifts on Campus

In addition to changing the FCRA, Congress amended the TILA to forbid issuers from offering tangible gifts to students on or near campus or at student events in exchange for filling out a credit card application.80 The Federal Reserve has offered a variety of clarifications to this simple rule, explaining that on or near campus means within 1,000 feet of the campus; that a “tangible item includes any physical item” but not “non-physical inducements such as discounts, rewards points, or promotional credit terms;” that issuers can give tangible gifts as long as they also give them to those not filling out applications;81 and that the prohibition applies to consumers under twenty-one and those over twenty-one if they are students.82

Academics have expressed concern that colleges have permitted and even endorsed credit card marketing.83 One study demonstrated that students who obtained a credit card through

79. See 155 CONG. REC. H4964 (daily ed. Apr. 29, 2009) (statement of Rep. Keith Ellison) (“Let me say that I knew that we had a problem in America when my 19-year-old son . . . kept getting solicitations for credit cards; but I was quite convinced . . . when my 13-year-old son . . . started getting credit card solicitations.”).

80. See 15 U.S.C. § 1650f(2) (2011) (prohibiting card issuers and creditors from offering to “a student at an institution of higher education any tangible item to induce such student to apply for or participate in an open end consumer credit plan,” whether on campus, near campus, or at a school-sponsored event).

81. See 12 C.F.R. § 226.57(c) (2012) (clarifying the terms “tangible item,” “inducement,” “near campus,” and “related event,” requiring that the creditor take steps to determine whether someone is a student, and making clear that mailings are included in the prohibition).

82. See Truth in Lending, 75 Fed. Reg. 7658, 7756 (Feb. 22, 2010) (to be codified at 12 C.F.R. pt. 226) (stating that the definition of college student “is intended to be broad and would apply to students of any age attending an institution of higher education and applies to all students, including those enrolled in graduate programs or joint degree programs”).

83. See MANN, supra note 8, at 157 (“[I]t is plain that in many cases the marketing proceeds with the approval of the university administrators, who voluntarily permit issuers to implement card-issuance programs directly on university campuses.” (citation omitted)).
on-campus marketing had higher debt-to-income ratios and that students often believed that their college had screened creditors who were allowed to market on campus.84 Academics argued that young consumers are more responsive to truthful-but-incomplete advertising85 and that college credit card marketing tactics overshadowed the TILA disclosures.86

Members of Congress have echoed the fears of academics, stating that gifts preyed on “vulnerable” college students87 and that issuers aggressively preyed on students.88 One member of Congress went even further than researchers and claimed that

84. See Norvilitis et al., Factors Influencing Debt Levels, supra note 21, at 941 (finding that students who received credit cards from the student union had higher debt-to-income ratios than those who got credit cards elsewhere, and that most students believed the school evaluated companies soliciting students in the union).

85. See Laurie A. Lucas, Integrative Social Contracts Theory: Ethical Implications of Marketing Credit Cards to U.S. College Students, 38 AM. BUS. L.J. 413, 422–23 (2001) (discussing what qualifies as deceptive advertising and arguing that college students “lack sophistication and therefore deserve special protection in relation to credit”).

[Additionally], most of the concern about college credit cards is not about credit terms that rise to this level of deception or unfairness. Rather, the concern is about offering credit to people who might not understand the dangers of such credit at a time in their lives when they are unlikely to currently have sufficient income to keep the debt from escalating at high interest rates.

Gartner & Schiltz, supra note 63, at 410.

86. See Lucas, supra note 85, at 414–15 (describing an increased emphasis on promotional disclosures instead of on TILA disclosures).

[T]he specific practice of target marketing to U.S. college students using credit card solicitations . . . de-emphasize[s] the disclosures required under the Truth in Lending Act (TILA). Many such solicitations instead emphasize other promotional materials—like celebrity endorsements or offers of prizes, gifts or discounts—and have reduced the size of the required disclosures, or included them in inserts, in order to fit the promotional material in the text of the solicitation . . . .

87. See 155 CONG. REC. H5011 (daily ed. Apr. 30, 2009) (statement of Rep. Steve Cohen) (“College students are most vulnerable and shouldn’t be lured to credit cards at an early age and put into even more debt than student loans do by offering prizes and gifts.”).

88. See 155 CONG. REC. E1033 (daily ed. Apr. 30, 2009) (statement of Rep. John Lewis) (“Credit card companies aggressively prey on our young college students who are not yet working. These companies rove college campuses and entice students with gifts, with the intent of collecting interest payments as the student ravel[s] herself in debt.”).
the marketing techniques were deceptive.  

Like the ability-to-repay requirement, however, the fundamental concern about marketing both through mailed offers and campus advertising was the high level of student debt that these practices ultimately created.  

C. College–Issuer Marketing Agreements

In addition to restrictions on who can obtain credit cards and how issuers can market those cards, the CARD Act also took aim at the relationship between credit card companies and colleges and organizations related to colleges. This subpart describes this part of the CARD Act and uses the study of college–issuer agreements to empirically evaluate the justifications offered for it.

1. The Provision and Its Rationale

The CARD Act requires institutions of higher education to “publicly disclose any contract or other agreement made with a card issuer or creditor for the purpose of marketing a credit card.”  

In addition to colleges publicly disclosing these agreements, credit card companies are required to provide to Congress any agreements they have with colleges.  

This obligation requires disclosure of agreements beyond just those that market cards to young consumers, as long as students are possible targets. The Federal Reserve has clarified: “An agreement may qualify as a college credit card agreement even if marketing of cards under the agreement is targeted at alumni,  

---


90. See Nelson, Young Consumer Protection, supra note 9, at 375 (describing the CARD Act as a recent effort by lawmakers “to address young consumers’ escalating indebtedness”); id. at 396–97 (stating a great concern of lawmakers enacting the CARD Act was the “detrimental consequences due to . . . accumulation of credit card debt” by college students).


92. See id. § 1637(r)(2) (requiring each creditor to submit an annual report describing all college agreements and explaining the details of such reports).
THE CARD ACT ON CAMPUS

faculty, staff, and other nonstudent consumers, as long as cards may also be issued to students in connection with the agreement.93

Several academics have argued that agreements between colleges and credit card companies have engendered students’ debt problems. Because of the financial incentives credit card issuers offer to schools and university officials, academics have argued that college administrations were willing to lead their students into debt to capture the issuers’ incentives.94 As Robert Manning argues:

This Faustian pact includes sponsoring school programs, funding student activities, renting on-campus solicitation tables, and paying “kickbacks” for exclusive marketing agreements such as college or alumni affinity credit cards. As a result, rather than protecting the economic and educational interests of their students, college administrators are playing an active and often disingenuous role in promoting the societal acceptance of consumer debt as well as the prominence of credit cards in college life.95

From the college’s perspective, some researchers argued, it is better for students to be in debt.96 Similarly, members of Congress believed that the relationship between credit card companies and universities led to perverse incentives to facilitate debt.97

Members of Congress also emphasized the importance of this provision to provide “transparency in university marketing deals with credit card issuers.”98 Senator Feinstein went further,
arguing that requiring transparency may “act as a deterrent to
deals with highly unfavorable terms for students.”99 Part IV.C uses my study of 300 of these college–issuer agreements to evaluate whether this prediction materialized. The CARD Act’s disclosure requirement is a significant change because attempts to obtain information about these agreements were stymied in the past because the agreements forbid the parties from disclosing their terms.100

2. Assessing the Rationale for the College–Issuer Disclosures

Because the CARD Act requires issuers and college-related entities to disclose their agreements, we now have the information needed to see whether this disclosure requirement was justified in the first place. The information we obtained from our sample of 300 agreements suggests that some of the concerns animating the disclosure requirement were justified. On the other hand, our findings suggest that other concerns appear to be overstated.

First, concerns about college-related entities promoting student use of credit cards are well-founded. The agreements envision, for the most part, students obtaining credit cards because of the agreements. Of all the agreements, 72.67% include student cards, while the remaining 27.33% are aimed exclusively at alumni or other groups.

In addition, the agreements create an easy mechanism for issuers to use to reach students. Many of the agreements, 68.33%, require that the college-related entity provide a list of mailing addresses for students. This percentage is significantly higher than information reported before the CARD Act disclosures.101 An

---

Dianne Feinstein); see also 155 Cong. Rec. E1035 (daily ed. Apr. 30, 2009) (statement of Rep. Thomas Petri) (“Despite the fact that hundreds of schools throughout the country have such arrangements, very little is known about them . . . . This bill simply seeks greater transparency by requiring credit card companies to report these arrangements.”).


100. See U.S. Gen. Accounting Office, supra note 36, at 30 (describing how attempts to uncover information about credit card college–issuer agreements failed because alumni association officials stated that “their contracts with the credit card issuers precluded disclosure of the terms and conditions”).

101. See Cheryl Hystad & Brad Heavner, Graduating Into Debt: Credit
agreement between Dickinson College (DC) and MBNA America provides a common, albeit circuitous, provision. It states: “Upon the request of MBNA America, DC shall provide MBNA America with mailing lists free of any charge.” Mailing lists are defined as “updated and current lists and/or magnetic tapes (in a format designated by MBNA America) containing names, postal addresses and, when available, telephone numbers of members segmented by zip codes or reasonably selected membership characteristics.” The definition of “members” explicitly includes students: “Member means undergraduate students, graduate students, alumni of Dickinson College and/or other potential participants mutually agreed to by DC and MBNA America.” Thus, while the agreement does not come out and say so, it requires the college to provide students’ addresses to the credit card issuer.

In addition to student mailing lists, the agreements provide issuers with other advertising rights. Around half of the agreements, 47.33% (n=142), did not list any specific advertising the entity would provide or participate in. For the other half, Table 3 outlines the specific advertising arrangements between issuers and college-related entities.

CARD MARKETING ON MARYLAND COLLEGE CAMPUSES 8–9 (2004), available at http://www.marylandconsumers.org/LinkClick.aspx?fileticket=x4_rQermn7Y%3d&tabid=72 (reporting two of twelve schools interviewed admitted they sold student information to credit card companies and three of twelve schools surveyed sold a student list in some form).


103. Id. § 1(e).

104. Id. § 1(f). In the other 31.67% of the agreements, the addresses on mailing lists are limited to nonstudents. Alabama State University’s (ASU’s) agreement, for instance, states:

ASU shall provide the initial mailing list, containing at least thirty thousand (30,000) non-duplicate alumni names (of persons at least eighteen years of age) as well as additional names of donors and parents of students, with corresponding valid postal addresses and, when available, telephone numbers and e-mail addresses of Alumni Members as soon as possible but no later than thirty (30) days after ASU’s execution of this Agreement.

Alabama State University Affinity Agreement, supra note 50, § 2(e).
Table 3: Advertising Arrangements Between Issuers and College-Related Entities

<table>
<thead>
<tr>
<th></th>
<th>Percentage of Agreements with Advertising Details(^{105})</th>
<th>Number of Agreements (n=158)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issuer permitted to advertise on entity’s website</td>
<td>90.51%</td>
<td>143</td>
</tr>
<tr>
<td>Issuer permitted to solicit customers and/or have advertisements at sporting events or other major events</td>
<td>29.11%</td>
<td>46</td>
</tr>
<tr>
<td>College-related entity will send e-mails recommending the issuer</td>
<td>8.86%</td>
<td>14</td>
</tr>
<tr>
<td>College-related entity will include credit card applications in organization magazines, newspapers or e-newsletters</td>
<td>3.80%</td>
<td>6</td>
</tr>
<tr>
<td>Issuer will provide credit education on campus (e.g., in student welcome kits, at orientation events, in the student newspaper or in the campus book store)</td>
<td>2.53%</td>
<td>4</td>
</tr>
</tbody>
</table>

In addition to the arrangements in Table 3, two agreements stated that the college-related entity would place banner advertisements for issuers, two stated that the issuer would be promoted in materials at the alumni office or at alumni meetings, and one stated that the issuer could place information in store publications. Yale University’s agreement with Chase Bank USA provides an example of the two most common provisions:

Yale shall prominently place a jpeg image with an associated hyperlink above the fold on the homepage, and shall use

---

\(^{105}\) These percentages add up to more than 100% because some agreements provided issuers with multiple advertising rights.
reasonable efforts to obtain placement on the checkout or point-of-sale pages, if any, of the Association of Yale Alumni (AYA) Web site (www.aya.yale.edu), and shall prominently place a link on the sponsor page of the Yale Athletics Web site . . . .106

Consistent with Schedule 3(a), Yale shall also provide or cause to be provided to Chase, at no cost to Chase, with access to each Yale home athletic event identified on Schedule 3(a) to market the Program . . . . Yale shall provide a location that is prominent with respect to visibility and pedestrian foot traffic.107

Based on the fact that most agreements are aimed at putting credit cards in the hands of students and that most agreements actively involve the school in distributing the means for advertising those cards, policymakers’ concerns about the entanglement of college-related entities and credit card issuers appear justified.

Yet, in some ways, the agreements are not as problematic as people imagined. First, the terms outlined in the agreements do not have the most abusive characteristics critics associate with credit cards. For instance, none of the 300 agreements we reviewed created cards with teaser rates, a common credit card snare that consumer advocates and academics criticize.108 Additionally, the rates established by the agreement are not extremely high considering the nonexistent credit histories of many students. The median rate in our sample was 13.15%, but most credit cards have much higher effective rates, especially for poor credit risks.109

Second, the agreements do, for the most part, give the school the right to approve of any advertising the issuer does pursuant to the agreement. Again, Alabama State University’s affinity

---


107. Id. § 4(c).


109. See MANN, supra note 8, at 190 (stating that higher-risk, higher-default-rate borrowers often have higher interest rates); MANNING, supra note 8, at 218 (citing some examples of credit card interest rates, including 19.8 APR and 21.9 APR).
agreement provides a common example: “ASU shall have the right of prior approval of all program advertising and solicitation materials to be used by MBNA America, which contain ASU’s trademark; such approval shall not be unreasonably withheld or delayed.” Most of the agreements, 96.00% (n=288), contained a provision giving the school-related entity the right to approve ads, while the remaining 12 agreements simply did not address approval of advertising. As long as college-related entities exercise strong judgment in approving advertisements, these provisions provide a check on abusive marketing behavior.

Third, and most importantly, the relatively small amount of money paid to each college-related entity undermines one of the key rationales behind requiring disclosures by colleges. Prior research, which did not have the benefit of the disclosures required by the CARD Act, appears to have overstated the extent to which colleges have benefitted from marketing agreements. Based on the Federal Reserve System’s aggregation of the data provided by issuers in 2009, 604 of the college-related entities, or 57.85%, made less than $10,000 under their agreements with issuers, with 219 making less than $1,000 and 99 making no money at all. The median payment amount was $5,891. Thus, for most organizations, their agreement with the issuer had a negligible effect on their bottom line. If it is true that credit card debt causes students to withdraw from school and cease paying tuition, it seems most schools have a lot more to lose if students are over-indebted than they have to gain by encouraging students to use credit cards.

For a small minority of entities, however, the agreements were lucrative. In 2009, 143 entities made more than $100,000 from their arrangement with issuers, and 25 entities obtained even more than $1,000,000. For these schools, it appears there may be an incentive to encourage credit card use. Overall,

---

110. Alabama State University Affinity Agreement, supra note 50, § 2(d).

111. The figures in the remainder of this section are all based on my analysis of the Federal Reserve’s spreadsheet. See FRB 2009 Full Data Spreadsheet, supra note 43.

112. See Small CLAIMS, supra note 66, at 7 (“This is a terrible thing. We lose more students to credit card debt than academic failure.” (quoting an Indiana University administrator)).
however, the data from the Federal Reserve indicate that the link between college–issuer agreements and over-indebtedness is not as clear as prior research had supposed.

Along these same lines, it appears that most agreements did not result in a substantial number of credit card accounts being opened. Of all the agreements, 87.36% (n=912 of 1044) of the agreements resulted in fewer than 100 new credit card accounts being opened. The median number of cards opened pursuant to an agreement was 14. These data demonstrate that the idea that these agreements were causing students at most schools to open accounts and take on excessive debt is not true. The next section takes up the other key rationale behind the CARD Act—excessive student credit card debt.

D. Misplaced Reliance on Credit Card Usage and Debt Levels

The primary motivating factor behind each of the young consumer provisions was the belief that students were incurring substantial debt loads that caused them to experience financial distress. A large part of the basis for this concern is a series of studies\textsuperscript{113} by Sallie Mae, a financial services organization focused on education,\textsuperscript{114} and Nellie Mae, a Sallie Mae subsidiary, which

\textsuperscript{113} Nellie Mae began producing reports in 1998, but then Sallie Mae took over. Together, they have produced five reports. For the most recent report, see Sallie Mae, How Undergraduate Students Use Credit Cards: Sallie Mae’s National Study of Usage Rates and Trends 2009 available at http://www.salliemae.com/NR/rdonlyres/0BD600F1-9377-46EA-AB1F-6061FC763246/10744/SLMCreditCardUsageStudy41309FINAL2.pdf.

\textsuperscript{114} See id. at 2 (describing Sallie Mae); see also Sallie Mae, Corporate Overview, https://www1.salliemae.com/about/corp_leadership (last visited Sept. 24, 2012) (“Sallie Mae (NASDAQ: SLM) is the nation’s No. 1 financial services company specializing in education . . . . Sallie Mae turns education dreams into reality for its 25 million customers.”) (on file with the Washington and Lee Law Review).

Sallie Mae is the nation’s leading provider of saving- and paying-for-college programs. The company manages $180 billion in educational loans and serves 10 million student and parent customers. Through its . . . affiliates, the company also manages more than $17.5 billion in 529 college-savings plans, and is a major, private source of college funding contributions in America with 10 million members and more than $475 million in member rewards.

focuses on student loans. This subpart outlines the findings of the Sallie Mae/Nellie Mae studies, focusing on the most recent study, and explores how academics, the press, and policymakers commonly misused those findings.

The concern over the Sallie Mae/Nellie Mae studies might appear parochial or merely interesting to academics only, but it is not. Every day when a member of Congress argued for the CARD Act’s young consumer provisions, they appealed to the figures in these studies (except one instance when the young consumer provisions were just mentioned in passing). It is hard to overstate the extent to which these reports have been misused.

Although there are some differences, the most recent Sallie Mae study is similar in most ways to the prior reports from Nellie Mae. In the most recent study, Sallie Mae pulled data from 1,200 credit bureau reports of students who had applied for private student loans with Nellie Mae or Sallie Mae. The Sallie Mae study finds that many students who apply for private loans have credit cards and that many of these students have high debt loads:

115. See SALLIE MAE STUDY, supra note 113, at 2 (“Since 1982, Nellie Mae has focused exclusively on providing education financing for undergraduate and graduate students and families, through the Federal Family Education Loan Program and through privately funded loans .... Nellie Mae is a wholly-owned subsidiary of SLM Corporation, commonly known as Sallie Mae.”); see also Nellie Mae, About Us, http://nelliemae.com/aboutus (last visited Sept. 24, 2009) (providing background on the company) (on file with the Washington and Lee Law Review).


118. See SALLIE MAE STUDY, supra note 113, at 4 (explaining the differences between the current and prior reports).

119. See id. at 19 (explaining the study’s methodology, including the sample group).
Eighty-four percent of this student population overall have credit cards, an increase of approximately 11 percent since the fall of 2004, the last time the undergraduate study was conducted. Data collected in March 2008 show that the average (mean) amount of debt carried by undergraduate student cardholders increased from 2004 by 46 percent to $3,173. During the same time period, median debt increased by 74 percent to $1,645. The average number of cards carried per cardholder, those carrying four or more cards, and those with balances in the $3,000 to $7,000 range also increased. Based on these numbers, the study concludes that “[i]n this time of credit crunch and economic downturn, college students are relying on credit cards more than ever before.”

There are two problems with how the Nellie Mae/Sallie Mae data are commonly used, the first of which has been suggested by other researchers and the second of which I raise here for the first time. First, as others have noted, the Nellie Mae/Sallie Mae studies only reflect a small, unique group of students, not college students generally. The studies used the credit reports of students who applied for private student loans, not even the government-subsidized loans to which most students turn first to finance their education. It is not a stretch to think that this group would have higher debt loads than the general population of college students because students apply for private student loans when government-subsidized loans are insufficient. A
sample were applying for a special type of student loan because they did not qualify for more conventional student loans due to either excessive debt or incomes that exceeded qualifying thresholds.

126. See SALLIE MAE & GALLUP, supra note 124, at 42.
127. Id.
128. Id.
129. See id. at 26 tbl.2c.
130. See e-mail from Patricia Christel, Vice President, Corporate Commc’ns, Sallie Mae, to Jim Hawkins, Assistant Professor of Law, Univ. of Hous. Law Ctr. (Jan. 3, 2011) (explaining the process by which the study’s credit card debt figures are taken directly from students’ credit reports and stating that these figures are generally higher than the debt levels self-reported by students) (on file with the Washington and Lee Law Review).
The exact effect of including cosigner debts is not known, but we do know that some parents deal with student credit cards by adding students to the parents’ accounts.\textsuperscript{131} If a parent with a $15,000 balance on a credit card adds a new student to the account, most people would not consider the student’s credit card debt load to be $15,000, but that is how the Nellie Mae/Sallie Mae studies characterize the debt. Like using a subset of college students who applied for private loans, this inclusion of cosigner debts has the potential to artificially inflate student debt levels in the studies.

The surveys I conducted at the University of Houston and Baylor University yielded very different results than the Nellie Mae/Sallie Mae studies. My study did not attempt to be nationally representative,\textsuperscript{132} so it also does not offer definitive proof about levels of credit card use or debt among all college students.\textsuperscript{133} Still, the fact that my study, which drew from all undergraduate students, generated such divergent results indicates that the driving factor behind Sallie Mae’s high credit card use and debt levels is a limited sample. Figure 3 contrasts the number of students with a credit card in the Sallie Mae study with the Houston/Baylor surveys.


\textsuperscript{132} See supra Part II.A (describing the study’s sample).

\textsuperscript{133} In another nationally representative study, however, Troy Adams and Monique Moore found that “[o]nly 8.2% and 5% [of more than 45,000 students] had a credit card balance of $1,000 to $2,999 or a balance of $3,000 or more, respectively.” Adams & Moore, supra note 16, at 103.
Similarly, Table 4 compares Sallie Mae’s data about debt levels with the data from the Houston/Baylor study.

Table 4: Comparison of Credit Card Debt Levels Between Sallie Mae and Houston/Baylor Studies

<table>
<thead>
<tr>
<th></th>
<th>Sallie Mae 134</th>
<th>Totals from Houston/Baylor Survey</th>
<th>Univ. of Houston</th>
<th>Baylor Univ.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Freshman Credit Card Debt</strong></td>
<td>Median debt was $939</td>
<td>91.04% had $0–$500 in debt</td>
<td>92.20% had $0–$500 in debt</td>
<td>90.47% had $0–$500 in debt</td>
</tr>
<tr>
<td><strong>Average Debt of Seniors Who Carry Credit Card</strong></td>
<td>$4,100 (n=27) had $0–$500 in debt and</td>
<td>43.14% (n=22) had $0–$500 in debt and</td>
<td>62.50% (n=5) had $0–$500 in debt and</td>
<td>25.00% (n=2) had over $3,000 in debt</td>
</tr>
<tr>
<td></td>
<td>15.25% (n=9) had over $3,000 in debt</td>
<td>13.73% (n=7) had over $3,000 in debt</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These results are consistent with the observation that students at private schools are more likely to have credit cards

134. The data in this table has been compiled from the Sallie Mae report. See SALLIE MAE STUDY, supra note 113, at 3, 8 (providing the median freshman credit card debt and the average debt of graduating seniors who are cardholders).
and higher debt loads. But, in every case, my survey found significantly lower debt levels than Sallie Mae's report.

Academics, the media, and consumer advocates have repeatedly misused the data when making arguments in favor of specific regulatory reforms. More alarming, however, is the fact that every time a member of Congress mentioned the Sallie Mae study in conjunction with arguments for the CARD Act's young consumer provisions, the member inappropriately used the findings as representative of all college students.

This chronic misuse of the Nellie Mae/Sallie Mae studies needs to be corrected to ensure that optimal student-card policies are enacted. By pointing out the inaccuracies in the ways that academics and legislators currently use the studies, I do not intend to show that student credit cards are benign. Indeed, I have argued at length elsewhere that credit cards cause financial

135. See Manning & Kirshak, supra note 35, at 41 (“The highest proportions [of students with credit cards] are at more affluent, private universities and the lowest in predominantly minority colleges and public universities that enroll high percentages of students from lower income households.”). But see Norvilitis et al., Personality Factors, supra note 21, at 1404 (reporting students in that study at state schools had higher debt loads than students at private schools).

136. To locate instances where members of Congress used the Nellie Mae/Sallie Mae studies in conjunction with the CARD Act, I searched Westlaw's “Congressional Record” database using the follow search command on January 21, 2012: “(sallie-mae or nellie-mae) /50 credit-card". It generated twenty-six results, but only five results involved the CARD Act; the other twenty-one were about other issues. In each case, the member of Congress misrepresented the studies' findings. See 155 CONG. REC. H5814 (daily ed. May 20, 2009) (statement of Rep. Earl Blumenauer) (“A recent Sallie Mae survey indicated that 84% of undergraduates had at least one credit card and that, on average, students have 4.6 credit cards.”); 155 CONG. REC. S5493 (daily ed. May 14, 2009) (statements of Sen. Dianne Feinstein) (stating a number of figures from the Sallie Mae report as though they represent all students and using them as evidence of a need for Congress to take action); 155 CONG. REC. S5316 (daily ed. May 11, 2009) (statement of Sen. Christopher Dodd) (“According to Sallie Mae, college students graduate with an average credit card debt of more than $4,000. That is up from $2,900 just 4 years ago. Nearly 20 percent of college students have credit card balances of over $7,000.”); 155 CONG. REC. H5020 (daily ed. Apr. 30, 2009) (statement of Rep. Louise Slaughter) (stating a number of Sallie Mae statistics as though they represent all students and rationalizing credit card debt as a reason for an increase in bankruptcy filings among young people); 155 CONG. REC. E1026 (daily ed. Apr. 30, 2009) (statement of Rep. Patrick Murphy) (blaming college credit card agreements for students’ “racking up debts that can take years to pay off,” and citing a “recent Sallie Mae study” as proof that graduating seniors have, on average, “more than $4,100 in credit card debt”).
Yet, if we really want to protect students who use credit cards, we need to use the information appropriately. Also, if legislators have incorrect beliefs about the sources of student debt problems, they may enact policies aimed at the wrong credit vehicles. For instance, perhaps if legislators had better data on the true amount of credit card debt, they would focus more attention on reforming student loans. Additionally, with a better understanding of debt loads, policymakers might be able to determine the average optimal amount of credit card debt for students and enact rules that limit balances at those levels.

IV. Measuring the Effectiveness of the CARD Act’s Young Consumer Provisions

The prior Part explored the pre-CARD Act era—the reasons for the CARD Act’s young consumer provisions, the flaws in those rationales, and the provisions themselves. This Part looks at the world after the CARD Act and reports data that measure the effects of the CARD Act for each of its provisions.

A. The Effects of the Ability-to-Pay Provision

The provisions requiring credit card companies to evaluate young consumers’ abilities to repay their debt have generated a variety of predictions. Some people have claimed that the CARD Act will eliminate access to credit because its ability-to-repay

---

137. See generally Jim Hawkins, Regulating on the Fringe: Reexamining the Link Between Fringe Banking and Financial Distress, 86 Ind. L.J. 1361 (2011) (surveying the evidence that links credit cards and financial distress).

138. See Tamar Lewin, Student Loan Default Rates Rise Sharply in Past Year, N.Y. Times, Sept. 12, 2011, at A14 (detailing an increase in student loan default rates from 7% to 8.8% in a single year).

139. Congresswoman Louise Slaughte r offered an amendment to the CARD Act with caps on the amount of debt students could accumulate, but these amendments did not make it into the final Act. See Bill Swindell, House Close To Passage Of Maloney Credit Card Measure, Nat’l Js. Congress Daily, Apr. 30, 2009, at 6 (describing Maloney’s and Slaughter’s proposed amendments, and explaining that Slaughter’s amendment would set standards for underwriting student cards, including limiting the available credit to the greater of 20% of income or $500 per month).
standard is too strict. Others, however, have asserted that the Act will not stop anyone who wants a credit card from obtaining one because the standards are too lenient. This section analyzes these and other predictions with the findings of my student surveys. My research resolves this apparent contradiction in the current literature on ability to repay by revealing how the CARD Act’s standards appear strict but actually contain many loopholes that students have discovered and exploited. I discovered that students have found creative ways to avoid the strictures of the independent-ability-to-pay provisions, although they have not enlisted peers as cosigners as many observers expected.

According to the survey results, not many students under the age of twenty-one had applied for a card since the beginning of the school year, creating a relatively small sample size in this limited demographic. Of the total sample of students under twenty-one, 11.05% (n=44) had applied for a card in the semester before the survey was taken. Three students erroneously failed to complete the detailed questions about their application, leaving forty-one students. Of those forty-one students, 56.09% (n=23) applied by themselves, and 43.90% (n=18) applied with a cosigner. Although the sample size is small, the results are interesting and touch on many of the empirical questions debated among the CARD Act’s opponents and proponents.

1. Qualifying as an Individual

Scholars raised a variety of concerns about loopholes around the provisions requiring young consumers to show an

140. See Schwartz, supra note 9, at 424 (“The upshot is that independently wealthy eighteen-year-olds, or those whose parents are willing and able to accept joint liability, will still be able to obtain a credit card. But poor and middle-income applicants may not.” (citation omitted)); Wood, supra note 9, at 172–74 (arguing that the standards for ability to repay are more stringent on those under twenty-one than on those over twenty-one and, therefore, this will prevent young consumers from building credit histories).

141. See Nelson, Young Consumer Protection, supra note 9, at 402–03 (“Considering the likelihood that young consumers will have little trouble meeting the CARD Act’s eligibility requirements as interpreted by the Federal Reserve Board, the door remains open for college-aged consumers to continue amassing significant amounts of debt.”).
independent ability to repay the debt if they apply by themselves, and responses to my survey indicate those concerns were valid. Part of the problem is that the concept of a student’s “income” eludes simple definition. 142 For instance, several sources have suggested the possibility that young consumers could use loan proceeds as income to obtain a credit card, although none of these sources offer evidence of this phenomenon. 143 In my survey, I found that 27.27% (n=6) of students under twenty-one who were applying by themselves listed loans as part of their income to qualify for the credit card. If we also include students over twenty-one, 30.56% (n=11) listed loan proceeds as part of their income. While the number of students in this category is small, this finding demonstrates that concerns about using one type of debt to qualify for another type of debt are plausible.

Another problem I discovered through the surveys was the extent to which students used money from their family as their income or assets to qualify for a credit card. Of the students under twenty-one who applied by themselves, 34.78% (n=8) listed money from relatives as income. For those who hoped the CARD Act would ensure young consumers could pay their debts on their own, this number is troubling.

Finally, in general, students did not use earned wages as income as often as proponents of the Act had hoped. Of the students under twenty-one applying for a card on their own, 68.42% (n=13) 144 reported having income below $10,000 a year, and only 45.45% (n=10) listed income as the sole means for obtaining a card. More surprising, only 52.27% (n=12) stated that they used earned income at all to qualify for the card, with the remainder relying on other sources.

142. See Joo et al., supra note 24, at 418 (“College students’ income is hard to measure because the definition of income varies from student to student.”).

143. See Nelson, Schoolhouse to Poorhouse, supra note 9, at 28 (“Moreover, if this incoming student takes out student loans to fund his or her educational expenses, these loans can be treated as ‘income’ to independently qualify for a card—a disconcerting practice that some student consumers have already begun to implement.” (citation omitted)); Susan Tompor, Credit Card Offers Still Contain Trouble Spots for Consumers, DETROIT FREE PRESS, Sept. 30, 2010, at B4 (noting that “students [are] reporting a college loan as ‘income’ and some card issuers [are] accepting that claim”).

144. Some respondents did not answer this question, which changes the number and percentage ratios.
Overall, these statistics paint a disturbing picture of the effectiveness of the ability-to-pay provision. The lenient requirements set up through Regulation Z have provided numerous ways to qualify for credit outside of actually being able to repay the debt.

2. Cosigning

Several news stories, academic articles, and a participant at an FDIC advisory meeting have raised the concern that the CARD Act would cause students to seek peer cosigners, but no data exist to confirm or refute this apprehension. Unlike fears about the inroads around the income or asset requirement, this fear appears misplaced, at least among the students I surveyed. Of the students under twenty-one who applied for a new card within the few months before completing the survey, 94.44% (n=17) used a parent as their cosigner.

145. See David Migoya, College Students Duck New Credit-Card Law with Friends, LOWELL SUN (MASS.), Sept. 8, 2010 (discussing college students’ cosigning for one another in order to get credit cards after the new law); see also John C. Ninfo II, Commentary: An 18-Year-Old Needs a Credit Card?, DAILY REC. (ROCHESTER, N.Y.), Jan. 26, 2010 (“Another concern is that parents, family members or friends will not stop and think twice before co-signing for a credit card for a young person who cannot meet the ‘independent means’ test . . . .”); Tompor, supra note 143, at B4 (“[S]ome college students who are 18 or 19 are asking friends twenty-one or older to co-sign their credit card applications.”).

146. See Nelson, Schoolhouse to Poorhouse, supra note 9, at 32 (“[O]lder students, who may already have student loan and/or credit card debt, are permitted to sign if they meet the issuers’ requirements. There have already been reports of some college students paying older students and friends to serve as cosigners.” (citations omitted)); Schwartz, supra note 9, at 427 (“Some eighteen- to twenty-year-old students . . . simply ‘ask classmates or fraternity brothers to co-sign’ their credit card application, ‘sometimes for a small fee.’” (citations omitted)).

147. See Ted Beck, President & CEO, Nat’l Endowment for Fin. Educ., statement at the Federal Deposit Insurance Corporation Meeting of the Advisory Committee on Economic Inclusion to Discuss Children’s Savings and Underserved Studies (Nov. 16, 2010) (“[W]e just did a survey, and 61 percent of parents don’t want to co-sign their credit card. So what we’re finding—and I don’t have a statistic for this—but young adults are going to their friends and saying, ‘Would you co-sign for me?’ who are over 21.”).

148. See Migoya, supra note 145 (“There are few hard numbers on how the trend is developing, but enough anecdotal evidence that it’s beginning to creep upward.”).
cosigner, while one student had a sibling cosign. No students under the age of twenty-one reported having a friend or someone other than a parent or sibling cosign for them.

One benefit of the CARD Act’s qualifications for young consumers is that it may have helped formalize the surety relationship that existed implicitly between many parents and their children. Of the students under twenty-one that I surveyed, 31.33% (n=104) expected someone else, most likely their parents, to pay their credit card debt. Prior to the CARD Act’s requirements about young consumers having cosigners, parents may have been informally drafted into the debtor-creditor relationship. Indeed, several members of Congress expressed this concern when debating the CARD Act’s young-consumer provisions,149 and one academic has criticized the practice of credit card companies “exploiting familial ties to reach into the pockets of those with whom there is no formal contract.”150

Findings in a study conducted by the Education Resources Institution and Institute for Higher Education Policy showed that 63% of students obtained their first card without a cosigner.151 That number is slightly higher than my study, in which 56.09% of students applied on their own, perhaps suggesting that the CARD Act is making cosigning more common. Although it is impossible to draw out any causal inferences with my survey’s data, the high number of parents listed as cosigners suggests that more parents will be formal, not merely informal, guarantors of student debt.

149. For instance, the CARD Act was originally introduced “to prevent credit card issuers from taking unfair advantage of college students and their parents.” 155 Cong. Rec. S174 (daily ed. Jan. 7, 2009) (statement of Sen. Herb Kohl) (emphasis added) (proposing to amend the TILA for the purpose of preventing this exploitation).


151. See EDUC. RES. INST. & INST. HIGHER EDUC. POLICY, CREDIT RISK OR CREDIT WORTHY? COLLEGE STUDENTS AND CREDIT CARDS 9 (1998) (“A majority of students, 36%, received their credit cards by applying on their own.”).
3. Access to Credit

One persistent criticism of the CARD Act’s restriction of access to credit cards has been that it will prevent those under twenty-one from getting credit cards, starting small businesses, establishing a credit history, or having access to an important source of credit. Also, some research suggests that borrowers may turn to alternative financial service providers if credit is unavailable.

In my survey, I asked whether students thought they needed a credit card to make purchases while they were in college. Students were less likely than some academics to think they needed credit cards. For students under twenty-one, only 38.85% (n=155) answered that they thought they needed a credit card. For students over twenty-one, the number jumped to 50.78% (n=65), but the CARD Act’s ability-to-repay requirements do not apply to that latter group.

I did not ask about whether students had turned to fringe bankers for credit, but some evidence indicates that students are turning to these lenders. In general, fringe lenders have gained business because of the tightening of credit and stricter regulations. More pointedly, one fringe lender has stated in a

152. See Brian Burnsed, New Rules Place Barriers Between Students, Credit Card Issuers, U.S. NEWS (Feb. 19, 2010), http://www.usnews.com/education/articles/2010/02/19/new-rules-place-barriers-between-students-credit-card-issuers (last visited Sept. 24, 2012) (“Peter Garuccio, a spokesman for the American Bankers Association, a banking lobby group, says, ‘It’s pretty clear that it will be tougher for people in this group to get credit cards. I think that you’ll probably see a decline in the number of cards in this segment.’”) (on file with the Washington and Lee Law Review).

153. See Schwartz, supra note 9, at 432 (“By categorically withholding credit cards from eighteen- to twenty-year-olds, section 301 seriously impedes their ability to start up a business.”).

154. See Wood, supra note 9, at 175 (stating that young people must build credit before they can buy cars or houses, and therefore, by preventing access to credit cards, the Act “hampers the ability of . . . eighteen- to twenty-one-year-olds to become fully independent adult consumers”).

155. See Williams & Emley, supra note 9, at 1421–22 (noting that the rules could prevent a spouse from opening a credit card because of a lack of personal income or assets, even if the other spouse had substantial income and assets).

156. See Schwartz, supra note 9, at 430–31 (listing possible alternatives to credit cards but also pointing out some problems with them).

157. See Jessica Silver-Greenberg, Payday Lenders Go Hunting, WALL ST. J.,
public securities filing that the CARD Act has increased demand for its product:

In some cases we believe regulatory changes have resulted in a constriction of the availability of unsecured credit for consumers with poor or no credit history (for example, the Card Accountability Responsibility and Disclosure Act of 2009, which, among other things, disallowed the issuance of a credit card to anyone under 21 without a co-signer or proof of ability to repay and also curtailed the amount of fees that banks can assess on cardholders). The Company believes that this constriction in available sources of credit has resulted in, and will continue to result in, an increased demand for our services, which has produced a corresponding growth in our fee and interest income, as well as an increase in our need for employees and opportunities for opening new stores. 158

Whether this is a harmful or salutary development is, of course, highly debatable, but it is worth noting that fringe credit products are generally much less likely than credit cards to cause borrowers to become over-indebted. 159 Thus, if the Act is leading students to alternative financial service providers, it may be doing them a favor.

B. The Effects of Restrictions on Marketing to Young Consumers

As I mentioned in the introduction, legislators are proud of the effects that they think the CARD Act’s provisions on marketing are having. This subpart discusses data from my surveys about offers students had recently received and credit card marketing they had observed. I found that the number of students receiving offers through the mail and being subjected to

---

Dec. 24, 2010 (“Payday lenders like Advance America are pushing hard to lure away customers from traditional banks. The effort is getting a boost from the industry’s loan crunch, especially for borrowers with blemished credit, and toughened regulation of fees and interest rates charged by the nation’s 7,760 banks and savings institutions.”).


159. See Hawkins, supra note 137, at 1399–1402 (arguing that the structure of fringe credit transactions makes them difficult to link to financial distress).
marketing remains high. The number appears, however, to have decreased in the two years since the CARD Act was enacted.

To understand the effects of the CARD Act’s provision on credit card marketing, I did not ask questions that tried to indict companies for breaking the law. For some provisions of the Act, this would be impossible because I could not measure how often, for instance, credit reporting agencies were giving information about young consumers to companies.160 More importantly, the real purpose of the Act was to decrease harmful advertising and student over-indebtedness, so the real measure of its effectiveness requires a larger consideration of its effects than mere compliance or noncompliance with the technical requirements of the law.

1. Prescreened Mail Offers

Commentary on the CARD Act has praised it for “protect[ing] students from insidious pre-screened offers with which they are consistently bombarded.”161 The truth, however, is much more nuanced. While it appears that the number of students reporting credit card offers has dropped, it remains quite high.

I asked students whether they had “received any credit card offers in the mail since the beginning of” either 2010 or 2011, depending on the year I administered the survey. Overall, 68.92% (n=275) of students under twenty-one reported receiving credit card offers in the mail during the preceding year. Thus, a large majority of students were still subjected to marketing through mailed offers, despite the CARD Act going into effect in February 2010. The number of students reporting offers, however, did decrease between 2010 and 2011. Of the students under twenty-one, 76.13% (n=185) reported having received an offer in 2010, but only 57.69% (n=90) indicated they had received an offer in 2011.

The fact that students are still receiving card offers does not necessarily indicate that credit reporting agencies or credit card companies are violating the CARD Act. Instead, credit card

160. For a description of this provision, see supra Part III.B.
161. Wood, supra note 9, at 170.
companies are likely obtaining information for consumers who are under twenty-one from sources other than credit reporting agencies, such as “commercial mailing lists through memberships to music or book clubs, magazine subscriptions, or by completing sweepstakes entry cards.” As the study of agreements between college-related entities and credit card issuers in this Article found, 68.33% of such agreements require that the college-related entity provide the issuer with student addresses. Issuers do not rely exclusively on credit bureaus for student addresses, which undermines the successfulness of the CARD Act’s attempt to cut off offers. The CARD Act’s approach was purposefully indirect, but its circularity appears to have undermined its effectiveness.

2. Marketing on and off Campus and Marketing Using Gifts

To measure the effectiveness of the CARD Act’s provision forbidding issuers from offering tangible gifts to students, I asked about whether students had seen any credit card marketing on campus and off campus and not just about the specific marketing prohibited by the Act of offering tangible gifts. I asked the following questions:

During your time in college, have you seen any credit card companies advertising ON or NEAR campus or at a student event?

During your time in college, have you seen any advertising by credit card companies OFF campus that appears to be directed at college students?

During your time in college, have you seen any credit card companies offering a gift (like a T-shirt or food) if you sign up for a credit card?

I focused on marketing efforts in general because prior research indicates credit cards draw in students for a variety of reasons. In reporting on his extensive groundbreaking qualitative research, Robert Manning describes the reasons his interviewees were attracted to their credit cards. None of his interviewees report that gifts were important to cards’ appeal. Often just the

162. Warwick & Mansfield, supra note 72, at 621.
163. See infra Part IV.C.
advertisements were sufficient, and the motivation for some students was credit issuers' appeals to responsible uses of open-ended credit. Tangible gifts were “unnecessary.” Thus, I wanted to measure the overall effect of the Act on marketing activities, not a single subset of student marketing.

I asked about whether students had seen credit card companies offering tangible gifts at all and not just offering tangible gifts on campus because I wanted to capture what effect the Act was having on tangible gifts whether on or off campus. First, issuers could easily evade regulations without having any meaningful effect on students’ experiences by offering gifts 1,001 feet from campus. Second, issuers can reach students on campus with offers of tangible goods through electronic sources like e-mail or social media. Companies cannot mail offers of tangible goods, but because an e-mail address does not have a physical location, credit card companies can send solicitations offering tangible gifts to students sitting in their college dorm rooms on their school “.edu” e-mail addresses.

164. See MANNING, supra note 8, at 172 (“Jeff’s first credit card was an impulsive response to a Citibank advertisement ‘that was hanging on the wall in the dorm.’”); id. at 178 (“I saw advertisements in the [student] newspaper, sign-up tables [in the student center], and applications [inserted] with my textbooks [from the bookstore].”); id. at 181 (“Citibank Visa advertisements ‘were plastered all over the university . . . .’”).

165. See id. at 175 (“He is most angered about how the credit card companies’ marketing literature on campus praises the benefits of ‘responsible use’ but neglects to inform impressionable and inexperienced students about the downside, such as the impact of poor credit reports on future loans and even potential employment.”).

In order to stretch her limited resources, Kristin decided to get her own credit card, since she no longer had access to her parents’ plastic. The slogan “It pays to Discover” was appealing because the “no annual fee,” “build your own credit history,” and “cash back bonus” features satisfied her need for financial control.

Id. at 188–89.

166. Id. at 190.

167. See 12 C.F.R. § 226.57(c)(3) (2012) (“A location that is within 1,000 feet of the border of the campus of an institution of higher education, as defined by the institution of higher education, is considered near the campus of an institution of higher education.”).

168. See id. § 226.57(c)(4) (clarifying the prohibition on inducement by stating that it includes mailing tangible goods to locations on or near campus).

To measure the effect of the Act, I compared the responses to several questions about credit card marketing from students who had only been in college during the time the CARD Act was in effect with those who had been in college for at least some time before the Act’s effective date. If the CARD Act was being effective, I posited that students who had been in college only during the time in which the Act was in effect should report seeing credit card marketing at a substantially lower rate than students who had been in college both during the time the Act was in effect and the time it was not in effect.

Of students who had only been in college while the CARD Act was in effect, 22.37% (n=68 of 304 students) reported seeing credit card companies marketing on campus, while 49.10% (n=109 of 222 students) of students who had been in school while the Act was not in effect reported seeing on-campus marketing efforts. This result is a statistically significant difference under the chi-squared test. Similarly, 67.21% (n=205 of 305 students) of students who had only been in college under the Act responded that they had seen credit card marketing off campus directed at students, while 81.07% (n=167 of 206) of those in school without the Act had observed this type of marketing. Finally, 40.33% (n=123 of 305 students) of students in school under the Act reported seeing credit card companies giving gifts to students, while 59.71% (n=123 of 206 students) in school without the Act reported this conduct.

These results are summarized in Figure 4:

---

170. For the first year I conducted the survey, I compared freshmen versus all other students, and for the second year, I compared freshmen and sophomores versus all other students since the Act had been in effect for two years.
171. $\chi^2(1, N=524)=41.55, p < 0.001, \text{Cramér's } V=0.28.$
172. $\chi^2(1, N=524)=11.97, p < 0.001, \text{Cramér's } V=0.15.$
173. $\chi^2(1, N=524)=19.50, p < 0.001, \text{Cramér's } V=0.19.$
In each case, it appears that the number of students observing marketing decreased by around 15–20% after the Act. This difference, however, could be attributed to being in school for a longer period of time and thus having more opportunities to see marketing activity than the effectiveness of the CARD Act’s provisions. In the first year I conducted the survey, 43.30% of sophomores reported seeing credit card marketing on campus, while 71.11% of seniors reported seeing credit card marketing on campus. Each of those groups had been in school while the CARD Act was not in effect, yet their answers differed by 30%. The chi-squared test confirms that this difference between sophomores and seniors is unlikely to be a result of chance.\textsuperscript{174} Thus, while the numbers appear to suggest the Act is having an effect, I think that effect is unlikely to be attributable to the Act.

The survey data do indicate, however, a decrease from 2010 and 2011 in the marketing reported by respondents. Of all the freshmen surveyed in the first year I conducted the survey, 32.19% had seen credit card companies marketing on campus, while only 12.20% of freshmen in the second year observed that marketing. The difference between the first and second year is

\begin{equation}
\chi^2(1, N=121)=5.84, \ p < 0.05, \text{ Cramér’s } V=-0.22.
\end{equation}
not attributable to adding a second school to my survey pool because the number of only University of Houston students reporting marketing on campus dropped to 15.25% in the second year of surveying. The number of freshmen reporting off-campus marketing dropped from 73.29% in the first year to 62.60% in the second year (or 64.41% if only University of Houston freshmen are included). Finally, the number of freshmen reporting companies handing out gifts decreased from 47.26% in the first year to 32.52% in the second year (or 25.42% if only University of Houston freshmen are included).

Together with the responses to the questions about mailed offers, I summarize these results in Figure 5:

**Figure 5: Percentage of Freshmen Reporting Credit Card Marketing Over Two Years**

As depicted in Figure 5, it appears that for all categories, credit card marketing directed at students is declining, and this decline is statistically significant for all categories except mailed offers. Thus, while the level of students reporting credit card marketing remains higher than proponents of the CARD Act may have

---

175. The chi-square test yielded the following results: for reports of mailed offers ($X^2(1, N=266)=2.82, p < 0.1, \text{Cramér's V}=0.10$); for reports of gifts ($X^2(1, N=266)=5.7, p < 0.05, \text{Cramér's V}=0.15$); for reports of on-campus marketing ($X^2(1, N=266)=16.74, p < 0.001, \text{Cramér's V}=0.25$); for reports of off-campus marketing ($X^2(1, N=266)=4.49, p < 0.05, \text{Cramér's V}=0.13$).
predicted, the trend of fewer students reporting instances of marketing suggests the Act may be effectively, although slowly, curbing such marketing.

C. The Effects of the College–Issuer Marketing Agreements Provisions

A variety of sources predicted before the Act passed that forcing issuers to disclose their agreements with colleges would affect the terms of those agreements. Others have since claimed that, in response to the Act, issuers are amending their agreements with organizations to remove marketing to students. Based on my study of the agreements between issuers and college-related entities, however, it appears that these predictions are wrong.

The majority of agreements did not change at all between 2009 and 2010, despite the fact that the CARD Act was passed in 2009 and went into effect in February 2010. Of the 300 agreements I studied, 64.33% (n=193) remained exactly the same in 2010 as they were in 2009, and for many agreements, the same as they had been for years before that. Only 1% of the agreements were first signed in 2009 or 2010.

It does appear that more agreements were terminated or expired after the CARD Act went into effect, but it is impossible to tell whether this change is merely correlative or if the CARD

176. See 155 Cong. Rec. S5493 (daily ed. May 14, 2009) (statement of Sen. Dianne Feinstein) (arguing that transparency might “act as a deterrent to deals with highly unfavorable terms to students”); see also Wood, supra note 9, at 171 (“Exposing the agreements will not only increase public awareness about these practices but may also deter the more unconscionable aspects of these agreements.” (citation omitted)).

177. See James Goodman, Credit Card Companies Adapt to College Rules, ROCHESTER DEMOCRAT & CHRON., Nov. 1, 2010 (describing issuers’ new focus on alumni rather than current students).

Given such restrictions, and with student loan debt at an all-time high, credit card companies have shifted their focus to alumni. Betty Riees, Bank of America spokeswoman, said in the past several years the bank has been amending several agreements to eliminate marketing to students. The bank has about 700 “collegiate affinity” agreements, she said, and 98 percent of open accounts are “non-student.”
Act caused companies or college-related entities to end the agreements. For the agreements that were in effect at the start of 2009, the study revealed that 24 of 300 were terminated or expired in that year, but in 2010, 63 were terminated or expired.

The mere fact that the number of terminations increased, however, obviously does not mean the CARD Act caused the change. It appears from the documents associated with the agreements that all of the relationships ended in a normal course of events, either because the agreement was set to expire (18 agreements) or the parties terminated the relationship pursuant to the agreement (45 agreements). Often, the issuer initiated the termination. Many of the termination letters followed this passage, taken from a letter from FIA Card Services to an organization, almost verbatim:

I am writing to inform you that following a comprehensive review of the American Institute of Chemists, Inc. credit card program, FIA Card Services, N.A. (f/k/a MBNA America Bank, N.A.) (“FIA”) has decided to terminate our Amended and Restated Affinity Agreement dated as of July 13, 1994, as the same may have been amended (“Agreement”).

Administrative reviews and decisions such as these seem to have little to do with increased regulation. More than 70% of the agreements that were terminated generated less than $5,000 in 2009 for the college-related entity, indicating a low activity level, and thus low profitability for the issuer.

In only two cases in all of the 300 agreements that I reviewed did I observe any mention of regulations influencing the decision to end the arrangement. In one example from March 2009, before the CARD Act was actually passed, an agent of the Tulane University Alumni Association terminated its agreement with Bank of America, explaining:

We have enjoyed our seven year affiliation with Bank of America and we have been satisfied with our relationship with the bank and especially with you as our account executive. Our termination is rather a sign of the economic times, the

increased projected scrutiny of university affinity programs
and the mission-relatedness of these affinity arrangements."\footnote{179}

Similarly, in November of 2008, just after the CARD Act was first
introduced, an alumni group associated with the University of
Houston indicated that regulatory pressure was influencing its
decision to end the agreement: “We believe this program is
counter to our mission to contact, engage, serve, empower and
acknowledge UHCL alumni. We also made this decision out of
appreciation for the pressures of credit and the frustration of
mailings."\footnote{180} The fact that only two agreements among 300
mention regulatory pressure suggests that based on outward
appearances, the CARD Act has had a negligible effect on the
number of agreement terminations.

Like the number of terminations, changes to the agreements
demonstrate little salutary effects from the CARD Act. Figure 6
summarizes the changes that occurred in the 83 agreements that
changed in 2010.

\footnote{179. Letter from Charlotte B. Travieso, Dir., Office of Alumni Affairs,
Tulane Alumni Ass’n, to Nazanin Rad, Bank of Am. Bus. Dev. (Mar. 9, 2009),
available at http://www.federalreserve.gov/CreditCardAgreementsContent/
CollegeAgreement_25.pdf.}

\footnote{180. Letter from Charity Ellis, Dir. of Alumni & Cmty. Relations, Univ. of
Hous. Clear Lake, to Peggy Fullett, Vice President, Bank of Am. (Nov. 6, 2008),
available at http://www.federalreserve.gov/CreditCardAgreementsContent/
CollegeAgreement_784.pdf.}
Twenty of the 300 agreements were amended in 2010. Of those 20 amendments, 60.00% (n=12) were ministerial or added provisions that likely have little effect on students’ experiences with the agreements, such as establishing a web portal to access accounts, extending the life of the agreement, or making small changes to the royalties. For 25.00% of the amendments (n=5), the changes were in line with the hopes of the CARD Act’s sponsors. For instance, in several agreements, students were omitted from the mailing lists that the college-related entity was obligated to provide or the issuer stopped paying any royalties for student accounts, taking away the incentive for the entity to promote them. For the other 15.00% (n=3) of the amendments, however, provisions were added that contradict the statute’s intent. Several agreements, for example, added an obligation that the college-related entity advertise for the issuer on its website, and others added students to those persons covered by the agreements. Thus, based on the continued stability of the number of college–issuer agreements and the fact that very few agreements changed to protect students from abuses, it appears that the disclosure requirement has failed to meet its goals.
V. Conclusion: Lessons from the CARD Act’s Young Consumer Provisions

The survey of students and study of college–issuer agreements in this Article have suggested that the CARD Act has not quelled marketing to young consumers or ensured that young consumers could repay their debts in the ways that proponents of the Act had hoped. Survey data demonstrate that students are using student loans to obviate the need to prove an ability to repay credit card debt. Responses to the survey also reveal that a high number of students are still receiving credit card offers in the mail and are still observing credit card issuers on and off campus targeting students with marketing, although the numbers appear to be declining. Similarly, requiring the disclosure of agreements between issuers and college-related entities has had almost no effect on either the number of those agreements or the terms of those agreements.

These results are significant if policymakers want more from the CARD Act than a political victory. The empirical work in this Article offers the first measurement of the Act’s actual effects, and the reality is not as rosy of a picture as many of the predictions about the Act had painted. More work needs to be done to correct the inefficiencies in this market.

Future attempts to establish student credit card policies need to adapt based on the lessons learned through the CARD Act. Primarily, several provisions of the CARD Act failed because they did not directly regulate the behavior that concerned policymakers. For instance, the provision forbidding credit bureaus from giving addresses for consumers under the age of twenty-one did not have the desired effect of curtailing credit card offers in the mail because it only addressed the problem tangentially. Instead, if Congress really wants to prevent offers in the mail, it could directly regulate the conduct, like it did to prevent junk faxes.181 If young consumers were given a private

181. See 47 U.S.C. § 227(b)(1)(C) (2011) (making it unlawful “for any person within the United States, or any person outside the United States if the recipient is within the United States—(C) to use any . . . device to send, to a telephone facsimile machine, an unsolicited advertisement” without meeting the narrow exceptions in the statute).
cause of action against issuers who violated this provision, the number of students reporting instances of being mailed credit card offers would likely drop significantly. In the same way, if members of Congress want to alter the terms of agreements between issuers and college-related organizations, they could do so directly, instead of relying on disclosures to incentivize the parties to change the agreements.

A second way the CARD Act was misguided was its failure to appreciate and respond to the business incentive of establishing students as new credit card customers. Credit card companies have an enormous stake in gaining college students as customers. The strength of this incentive causes issuers to seek creative ways for penetrating the student-card market despite new regulations. As we have seen in a variety of markets, creditors are nimble in avoiding unwanted regulation. Because of this, a regulatory strategy that permits legitimate business purposes while minimizing harms to consumers is preferable. In the case of the CARD Act, the misuse of Sallie Mae’s figures on student debt likely led Congress to regulate with a supposed—and inaccurate—harm in mind instead of legitimate harms. Because the empirical work in this Article has removed this barrier, legislators should reconsider amendments to cap total balances on student cards. Such a regulation would allow issuers to pursue student customers and make credit cards available to students without the risk that students will be buried in debt.

182. See, e.g., id. § 227(b)(3) (setting out a private cause of action for violations of the junk fax statute).
183. Cf. MANN, supra note 8, at 154 (suggesting “a ban on marketing directed at minors and college-age persons”).
184. See, e.g., Jill M. Norvilitis & Phillip Santa Maria, Credit Card Debt on College Campuses: Causes, Consequences, and Solutions, 36 C. STUDENT J. 356, 361 (2002) (suggesting “changing how fees are paid to colleges or student organizations” and providing the example that “if student groups received a flat fee for sponsoring a table rather than an amount per completed application, there might be less pressure on students to complete applications”).
185. See MANNING, supra note 8, at 167 (explaining the important role students play in maintaining credit card companies’ market share).
187. See supra Part III.D.
188. See supra note 139 and accompanying text (discussing Rep. Slaughter’s
The harms that financial distress and misguidance cause to young consumers are important and require a regulatory response. In order to shape that response, however, we need to evaluate the empirical claims behind policy prescriptions and learn from the failures of the CARD Act. As policymakers take on the student debt crisis, these lessons can help establish optimal student credit policies.
Appendix A: The CARD Act Student Survey

Survey on the Effects of the CARD Act
Contact: Asst. Professor Jim Hawkins, 713-743-5018

Please circle your answer:

1. How many years have you been attending any college full-time?
   A. This semester is my first year
   B. This is my second year
   C. This is my third year
   D. This is my fourth year
   E. I have been attending college for more than four years

2. Are you under 21?  A. Yes  B. No

3. What is your gender?  A. Male  B. Female

4. What is your race?
   A. Non-Hispanic White
   B. Non-Hispanic Black/African American
   C. Latino
   D. Asian
   E. Other

5. During your time in college, have you seen any credit card companies advertising ON or NEAR campus or at a student event?
   A. Yes
   B. No

6. During your time in college, have you seen any advertising by credit card companies OFF campus that appears to be directed at college students?
   A. Yes
   B. No

---

189. This survey is the version used in fall 2011 and spring 2012. The earlier version of the survey is identical except that the dates are changed and some of the language is aimed only at students at the University of Houston.
7. During your time in college, have you seen any credit card companies offering a gift (like a T-shirt or food) if you sign up for a credit card?
   A. Yes
   B. No

8. Have you received any credit card offers in the mail since the beginning of 2011?
   A. Yes
   B. No

9. How many credit cards do you currently have?
   A. 0
   B. 1
   C. 2
   D. 3
   E. More than 3

10. Approximately how much do you currently owe on your credit cards?
    A. 0–$500
    B. $500–$1,000
    C. $1,000–$2,000
    D. $2,000–$3,000
    E. More than $3,000

11. Do you expect to pay off these balances yourself or do you expect someone else will pay them off (like a parent)?
    A. I expect to pay them off
    B. I expect someone else will pay them off

12. Do you think you need a credit card to make purchases during your time in college?
    A. Yes
    B. No

13. Since you started school this Fall, have you opened a new credit card account?
    A. Yes
    B. No
ONLY ANSWER THE FOLLOWING QUESTIONS IF YOU HAVE OPENED A NEW CREDIT CARD ACCOUNT SINCE STARTING SCHOOL THIS FALL (I.E., YOU ANSWERED YES TO QUESTION 13).

14. Did you apply for the new credit card by yourself or with a cosigner?
   A. By myself
   B. With a co-signer
   C. Not applicable

15. If you had a cosigner, who was your cosigner?
   A. Parent
   B. Spouse
   C. Friend
   D. Sibling
   E. Other
   F. Not applicable

16. If you applied by yourself, what is your approximate annual income?
   A. Less than $10,000 a year
   B. $10,000 - $20,000 a year
   C. $20,000 - $30,000 a year
   D. $30,000 - $40,000 a year
   E. More than $40,000 a year
   F. Not applicable

17. If you applied by yourself, circle all the answers that you used as part of your “income” when applying for the credit card:
   A. Income from a job
   B. Student loan proceeds
   C. Money from parents/family
   D. Other: _______________________________
   E. Not applicable
Congress, the Constitution, and Supreme Court Recusal

Louis J. Virelli III*

Abstract

Recusal is one of the most hotly contested issues facing the Supreme Court. From the wide-ranging debate over Supreme Court recusal, however, a singular theme has emerged: Congress must do more to protect the integrity and legitimacy of the Court by regulating the Justices’ recusal practices. Herein lies the problem. Rather than solve the puzzle of Supreme Court recusal, direct congressional regulation has created an impasse between Congress and the Court that has consequences for the reputation, efficacy, and legitimacy of both Branches. In a precursor to this Article, I recast the issue of Supreme Court recusal as a constitutional question and argued that direct congressional regulation of Supreme Court recusal violates the separation of powers. This Article builds on that prior work and argues that separation of powers principles are critical to understanding and alleviating the inter-branch impasse over recusal. It contends that Congress, rather than the Court, should take the lead in resolving that impasse and that the separation of powers requires Congress to use indirect constitutional mechanisms to do so. Specifically, Congress should repeal the current statutory provision directly regulating Supreme Court recusal and focus instead on more

* Associate Professor of Law, Stetson University College of Law. J.D. and M.S.E., University of Pennsylvania; B.S.E., Duke University. An earlier version of this paper was part of a Hot Topics Panel at the 2012 Annual Meeting of the Association of American Law Schools, and benefitted immensely from the feedback of the other panelists, Sherrilyn Ifill, Dahlia Lithwick, Jeffrey Stempel, and Russell Wheeler, as well as the attendees. I would also like to thank Mike Allen, Bradley Areheart, Brannon Denning, Jamie Fox, Joel Goldstein, Blake Hudson, Meghan Irmler, and Marco Jimenez for their helpful comments and suggestions. Any errors are entirely my own. This research was generously supported by a grant from Stetson University College of Law.
indirect constitutional tools—such as impeachment, procedural reform, judicial confirmation, appropriations, and investigation—to influence the Justices’ recusal practices. This effort to frame the recusal debate within its proper constitutional context permits a more robust and productive dialogue about both the Justices’ recusal practices as well as the broader question of the nature and dynamics of inter-branch relations in our tripartite government.

Table of Contents

I. Introduction ................................................................. 1537

II. The Debate over Recusal at the Court ......................... 1544
   A. Supreme Court Recusal in Action ............................. 1544
      1. Congressional Regulation of Supreme Court Recusal .......... 1545
      2. The Justices’ Recusal Practices ............................. 1547
   B. The Arguments over Reform ...................................... 1551
   C. The Impasse .......................................................... 1556

III. Recusal and the Separation of Powers ......................... 1558
   A. Congress’s Constitutional Responsibilities ..................... 1558
   B. Constitutional Limits on Congressional Regulation ............... 1562
   C. Why is this Congress’s Problem? ............................. 1575

IV. Congress as an Effective Constitutional Actor .............. 1587
   A. Congress’s Indirect Constitutional Tools ..................... 1587
      1. Impeachment ...................................................... 1587
      2. Procedural Reform ............................................. 1590
      3. Judicial Confirmation .......................................... 1594
      4. Appropriations .................................................. 1595
      5. Investigation ..................................................... 1597
   B. Political and Institutional Benefits ............................ 1599
      1. Systemic Benefits .............................................. 1599
      2. Institutional Benefits ........................................ 1602

V. Conclusion ...................................................................... 1605
I. Introduction

Supreme Court recusal—the question of whether an individual Justice may participate in a case—is currently among the most high profile and controversial issues involving the Court. Chief Justice Roberts’s 2011 Year-End Report on the Federal Judiciary (“Year End Report”) focused entirely on Supreme Court ethics and recusal, and was immediately met with sharp criticism by legislators and commentators. Sitting

1. As has become common practice in discussions about the disqualification of judges, the term recusal in this instance will be used interchangeably to include both the terms “disqualification,” which traditionally refers to involuntary removal of a judge from a case, and “recusal,” which is generally limited to a judge’s voluntary decision to withdraw from a case. See Richard E. Flamm, Judicial Disqualification: Recusal and Disqualification of Judges § 1.1, at 4 (Banks & Jordan, 2d ed. 2007) (“In fact, in modern practice ‘disqualification’ and ‘recusal’ are frequently viewed as synonymous, and employed interchangeably.”).


Justices were called to testify before Congress twice in the past year about recusal issues, and more than 135 law professors signed a letter to the House and Senate Judiciary Committees in March of 2011 outlining the need for new recusal legislation to “protect the integrity of the Supreme Court.” That letter led to the introduction of a bill in the House of Representatives that would establish new substantive and procedural recusal standards for the Court. The Justices’ individual conduct has also been the subject of recent criticism, including a heated


8. Justices Samuel Alito, Stephen Breyer, Ruth Bader Ginsburg, Antonin Scalia, Sonia Sotomayor, and Clarence Thomas have all recently been criticized for their interactions with politically interested entities that either have been or are likely to come before the Court. Justices Scalia and Thomas were criticized for attending Federalist Society fundraisers and dinners sponsored by the conservative Koch brothers on the basis that ties to politically active organizations could negatively impact their ability to remain impartial in future cases. See Nan Aron, An Ethics Code for the High Court, WASH. POST, Mar. 14, 2011, at A19; Nina Totenberg, Bill Puts Ethics Spotlight on Supreme Court Justices, NPR (Aug. 17, 2011), http://www.npr.org/2011/08/17/139646573/bill-puts-ethics-spotlight-on-supreme-court-justices (last visited Sept. 18, 2012) (on file with the Washington and Lee Law Review); Yeomans & Schwartz, supra note 4 (demanding a “reasoned explanation . . . of the propriety of the recent decision by Justices Clarence Thomas and Scalia to headline a fundraiser for the Federalist Society”). Justice Thomas was similarly criticized for his relationship with a wealthy conservative contributor who allegedly provided funding for projects of interest to the Justice and his wife. See Mike McIntire, The Justice and the Magnate, N.Y. TIMES, June 19, 2011, at A1. Justice Alito has drawn scrutiny for his attendance at fundraising dinners for the conservative American Spectator magazine. See Yeomans & Schwartz, supra note 4. Justices Breyer, Ginsburg, and Sotomayor were singled out for having accepted paid trips from organizations with political viewpoints. See The Justices’ Junkets, WASH. POST, Feb. 21, 2011, at A14 (noting that Justices Breyer, Ginsburg, and Sotomayor accepted trips paid for by the American Bar Association, the American Sociological Association, and the American Civil Liberties Union, respectively).
public debate over Justices Elena Kagan and Clarence Thomas's participation in the Court’s review of the controversial Affordable Care Act.\(^9\) Finally, exhaustive coverage in the news media and legal academy has made recusal an unavoidable part of any modern discussion of the Court.\(^{10}\)

---


Unfortunately, this intense focus on Supreme Court recusal largely misses the point. By exclusively treating recusal as a question of judicial ethics, commentators and legislators have overlooked the fundamental constitutional questions that ultimately drive the ongoing conflict between Congress and the Court over the Justices’ recusal practices. This Article is the second installment in a project designed to address that oversight by recasting Supreme Court recusal as a constitutional separation of powers issue. The precursor to this Article argued that the separation of powers constitutionally precludes Congress from creating legally binding recusal standards for the Justices—what I will refer to here as direct regulation of Supreme Court recusal decisions. This Article uses separation of powers principles to move past the identification of constitutional issues

and to offer solutions to the systemic constitutional problems surrounding recusal at the Court. It recognizes that although Congress may be constitutionally prohibited from directly regulating Supreme Court recusal, Congress retains constitutional authority and responsibility to indirectly influence recusal in ways that may better comport with our constitutional scheme.

There are at least three major benefits to reexamining recusal in the light of constitutional structure. First, such a review exposes the most serious problem with the current Supreme Court recusal regime, the (largely overlooked) constitutional impasse between Congress and the Court over recusal. Congress’s prior attempts to directly regulate recusal have been met with indifference or polite disregard by the Court, which in turn triggered calls for even more rigorous congressional regulation. There are significant problems with

12. The constitutional “impasse” advanced here is similar to the first element of a “constitutional showdown” as defined by Professors Posner and Vermeule: “[A] disagreement between branches of government over their constitutional powers.” Eric A. Posner & Adrian Vermeule, Constitutional Showdowns, 156 U. PA. L. REV. 991, 997 (2008). The conflict over recusal does not yet rise to the level of a “showdown,” however, because it neither “ends in the total or partial acquiescence by one branch in the views of the other [nor] . . . creates a constitutional precedent.” Id.

13. See 79 U.S.L.W. 2389 (Apr. 19, 2011) (describing Justice Kennedy’s testimony before the House Subcommittee on Financial Services and General Government, in which the Justice expressed the view that “there is a constitutional problem” with Congress prescribing ethical rules for the Court); 2011 YEAR-END REPORT, supra note 3, at 7 (noting that “the limits of Congress’s power to require recusal [of Supreme Court Justices] have never been tested”). The Court’s position was confirmed in its recent refusal of a congressional request to voluntarily adopt its own recusal reforms. See 2011 YEAR-END REPORT, supra note 3, at 8–9 (explaining that the Justices’ recusal decisions are unreviewable by the other Justices or another judicial body, and that those decisions are materially different from those of lower-court judges and thus harder to comport with the existing law on recusal); Letter from Chief Justice Roberts to Five Members of the Senate Judiciary Committee (Feb. 17, 2012) [hereinafter Chief Justice Letter], available at http://big.assets.huffingtonpost.com/LtrtoChairmanLeahyonYear-EndReport02172012.pdf (explaining that, despite the Senators’ request, “for the reasons explained in my year-end report, the Court does not plan to adopt the Code of Conduct for United States Judges”). For a more detailed discussion of Supreme Court recusal practice, see infra Part II.A.2.

14. See, e.g., Supreme Court Transparency Disclosure Act of 2011, H.R.
this approach. Direct regulation of recusal is legally and practically unenforceable against the Justices. Furthermore, the Court has never conceded that Congress has the constitutional authority to directly regulate its recusal practices, and has consistently acted as if that authority is lacking.\textsuperscript{15} The result is an effective impasse between Congress and the Court over recusal that has implications for the legitimacy, integrity, and efficacy of both Branches.

Second, reexamining recusal in the light of constitutional structure better reveals the nature and dynamics of the inter-branch conflict over recusal as well as the potential solutions to that conflict. These solutions depend on a fundamental, yet previously unanswered, question relating to Supreme Court recusal—how should the impasse be resolved, and who is responsible for resolving it? The answers lie with the separation of powers and, ultimately, with Congress. Viewing recusal as a separation of powers problem reveals that Congress and the Court are not equally capable of facilitating a successful resolution of the current impasse over recusal. The judicial options delegated to the Court by the Constitution are likely to harm the public perception, and thus democratic legitimacy, of both Branches by making the Court seem self-serving or obstinate. The constitutional choices assigned to Congress, by contrast, offer greater possibilities for coordinated efforts between the two Branches in pursuing their respective agendas regarding recusal. In terms of how a resolution should be achieved, focusing on Congress’s role as a constitutional actor allows us to see the entire range of constitutional options available to it, including

\textsuperscript{862, 112th Cong. § 3 (2011) (requiring recusal reform); Robert Barnes, Roberts: Justices Won’t Adopt Code of Conduct, WASH. POST, Feb. 22, 2012, at A7. (explaining that “[m]embers of Congress, a group of law professors and outside groups have called upon the court to adopt” ethics reform); Law Professor Letter, supra note 2, at 1 (advocating for recusal reform); Senator Letter, supra note 4 (requesting the Supreme Court to formally adopt the Code of Conduct for United States Judges).

15. These factors are also relevant to the constitutional question of whether Congress has the authority to directly regulate Supreme Court recusal in the first instance, a question that is addressed \textit{infra} at Part II.A and in my previous work. See Virelli, supra note 11, at 1185 (stating that separation of powers principles prevent Congress from directly regulating Supreme Court recusal).
various indirect approaches—such as impeachment, procedural reforms, judicial confirmation, appropriations, and investigations—that permit Congress to influence the Justices' recusal decisions without upsetting the delicate balance and coordination between the Branches that our constitutional scheme requires.

Third, reevaluating recusal in terms of constitutional structure provides insights into much larger questions about how best to handle inter-branch disputes. The separation of powers anticipates conflict among the coordinate Branches of the federal government, but it does not always provide easy answers as to how those conflicts should be treated or resolved. When two Branches reach an impasse over which has the power to act, is the answer a matter of law or constitutional politics? Which branch must or should take the initiative in promoting an amicable and workable resolution? Is the answer more difficult when the disagreement involves the Court? The answers to these questions are pivotal to understanding more fully the intricacies of our constitutional structure, as well as to maintaining a legitimate and effective constitutional democracy. The instant analysis of the impasse between Congress and the Court over recusal offers a useful view into the various approaches to be taken and political and institutional benefits to be sought with regard to other inter-branch conflicts.

This Article draws on the literature regarding Congress's role as a constitutional interpreter to argue that the separation of powers is critical to resolving the current impasse between Congress and the Court over recusal because it offers unique answers as to both why Congress must take the lead in resolving the impasse and how that resolution should be achieved. It contends that Congress should use indirect constitutional approaches, rather than direct regulation, to influence the Justices' recusal practices because doing so will alleviate the intractable tension created by direct congressional regulation of the Court and promote the democratic legitimacy of both Branches by reflecting a better balance of power between them. Part II describes the current Supreme Court recusal debate, including the constitutional impasse that has arisen between Congress and the Court over recusal. Part III draws on the literature regarding Congress's role as a constitutional interpreter to more closely examine its role regarding recusal. It
identifies the constitutional limits on Congress’s power over recusal and explains why Congress, rather than the Court, must be primarily responsible for breaking the constitutional impasse over recusal, albeit through indirect measures. Part IV addresses the array of indirect constitutional tools available to Congress regarding recusal and makes the normative case for why Congress should take the initiative to resolve the current impasse over recusal at the Court. Part V concludes and discusses some of the broader implications of the recusal analysis for the separation of powers.

II. The Debate over Recusal at the Court

The debate over recusal at the Supreme Court can be described in terms of two distinct viewpoints. The first is a reformist view. The overwhelming majority of commentary from legal academics, members of Congress, and journalists supports some measure of recusal reform for the Court. Proponents of reform express concern about the impact of the Justices’ recusal practices on the legitimacy of the Court and on the due process rights of individual litigants. The opposing view is focused on judicial independence. It is primarily represented—and occasionally, if rarely, put forth—by the Justices themselves. It is grounded in concerns about constitutional structure and function, and as such, centers on the unique role of the Court within our constitutional system. The interaction of these differing perspectives has created not only a heated controversy, but also an impasse in the conversation about Supreme Court recusal that, in my view, must be addressed if either set of concerns is to be alleviated.

A. Supreme Court Recusal in Action

Before fleshing out these contrasting views on Supreme Court recusal, it is useful to explain how recusal at the Court works. Supreme Court recusal has consistently operated on parallel tracks of congressional regulation and Supreme Court adjudication.
1. Congressional Regulation of Supreme Court Recusal

Since the beginning of the Republic, Congress has engaged in direct regulation of federal judges’ recusal practices. The Judiciary Act of 1792 contained the first statutory recusal standards.\(^\text{16}\) From 1792 through 1948, Congress revised the recusal standards for federal judges several times, and “in each instance... enlarged the enumerated grounds for seeking disqualification.”\(^\text{17}\) None of these revisions, however, applied to Supreme Court Justices. In 1948, Congress amended the existing recusal statute to include “justices” in addition to “judges” and required recusal where a judge or Justice had been a material witness or of counsel, possessed a “substantial” interest in the case, or was related to an attorney or party in the case such that it would be “improper, in [the judge’s] opinion” for the judge to hear that case.\(^\text{18}\) In 1972, the American Bar Association (ABA) adopted a new recusal rule in its Model Code of Judicial Conduct that, *inter alia*, required recusal of a judge “in a proceeding in which the judge’s impartiality might reasonably be questioned.”\(^\text{19}\) Two years later, Congress amended the federal recusal statute to codify the ABA Model Code’s objective reasonableness standard.\(^\text{20}\) This is the version of the federal recusal statute that applies to Supreme Court Justices today.

\(^{16}\) Act of May 8, 1792, ch. 36, § 11, 1 Stat. 275, 278–79. It was not until 1821 that Congress offered recusal standards that added anything significant to the common law criteria for judicial recusal, and even then the question of enforcement was left entirely to the judge being asked to recuse himself and was not applied to the Justices of the Supreme Court. See Act of Mar. 3, 1821, ch. 51, 3 Stat. 643.


\(^{18}\) Id. (emphasis added).

\(^{19}\) MODEL CODE OF JUDICIAL CONDUCT Canon 3E (2004). The Code also mandated recusal when a judge was personally biased, had served as a lawyer in the controversy, had a financial interest in the outcome of the case, or was within the third degree of relationship with a party, lawyer, interested person, or material witness in the case. See id. at 3E(1)(a)–(e).

In March 2011, however, thirty-three members of the House of Representatives sought to further expand the requirements for Supreme Court recusal with the introduction of the Supreme Court Transparency and Disclosure Act.\footnote{See Supreme Court Transparency Disclosure Act of 2011, H.R. 862, 112th Cong. (2011).} The bill was inspired in part by a letter from 138 law professors to the House and Senate Judiciary Committees outlining the need for legislation to “protect the integrity of the Supreme Court.”\footnote{Law Professor Letter, supra note 2, at 1.} The letter argued that the Code of Conduct for United States Judges should be applied to Supreme Court Justices.\footnote{Id. at 2.} It then stated that “justices must be subject to an enforceable, transparent process governing recusal”\footnote{Id.} and suggested some procedural reforms to achieve that goal, such as requiring a “written opinion when a Supreme Court justice denies a motion to recuse” and instituting “a procedure . . . for review of a decision by a Supreme Court justice not to recuse.”\footnote{Id. at 3.} The bill, in turn, mandated that “[t]he Code of Conduct for United States Judges . . . shall apply to the justices of the United States Supreme Court.”\footnote{H.R. 862, § 2(a).} It required the Justices to disclose “in the public record of the proceeding the reasons for the denial of [a recusal] motion” and established “a process under which . . . other justices or judges of a court of the United States” shall review a Supreme Court Justice’s denial of a recusal motion.\footnote{Id. § 3(a)(2), (b).} Although the bill has not yet been adopted, it evidences

Congress has begun in recent months to utilize some of its other constitutional tools with regard to the Justices’ recusal practices, including asking sitting Justices to testify about recusal matters. On April 14, 2011, Justices Breyer and Kennedy testified before a House subcommittee about Supreme Court ethics and recusal standards, and on October 5, 2011 Justices Breyer and Scalia testified about the same topic before the Senate Judiciary Committee. See supra note 5. Although both instances added public attention and awareness to the question of Supreme Court ethics and recusal as well as engaged the Court and its individual members actively and personally in the debate in a way that statutory prescriptions do not, they were relatively low-profile events that were not entirely dedicated to ethics and recusal questions.
the ongoing interest in legislative solutions to Supreme Court recusal issues.

Congress's focus on direct regulatory solutions to the perceived problems with Supreme Court recusal represents a popular and well-intentioned attempt to promote the important features of legitimacy and integrity at the Court. This regulatory approach has nevertheless failed to effectively constrain the Justices.

2. The Justices' Recusal Practices

A brief examination of the history of the Court's recusal practices reveals that Supreme Court recusal operates in almost precisely the same way today as it did at the Founding—as a personal, independent, unreviewable decision by an individual Justice whether to participate in an individual case. Around the time of the Founding, recusal was both procedurally and substantively a purely judicial question. Recusal doctrine was the product of judge-made common law, and judges were empowered to make the initial (and, in the case of United States Supreme Court Justices, the final) ruling as to their own recusal.29

Decisions by prominent members of the Court also reflect the Justices' unfettered approach to recusal. Chief Justice John

As such, they did not have the full impact that more formal, targeted hearings may provide. For a more detailed discussion of the promise of congressional hearings for seeking resolutions to the impasse over recusal at the Court, see infra Part IV.A.4.

28. See Flamm, supra note 1, §§1.2–1.4, at 5–8 (describing recusal practices as developing in the English common law and being adopted by the American colonial courts and, later, the federal Judiciary).

29. All of this information is consistent with historical accounts of the importance placed on the independence of judges in the period. As Professor Gerber explains in his detailed account of judicial independence in the colonial courts, a separate and independent judiciary was vitally important to each of the colonial and state systems that functioned as precursors to our federal government under the Constitution, as well as to the Framers as a check against potential overreaching by the Legislature, especially in the area of federal state relations. See Scott Douglas Gerber, A Distinct Judicial Power: The Origins of an Independent Judiciary, 1606–1787, 34–37 (2011).
Marshall presided over and wrote his famous opinion in *Marbury v. Madison*\(^{30}\) despite the fact that he was personally responsible for failing to deliver the judicial commission that gave rise to the case. Justice Oliver Wendell Holmes reviewed several cases as a Supreme Court Justice that he had participated in as a member of the Supreme Judicial Court of Massachusetts.\(^{31}\) Similarly, Justice Hugo Black sat for multiple cases reviewing legislation he had been instrumental in passing as a member of the Senate.\(^{32}\) Finally, Chief Justice Harlan Stone retracted his own recusal decision in a case, not because his views changed regarding his fitness to participate, but because he was concerned that without his participation the Court would not be able to achieve a quorum.\(^{33}\)

This sort of independent decision-making by the Justices did not change after Congress amended the recusal statute in 1948 to purportedly include Supreme Court Justices.\(^{34}\) The Justices continued to determine their own recusal status without any review or, in most cases, any public explanation for their decision. The rare instances where the Justices chose to publicly disclose

---

\(^{30}\) Marbury v. Madison, 5 U.S. (1 Cranch) 137 (1803).


their views regarding recusal confirm that the Justices’ decisions were made individually and independently. In *Laird v. Tatum*, then-Associate Justice Rehnquist published a memorandum explaining his decision not to recuse himself from a case challenging the constitutionality of a domestic surveillance program that he had been involved with (and testified about before Congress) during his time in the Justice Department’s Office of Legal Counsel. Although Justice Rehnquist cited the relevant recusal statute, he went on to remind the reader that “under the existing practice of the Court disqualification has been a matter of individual decision,” and pointed out that the unique nature of the Court makes Supreme Court recusal difficult as an institutional matter. In light of these special considerations, Justice Rehnquist argued that a Supreme Court Justice has a “duty to sit where not disqualified,” and as such should not “bend[] over backwards” to recuse himself. Justice Scalia described a similar process more than twenty years later in his memorandum in *Cheney v. U.S. District Court*. Justice Scalia was asked to recuse himself as a result of his recent participation in a hunting trip with the Vice President, who was a named party in the suit. In explaining his reasons for not recusing himself,

---

38. See id. at 837 (denying a motion for the Court to withdraw its opinion in *Laird*, 408 U.S. 1). Justice Rehnquist made the identical argument in another memorandum explaining a decision not to recuse himself in *Microsoft Corp. v. United States*, 530 U.S. 1301, 1303 (2000):

Finally, it is important to note the negative impact that the unnecessary disqualification of even one Justice may have upon our Court. Here—unlike the situation in a District Court or a Court of Appeals—there is no way to replace a recused Justice. Not only is the Court deprived of the participation of one of its nine Members, but the even number of those remaining creates a risk of affirmance of a lower court decision by an equally divided court.

40. *Id.* at 838 (internal quotation marks omitted).
Justice Scalia rejected the statutory presumption that he should “resolve any doubts in favor of recusal” because “the Supreme Court . . . is different.”

The Court reinforced the idea that the Justices have their own permissive view of recusal in its 1993 Statement of Recusal Policy. The Statement was signed by seven of the nine sitting Justices and only addressed cases where a Justice’s relative participates in a matter before the Court. As Professor Sherrilyn Ifill described it, “the Recusal Policy simply reflects the Justices’ own sense of what to them would constitute a reasonable basis upon which to question a judge’s impartiality.”

Most recently, Chief Justice Roberts explained in his Year-End Report that “[l]ike lower court judges, the individual Justices decide for themselves whether recusal is warranted.” He went on to explain that due to institutional differences between the Supreme Court and the lower courts, a “Justice accordingly cannot withdraw from a case as a matter of convenience or simply to avoid controversy. Rather, each Justice has an obligation to the Court to be sure of the need to recuse before deciding to withdraw from a case.”

All of these examples demonstrate that, at minimum, the Justices consider a host of prudential factors in making their recusal decisions, and they do so unencumbered by the prospect of review or the presence of statutory requirements. As Professor Jeffrey Stempel so aptly described it, “Supreme Court recusal practice provides an almost unique illustration in American government of substantive law without force when applied to a certain institution.”

---

42. Id. at 915.
44. Justices Blackmun and Souter did not sign the policy. See id. at 1101–03 & n.1.
46. 2011 YEAR-END REPORT, supra note 3, at 8.
47. Id. at 9.
48. Stempel, supra note 32, at 642.
B. The Arguments over Reform

The Justices' longstanding practice of making individualized, unreviewable recusal decisions has not, however, gone unnoticed. Critics of the Court's recusal practices argue that the Justices' failure to submit their decisions to more traditional legal processes has damaged the integrity and reputation of the Court. The Justices' increasingly publicized involvement in political causes and organizations has highlighted concerns about the fairness and impartiality of the Justices' decisions and has resulted in calls for more stringent ethics and recusal standards.49 The fact that the Justices' recusal decisions are unreviewable has incited discussion about the quality of those decisions, especially in fact-specific inquiries like recusal in which a Justice functions as both fact finder and adjudicator.50

Advocates of recusal reform have suggested mandating review of an individual Justice's decision not to recuse by the remainder of the Court or by a special committee of federal judges assembled specifically for that purpose. Professor Stempel has proposed a standard of review under which "[a]ny party aggrieved by the refusal of a Supreme Court Justice to disqualify himself may, on timely motion, obtain review by the full Supreme Court."51

49. The most recent and popular movement in this regard is to apply the Code of Conduct for United States Judges to the Justices. See Supreme Court Transparency Disclosure Act of 2011, H.R. 862, 112th Cong. § 2(a) (2011); Law Professor Letter, supra note 2, at 2.

50. As Professor Stempel explained:
The Court also lacks any formal rule, mechanism, or custom of permitting fact development in aid of a recusal motion...[T]hese facts concerning the impartiality of a Supreme Court Justice have never been permitted to develop the facts of the alleged conflict under the auspices of the Court. Occasionally, as in Tatum, a Justice will offer a version of the facts in answer to the motion, which hardly passes as meaningful discovery or even scrutiny.

Stempel, supra note 32, at 642; see also Amanda Frost, Keeping Up Appearances: A Process-Oriented Approach to Judicial Recusal, 53 U. Kan. L. Rev. 531, 576 (2005) (noting how in his memorandum in Cheney v. U.S. District Court, "Justice Scalia revealed facts about circumstances and logistics of the trip that previously had been unknown" as part of his explanation for deciding not to recuse himself from the case).

51. Stempel, supra note 32, at 644.
2011 House bill on Supreme Court ethics and recusal went further, allowing for review of a Justice's decision not to recuse by "other justices or judges or a court of the United States, among whom retired justices and senior judges . . . may be included." The most compelling criticism of the Court's recusal practices involves the Justices' reluctance to explain their decisions. The Court offers public explanations of its decisions as a means of remaining accountable and protecting its own legitimacy. The Justices' reluctance to explain their recusal decisions imperils accountability and legitimacy, especially because the practice is out of step with judicial behavior more broadly. The lack of transparency in the Court's current recusal practices has been described as indefensible in a modern democratic society, particularly in an age where other information about the Justices' practices is so readily available, and has inspired calls for statutory requirements that Justices publish their reasons for denying a recusal motion. Professor Debra Lynn Bassett has advocated for greater disclosure of the Justices' reasons for recusal through "statements of interest." More recently, Professors William Yeomans and Herman Schwartz have argued that:

"The courts' fundamental legitimacy rests on the notion that judges . . . explain what they have done in reasoned opinions for all to read. . . . This same transparency is even more essential when Justices apply the law to themselves. . . . Courts obviously need secrecy for their deliberations and decision making. But there can be no harm in a justice explaining why he or she withdraws from a case or refuses to withdraw."

52. H.R. 862, § 3(b) (emphasis added). The prospect of allowing lower court judges to review Supreme Court recusal decisions potentially runs afoul of the constitutional requirement that there be "one supreme Court." U.S. Const. art. III, § 1.
53. See H.R. 862, § 3(a)(2) ("If a justice of the Supreme Court denies a motion . . . that the justice should be disqualified . . . the justice shall disclose . . . the reasons for the denial of the motion."); Law Professor Letter, supra note 2, at 2.
54. Bassett, supra note 17, at 695.
55. Yeomans & Schwartz, supra note 4.
Finally, the above complaints evoke a common concern about the effects of the Justices’ recusal practices on the public perception of, and confidence in, the Court. For an institution like the Supreme Court that depends so heavily for its institutional effectiveness on public confidence, any damage done to that perception presents a significant problem for the Court specifically, for the other institutions of government, and for the principle of separation of powers more generally.

The responses to these arguments for recusal reform come principally from the Justices themselves. They focus, perhaps not surprisingly considering their source, less on the nature of the Justices’ recusal decisions and more on the effects of those decisions on the Court’s institutional mission. The central theme of this viewpoint is derived from the common law “rule of necessity,” under which a judge’s decision to recuse is overridden by the lack of an adequate replacement to hear the case. In the context of the lower federal courts, the rule of necessity is effectively a nullity. Congress has provided, pursuant to its power to create the lower federal courts, myriad options for calling into service replacement judges in the event a judge is recused. The


The Court’s power lies, rather, in its legitimacy, a product of substance and perception that shows itself in the people’s acceptance of the Judiciary as fit to determine what the Nation’s law means and to declare what it demands. . . . Thus, the Court’s legitimacy depends on making legally principled decisions under circumstances in which their principled character is sufficiently plausible to be accepted by the Nation.

57. See United States v. Will, 449 U.S. 200, 213 (1980) (“[A]lthough a judge had better not, if it can be avoided, take part in the decision of a case in which he has any personal interest, yet he not only may but must do so if the case cannot be heard otherwise.”(quoting F. Pollock, A First Book of Jurisprudence 270 (6th ed. 1929))); see also Flamm, supra note 1, § 20.2, at 576–77 (explaining that the “rule of necessity” involves the principle that disqualification will not be permitted to destroy the only tribunal with power to act in the premises—that is, “where disqualification would result in an absence of judicial machinery capable of dealing with a matter, . . . disqualification must yield to necessity”).

Supreme Court, however, is different. The constitutional mandate that there be “one supreme Court” creates significant—if not impenetrable—barriers to the availability of replacement Justices. The lack of replacement Justices in turn creates the possibility that the recusal of a single Justice will facilitate a tie vote, resulting in a non-precedential affirmation of the decision below that does not “resolve the significant legal issue presented by the case.” Recusal is also problematic for the litigants, the argument goes, because a decision by an irreplaceable Justice to recuse himself is “effectively the same as casting a vote against the petitioner.” In cases involving at least one recusal, the petitioner must garner a larger percentage of the available Justices’ votes both on the merits (five out of eight rather than five out of nine) and at the certiorari stage (four out of eight rather than four out of nine for a certiorari petition). The recusal of multiple Justices is even more problematic for the Court, as it could defeat quorum in a specific case and thus make the Court powerless to exercise its constitutional function as the final adjudicator of the “cases or controversies” properly before it.

60. See Letter from Chief Justice Hughes to Senator Wheeler, Mar. 22, 1937, available at http://newdeal.feri.org/court/hughes.htm (“The Constitution does not appear to authorize two or more Supreme Courts or two or more parts of a Supreme Court functioning in effect as separate courts.”); see also Ruth Bader Ginsburg, An Open Discussion with Ruth Bader Ginsburg, 36 CONN. L. REV. 1033, 1039 (2004) (explaining that “there’s no substitute for a Supreme Court Justice”); Wheeler, supra note 10, at 2 (contending that a court of lower court judges sitting in review of Supreme Court recusal decisions “would most likely violate the Constitution’s ‘one Supreme court’ mandate”). But see Lisa T. McElroy & Michael C. Dorf, Coming Off the Bench: Legal and Policy Implications of Proposals to Allow Retired Justices to Sit by Designation on the Supreme Court, 61 DUKE L.J. 81, 104 (2011) (arguing that retired Justices, but not those that have resigned their commissions, may be eligible to substitute for recused justices in Supreme Court cases).
63. Professor Lubet described this phenomenon as the “Certiorari Conundrum.” Steven Lubet, Disqualification of Supreme Court Justices: The Certiorari Conundrum, 80 MINN. L. REV. 657, 661–65 (1996).
64. Virelli, supra note 11, at 1213–17. These same arguments may apply in
There are other reasons to support the Justices’ retaining control over recusal that are not grounded in concerns about a lack of eligible replacements for a recused Justice. Justice Scalia has argued that limiting the Justices’ independence over recusal decisions could lead to politically motivated attacks designed to influence the outcome of a case by forcing unsympathetic Justices to recuse themselves.65 He explains that “[t]he people must have confidence in the integrity of the Justices, and that cannot exist in a system that assumes them to be corruptible by the slightest friendship or favor, and in an atmosphere where the press will be eager to find foot-faults.”66 The Justices have also raised constitutional concerns about congressional influence over recusal decisions at the Court. While testifying before the House Subcommittee on Financial Services and General Government,67 Justice Kennedy pointed out that “there is a constitutional problem” with making statutory recusal standards binding on the Court.68 In his Year-End Report, Chief Justice Roberts made a point of explaining that “the limits of Congress’s power to require recusal [of Supreme Court Justices] have never been tested. The Justices follow the same general principles with respect to recusal as other federal judges, but the application of those principles can differ due to the unique circumstances of the Supreme Court.”69

65. This phenomenon is considered by many to be the driving force behind the recent calls for Justices Kagan and Thomas to recuse themselves from cases reviewing the Affordable Care Act. See, e.g., Mukasey, supra note 10 (“The persistence of recusal issues appears to have little to do with the legal merits . . . but a great deal to do with . . . agenda-driven politics . . . ”).
66. Cheney, 541 U.S. at 928.
67. Malloy, supra note 5, at 2389.
68. Id.
69. 2011 YEAR-END REPORT, supra note 3, at 7.
C. The Impasse

The ongoing debate over the proper role of legislative involvement in Supreme Court recusal practice reflects a constitutional impasse between Congress and the Court. Although proponents of statutory recusal standards acknowledge the important differences between recusal in the lower courts and the Supreme Court, concerns over the integrity, fairness, and transparency of the Court have caused them to look to Congress to standardize and strengthen the Court’s recusal standards. But Congress’s attempts to regulate the Court’s recusal practices have inspired both direct and indirect resistance from the Justices. In February of 2012, Chief Justice Roberts responded to a request from five members of the Senate Judiciary Committee asking the Court to formally adopt the Judicial Conference of the United States Code of Conduct for United States Judges, which sets ethical and recusal standards for the lower federal courts. The Chief Justice plainly stated in a letter to the Senators that “the Court does not plan to adopt the Code of Conduct.” The Chief Justice’s remarks came in the wake of his Year-End Report, in which he highlighted the unique role of the Supreme Court within the Judiciary and made clear that the Justices consider their ethical and recusal obligations to be individual and independent.

These recent (and defiant) comments about recusal are the most direct in a long list of indications that the Justices do not feel bound by congressional attempts to regulate their recusal decisions. For instance, the Court has never conceded that Congress has the constitutional authority to set recusal standards. Moreover, although recusal remains a relatively

70. See, e.g., Flamm, supra note 1, § 29.4, at 916–17 (suggesting that there may be a “more compelling ‘duty to sit’ for Supreme Court Justices than for other judges”); Bassett, supra note 17, at 682–93 (discussing proposals for standards of Supreme Court recusal); Ifill, Do Appearances Matter?, supra note 45, at 619 (“To be certain, Justices on the Supreme Court face legitimate concerns that are not at issue for judges on other courts who are faced with recusal motions.”).


72. See id. at 7 (“[T]he limits of Congress’s power to require recusal have never been tested.”).
common practice at the Court, the Justices rarely explain their
decisions. When explanations are proffered, institutional
concerns about the effective functioning of the Court predominate
over the statutory standards put forth by Congress. Most
importantly, because the statutory standards are not enforceable
against the Court and because the Court’s decisions are not
reviewable, there is no coercive legal action available to Congress
to force the Justices to change their practices. In the words of
Professor Ifill, “the Justices encourage and protect a fiercely
independent approach to their recusal determinations.”

This impasse is more than a legal abnormality or
inconvenience. In many ways, it is the source of much of the
public frustration with, and criticism of, the Court and Congress
over recusal. On a small scale, the impasse makes it more
difficult to achieve any sort of lasting resolution or thoughtful
treatment of the recusal question as the two sides effectively talk
past one another. On a much grander scale, the impasse over
recusal raises the same systemic problem as any seemingly
irreconcilable dispute between two coequal Branches—doubts
about the legitimacy of our constitutional democracy. As the two
Branches continue to appear at odds over when and how
Supreme Court Justices should refrain from participating in
specific cases, both sides appear to be obstinate, arbitrary,
ineffective, or some combination thereof. This promotes an air of
lawlessness around the issue that weakens confidence in our
public institutions and creates problems for our constitutional
structure. Finally, the recusal impasse is significant because it
could be the first step in what Professors Posner and Vermeule

73. See McElroy & Dorf, supra note 60, at 99 (explaining that “Justices
would almost certainly recuse themselves in clear-cut conflict situations”);
see also Times Topics, Elena Kagan, NYTIMES.COM (Oct. 4, 2010)
(last visited Sept. 18, 2012) (noting that “[b]ecause of her tenure as solicitor
general, Justice Kagan has recused herself from about half of the 54 cases”
so far on the Court’s docket for the 2010 term) (on file with the Washington
and Lee Law Review).

74. See, e.g., Frost, supra note 50, at 569 (“Judges who recuse themselves
rarely issue a decision explaining why.”).

75. See infra notes 111–17 and accompanying text.

76. Ifill, Do Appearances Matter?, supra note 45, at 622.
call “constitutional showdowns”—inter-branch disputes over constitutional authority that end in the development of new constitutional precedents.\textsuperscript{77}

The harmful effects of the impasse over recusal represent a serious and heretofore unforeseen or unmentioned problem within Supreme Court recusal. The existing debate over what the Court should be required to do and whether those requirements would lead to untenable consequences largely misses the point. The impasse between Congress and the Court over recusal should first be viewed as a constitutional separation of powers problem in order to properly contextualize the entire range of tensions and issues at stake. Only then can we envision a solution that is worthy of a conflict between two coequal Branches of government.

\textbf{III. Recusal and the Separation of Powers}

\textbf{A. Congress’s Constitutional Responsibilities}

Resolving a conflict between two Branches of government necessarily evinces the separation of powers. The Framers’ concept of the separation of powers is often attributed to Montesquieu,\textsuperscript{78} and is embodied in a tripartite constitutional structure that envisions constitutional responsibilities for each branch that are separate and apart from its counterparts in order to protect against overreaching or aggrandizement by any single branch. This has remained a foundational principle of American

\textsuperscript{77} See Posner & Vermeule, supra note 12, at 997.

\textsuperscript{78} See Baron de Montesquieu, The Spirit of the Laws 149–62 (Thomas Nugent trans., Hafner Publishing Co. 1949) (1748) (discussing separation of powers as a concept for ensuring liberty). In addition to being an implicit part of many of the arguments about governmental structure at the Convention, Montesquieu’s views regarding the separation of powers were explicitly relied upon by several of the delegates. See 1 The Records of the Federal Convention of 1787, at 391 (Max Farrand ed., 1911) (remarks of Butler) (“The great Montesquieu says, it is unwise to entrust persons with power, which by being abused operates to the advantage of those entrusted with it.”); 2 The Records of the Federal Convention of 1787, at 34 (Max Farrand ed., 1911) (remarks of Madison) (“[A]ccording to the observation of Montesquieu, tyrannical laws may be made that they may be executed in a tyrannical manner.”).
government ever since. As the Court recently reiterated in *Stern v. Marshall*, the separation of powers, at minimum, anticipates a division of labor among the Branches, such that some spheres of authority are reserved exclusively to individual Branches. The *Stern* Court went on to acknowledge, however, that “the three Branches [of the federal government] are not hermetically sealed from one another.” In fact, some redundancies are necessary in order to effectively protect against the aggrandizement and encroachment that the Framers saw as crucial to warding off tyranny. Chief among these redundancies is the shared responsibility of all three constitutional Branches of government to interpret the Constitution, especially with regard to the separation of powers. As James Madison explained:

> I beg to know upon which principle it can be contended that any one department draws from the Constitution greater powers than another in marking out the limits of the powers of the several departments. The Constitution is the charter of the people to the government; it specifies certain great powers as absolutely granted and marks out the departments to exercise them. If the constitutional boundary of either be brought into question, I do not see than any one of these independent departments has more right than another to declare their sentiments on that point . . . .

---


80. As Chief Justice Roberts articulated in *Stern*:
   > Under “the basic concept of separation of powers . . . that flow[s] from the scheme of a tripartite government” adopted in the Constitution, “the ‘judicial Power of the United States’ . . . can no more be shared” with another branch than “the Chief Executive, for example, can share with the Judiciary the veto power, or the Congress share with the Judiciary the power to override a Presidential veto.”

81. Id. at 2609.


84. 12 THE PAPERS OF JAMES MADISON 238–39 (William T. Hutchinson et al.
Madison’s view was echoed by Presidents Jefferson and Jackson\(^{85}\) and nearly two centuries later by the Supreme Court in *United States v. Nixon*:\(^{86}\) “In the performance of assigned constitutional duties, each branch of the Government must initially interpret the Constitution, and the interpretation of its powers by any branch is due great respect from the others.”\(^{87}\)

The scholarly literature on Congress’s interpretive responsibilities is, not surprisingly, wide-ranging and diverse. Questions abound about, *inter alia*, Congress’s institutional competency to interpret the Constitution,\(^{88}\) the role of institutional design in improving that core competency,\(^{89}\) and how to balance the interpretive role of nonjudicial actors like Congress against that of the Judiciary.\(^{90}\) This robust debate, however, does

----

\(^{85}\) Jefferson wrote that “nothing in the Constitution has given [the judges] a right to decide for the Executive, more than to the Executive to decide for them.” Letter from Thomas Jefferson to Mrs. John Adams (Sept. 11, 1804), *in The Writings of Thomas Jefferson* 49, 50 (Andrew A. Lipscomb ed., 1905); see also President Jackson’s Veto Message to the Senate (July 10, 1832), *in Messages and Papers of the Presidents* 576, 582 (James D. Richardson ed., 1908) (“The opinion of the judges has no more authority over Congress than the opinion of Congress has over the judges, and on that point the President is independent of both.”).


\(^{87}\) *Id.* at 703.


\(^{89}\) See, e.g., Elizabeth Garrett & Adrian Vermeule, *Institutional Design of a Thayerian Congress*, 50 Duke L.J. 1277, 1278 (2001) (arguing that Congress can be improved by focusing “upon questions of institutional design”).

\(^{90}\) See, e.g., Larry Alexander & Frederick Schauer, *On Extrajudicial
not ultimately take issue with the narrower conception of legislative constitutional interpretation articulated by the Founding Fathers and the Court. As Professor James Bradley Thayer stated in his seminal article *The Origin and Scope of the American Doctrine of Constitutional Law*,91 “it is the legislature to whom this power is given—this power, not merely of enacting laws, but of putting an interpretation on the [C]onstitution which shall deeply affect the whole country.”92 In the more than 100 years since Professor Thayer’s pronouncement, a general consensus has emerged that the separation of powers includes an obligation of Congress to consider the constitutionality of its actions vis-à-vis the other Branches.93 That is not to say that Congress is a better source of constitutional understanding than the courts, or that Congress as a descriptive matter engages in robust constitutional interpretation regularly. None of these far more controversial issues is necessary to the task at hand. Because Supreme Court recusal has created an impasse between Congress and the Courts, it is enough for the present discussion to acknowledge that Congress is constitutionally authorized and (at least in part) responsible for evaluating the constitutionality

---


92. Id. at 136.

93. Professors Alexander and Schauer put it thusly:
A recurrent claim in American constitutional discourse is that judges should not be the exclusive and authoritative interpreters of the Constitution. The Constitution speaks to all public officials, it is said, and thus all officials, not just judges, must make their own decisions about what the Constitution commands. To hold otherwise, it is argued, is to fail to recognize the constitutional responsibilities of officials who happen not to be judges.

of its actions and, in turn, for resolving the constitutional tension created by Supreme Court recusal.

Reorienting Supreme Court recusal in a constitutional framework and acknowledging Congress’s constitutional responsibilities with regard to the separation of powers highlights two significant constitutional questions about the impasse over recusal. The first is whether Congress has the constitutional authority to limit a Justice’s participation in a specific case. The second is which branch of government is best suited to resolve that impasse. In answering each of these questions, the remainder of this Part highlights the need—as a matter of both constitutional law and policy—for alternative approaches by Congress to the issue of recusal at the Court.

B. Constitutional Limits on Congressional Regulation

Congressional regulation of Supreme Court recusal triggers serious concern about the separation of powers; 94 may Congress pass legislation affecting the Court’s recusal decisions, or does the Constitution vest sole authority for those decisions in the Justices themselves? Article III unequivocally vests the “judicial Power of the United States” in the Supreme Court and “in such inferior Courts as the Congress may . . . ordain and establish.” 95 It does not, however, offer any further explanation of the scope or substance of the judicial power, much less whether and to what extent that power includes the authority of Supreme Court Justices to decide questions of their own recusal. Without greater textual guidance on the matter, constitutional history, practice,

94. See Virelli, supra note 11 (advancing this argument in greater detail).
95. U.S. CONST. art. III, § 1. This express congressional power over the lower federal courts is not applicable to the Supreme Court, and is a significant limit on the inherent power of those lower courts. It shifts the balance of power between Congress and the lower courts toward Congress while leaving the balance between Congress and the Supreme Court undisturbed. This difference in the constitutional power sharing arrangement between Congress and the Supreme Court versus the lower federal courts is the main reason why the instant analysis focuses solely on recusal at the Supreme Court, rather than federal courts in general.
and structure become crucial to understanding the constitutional status of Supreme Court recusal.

The records of the Constitutional Convention and state ratifying debates are virtually silent on the meaning of the “judicial Power” generally, let alone the power of judges and Justices to decide questions about recusal. A historical analysis of Supreme Court recusal under Article III must therefore be based both on the sources of information most commonly understood to have been relied upon by the Framers in fashioning the judicial power—the “business of the Colonial courts and the courts of Westminster when the Constitution was framed”—as well as on “early congressional and judicial precedent” interpreting Article III. Although little direct evidence exists of the Framers’ view of Supreme Court recusal specifically, the evidence that is available supports the idea that recusal was, in and around the Founding, both procedurally and substantively a question for an independent Judiciary. English and colonial judges developed highly permissive common law standards for recusal and used those standards to decide their own recusals.


99. See FLAMM, supra note 1, § 1.4, at 8 (“In the pre-Revolutionary American Colonies, as in England, the only accepted ground for disqualifying a judge was pecuniary interest in a pending cause; and for years following
matters. With regard to the Supreme Court in particular, the early Congresses chose not to intervene in the Justices’ exercise of their recusal power even after they chose to do so for the lower federal courts. This practice indicates that, at the time the Constitution was drafted, Supreme Court recusal was a matter for the Court through its exercise of the judicial power granted to it by Article III.

Longstanding governmental practice further corroborates this view. Although the most well-known examples occur in disputes between the Legislative and Executive Branches,

100. See, e.g., John P. Frank, Disqualification of Judges, 56 YALE L.J. 605, 612 (1947) (“In the Supreme Court disqualification has always been the prerogative of each individual Justice . . . .”).

101. See Act of May 8, 1792, ch. 36, § 11, 1 Stat. 275, 278–79 (regulating only federal “judges,” rather than “justices”).

102. As Justice Frankfurter and Professor Landis explained: “The scope and qualities of a power which has been voluminously exercised since 1789 must be looked for in the cumulative proof of its exercise.” Frankfurter & Landis, supra note 96, at 1018. The Court has continually relied on traditional understandings and practices of the three Branches of government to determine the proper constitutional balance among them. Perhaps the most well-known statement regarding the division of authority between the Legislative and Executive Branches is found in Justice Jackson’s concurrence in Youngstown Sheet & Tube Co. v. Sawyer, in which he articulated a three-tier system for resolving separation of powers disputes. Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 635–38 (1952) (Jackson, J., concurring). Justice Jackson’s famous system considered evidence of historical congressional “inertia, indifference or quiescence” to the challenged executive action as important to understanding the proper balance of power between the Branches. Id. at 637. Although congressional intent regarding the conduct of a coordinate branch is difficult to determine, its relevance in Justice Jackson’s framework highlights the important fact for present purposes: that constitutional meaning in separation of powers disputes can be gleaned from established governmental practices in the contested area. Justice Frankfurter made this clear in his own Youngstown concurrence:

In short, a systematic, unbroken, executive practice, long pursued to the knowledge of the Congress and never before questioned, engaged in by Presidents who have also sworn to uphold the Constitution, making as it were such exercise of power part of the structure of our government, may be treated as a gloss on “executive Power” vested in the President by § 1 of Art. II.

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

103. See, e.g., Dames & Moore v. Regan, 453 U.S. 654 (1981); Pocket Veto
government practice has also played an important role in resolving conflicts between legislative and judicial power. In the case of Supreme Court recusal, longstanding legislative and judicial custom indicates that the decision to recuse a Supreme Court Justice is not only a judicial decision, but also one that is constitutionally protected from legislative interference.

Congress did not seek to regulate Supreme Court Justices’ recusal practices for the first 150 years of the Republic. In 1948, Congress amended the federal recusal statute to include, for the first time, “[a]ny justice . . . of the United States,” but it did not set any procedural standards for deciding recusal questions, disrupt the ongoing practice of judges resolving their own recusal questions in the first instance, or address the review of those initial decisions. With regard to the Supreme Court, the amended statute has led to the seamless continuation of the Justices’ historical recusal practices—unreviewable, individualized determinations by each Justice of their own qualification to sit in a particular case, without any obligation to justify or otherwise explain their decisions. As Professor Ifill described it, “[i]t appears that the Justices on the Court enjoy the unreviewable power to determine individually whether and when to disqualify themselves.” Nevertheless, in light of ample historical and


104. In Chambers v. NASCO, Inc., 501 U.S. 32 (1991), the Court confirmed judges’ ability to impose sanctions for bad faith conduct in the absence of any statutes or regulations to that effect. See id. at 42, 58. The Chambers Court explained that “[i]t has long been understood that ‘[c]ertain implied powers must necessarily result to our Courts . . . from the nature of their institution.’” Id. at 43 (alteration in original) (quoting United States v. Hudson, 11 U.S. (7 Cranch) 32, 34 (1812)). In Link v. Wabash Railroad Co., 370 U.S. 626 (1962), the Court also relied on longstanding practice to uphold judicial power to sua sponte dismiss for failure to prosecute despite an existing rule ostensibly requiring a motion by defendant to trigger dismissal. See id. at 628–30.


106. Ifill, Do Appearances Matter?, supra note 45, at 620. See also R. Matthew Pearson, Note, Duck, Duck Recuse: Foreign Common Law Guidance and Improving Recusal of Supreme Court Justices, 62 WASH. & LEE L. REV. 1799, 1813–14 (2005) ("The 'historic practice' of the United States Supreme Court has always been to refer motions for recusal to the Justice whose disqualification is sought. Thus . . . the actual procedure by which the decision is made is truly a creature of tradition.").
modern evidence of the Court’s erratic and at times seemingly unprincipled recusal decisions, Congress has not taken any definitive steps to enforce its statutory requirements or to indicate that it considers the Court’s ongoing recusal practices problematic. This acquiescence or indifference to Supreme Court recusal practice is evidence of the Justices’ constitutional authority to decide recusal questions independently.

The Justices’ longstanding practices suggest a similar constitutional interpretation. Prior to the recusal statute being amended to include them, several of our most esteemed Justices made highly controversial recusal decisions without any interference or objection from Congress. A similar pattern emerged after the statute was amended, as the Justices continued to make controversial decisions that were at best tangentially faithful to the statutory standard. In their memoranda in *Laird* and *Cheney*, respectively, Justices Rehnquist and Scalia relied heavily on extrastatutory factors in justifying their decisions not to recuse. They cited the Justices’ individual power to make recusal decisions, the recusal


108. The instant example of the Justices retaining dominion over their own recusal decisions despite congressional action to the contrary provides even stronger support for the Supreme Court’s exclusive power over recusal than evidence of judicial power in the absence of legislative activity. *See generally* Dames & Moore v. Regan, 453 U.S. 654 (1981).


110. The Court has never taken a case in which it has applied the federal recusal statute to one of its own members. See Bassett, *supra* note 17, at 676–80 (outlining the “four major opinions touching on section 455,” all of which dealt with recusal decisions by a lower court judge).

111. 409 U.S. 824 (1972).


114. *See Laird*, 409 U.S. at 833 (reminding the reader that “under the existing practice of the Court disqualification has been a matter of individual
SUPREME COURT RECUSAL

precedents of other Justices, and the Court’s unique institutional features—such as its lack of substitute Justices and the problems created when it issues equally divided decisions—to support their conclusion that the Justices should employ an effective presumption against recusal. These memoranda decision”).

115 See Cheney, 541 U.S. at 924–26 (citing decisions by Justices White and Jackson not to recuse themselves despite their close friendships with high-ranking members of the Administration); Laird, 409 U.S. at 829–30, 31 (explaining that “different Justices who have come from the Department of Justice have treated the same or very similar situations differently” from one another, and Justice Rehnquist’s “impression is that none of the former Justices of this Court since 1911 have followed a practice of disqualifying themselves” in analogous cases). It is worth noting that one of the decisions that Justice Scalia cited as useful precedent occurred before the federal recusal statute was amended to include Supreme Court Justices.

116. See Cheney, 541 U.S. at 915 (identifying the potential problem of a “tie vote [leaving the Court] unable to resolve the significant legal issue presented by the case”); id. at 916 (explaining that recusal is “effectively the same as casting a vote against the petitioner”). Some proponents of congressional reform of Supreme Court recusal, including H.R. 862, the recent proposal introduced in the House, have suggested using circuit court judges to review Supreme Court recusal decisions. See Supreme Court Transparency Disclosure Act of 2011, H.R. 862, 112th Cong. (2011). While such an arrangement may appeal to those concerned with the ethical ramifications of current Supreme Court recusal practices, there are significant constitutional problems with such an arrangement, including Article III’s mandates that there be only “one supreme Court,” and that Congress have power to create only “inferior courts.” U.S. CONST. art. III, § 1. Justice Rehnquist made the identical argument in another memorandum explaining a decision not to recuse himself in Microsoft v. United States:

Finally, it is important to note the negative impact that the unnecessary disqualification of even one Justice may have upon our Court. Here—unlike the situation in a District Court or a Court of Appeals—there is no way to replace a recused Justice. Not only is the Court deprived of the participation of one of its nine Members, but the even number of those remaining creates a risk of affirmance of a lower court decision by an equally divided court.


117. More specifically, Justice Rehnquist argued that a Supreme Court Justice has an “even stronger” “duty to sit where not disqualified,” Laird, 409 U.S. at 837, and as such Justices should not “bend [] over backwards” to recuse themselves. Id. at 838; see also Cheney, 541 U.S. at 915 (rejecting the federal recusal statute’s presumption in favor of recusal because the “Supreme Court . . . is different”). Justice Scalia added a warning that recusal would “harm the Court” by encouraging the use of recusal as a means of influencing the outcome of cases before the Court. Cheney, 541 U.S. at 927 (expressing
reveal the Justices’ general disregard for the statutory recusal standards. Although both Justices referred to the statutory recusal text in their analyses, they did not employ any of the traditional tools of statutory interpretation; they did not parse the statutory language, attempt to divine congressional intent, examine legislative history, or evaluate lower court interpretations of the statute. Apparently, neither Justice thought of their analysis as primarily statutory, but instead as a broader, independent examination of their fitness to participate in the case at hand.118

The Court’s 1993 “Statement of Recusal Policy”119 confirms the Justices’ commitment to independent, extrastatutory recusal decisions. The Policy claimed to apply the relevant statute, but ultimately relied on institutional concerns to read it very narrowly.120 As Professor Bassett observed, the Court’s Recusal Policy “re-emphasized [the Justices’] negative view of recusal” by “simply reflect[ing] the Justices’ own sense of what to them would constitute a reasonable basis upon which to question a judge’s impartiality and appl[y]ing that standard across the board.”121 Justices Breyer and Kennedy reinforced the idea that the Supreme Court feels ultimately unrestrained in its recusal practice, and that this feeling may be constitutionally justified, during their congressional testimony in April of 2011,122 and

---

118. This is particularly noteworthy in the case of Justice Scalia, whose well documented views on statutory interpretation seek to avoid, seemingly at all costs, policy-based analyses of statutory issues. See Antonin Scalia, Originalism: The Lesser Evil, 57 U. Cin. L. Rev. 849, 863 (1989) (“[T]he main danger in judicial interpretation . . . of any law . . . is that the judges will mistake their own predilections for the law.”).


120. See id. at 1103 (citing as support for its position the fear of strategic recusal motions, “the possibility of an even division on the merits of a case,” and a “distorting effect on the certiorari process” in cases where recusal occurs).


122. See 79 U.S.L.W. 2389, 2389 (Apr. 19, 2011). During the hearing, Justice Kennedy testified that “there is a constitutional problem” with making ethics rules binding on the Court, and Justice Breyer reiterated “that it is a more
Chief Justice Roberts did the same in his Year-End Report. In sum, the Justices’ conduct and statements regarding recusal have consistently indicated that the Court, as a matter of constitutional practice, considers recusal an individual matter to be decided independently by each Justice, free from congressional interference.

Constitutional structure offers a more complex analysis, but one that leads toward the same conclusion. There are no provisions in Article III that limit the Court’s exercise of its “judicial Power” over recusal. Neither of the two provisions that are commonly thought of as qualifying the judicial power, the provision empowering Congress to “ordain and establish” the inferior courts, and the clause enabling Congress to make “Exceptions” to the Court’s appellate jurisdiction, have any significant bearing on the Court’s power over recusal. The provision accounting for congressional creation of the inferior courts simply does not apply to the Supreme Court. The Exceptions Clause deals only with appellate jurisdiction, which is neither synonymous with recusal nor broad enough to include cases arising under the Court’s original jurisdiction. Finally, the Framers’ choice to empower Congress to affect Supreme Court justices to decide whether to recuse themselves than it is for lower court judges (who presumably are governed by § 455) due to a lack of replacements for the Justices. Id. at 2389, 2390.

123. See 2011 Year-End Report, supra note 3, at 3–5 (discussing the “fundamental difference between the Supreme Court and the other federal courts”).


125. U.S. Const. art. III, § 2, cl. 2.

126. Professor James Liebman and William Ryan have demonstrated that the Framers did not equate judicial power with jurisdiction. See Liebman & Ryan, supra note 96, at 708 ("We can confidently report . . . that . . . when Article III says 'judicial Power,' its drafters meant just that and not, e.g., 'jurisdiction' . . . ."). This is consistent with our common understanding of judicial authority; there are many instances where judges exercise their judicial authority that do not bear on whether the court has the power to decide the case before it. See Reed Elsevier, Inc. v. Muchnick, 130 S. Ct. 1237, 1243 (2010) ("Jurisdiction" refers to 'a court's adjudicatory authority.' Accordingly, the term 'jurisdictional' properly applies only to 'prescriptions delineating the classes of cases (subject-matter jurisdiction) and the persons (personal jurisdiction) implicating that authority.' (citations omitted) (quoting Kontrick v. Ryan, 540 U.S. 443, 455 (2004))). Recusal is precisely one of those instances.
appellate jurisdiction suggests a countervailing intent to preclude congressional intrusion into those exercises of the Judicial power—like recusal—that are not expressly subjected to congressional authority under Article III.127

The Impeachment Clauses in Article I128 are relevant to understanding the Justices’ constitutional authority over recusal because impeachment is the only textual authority for the removal of Article III judges, including Supreme Court Justices.129 This fact, however, triggers the interpretive canon of expressio unius est exclusio alterius (“to state the one is to exclude the other”); since recusal is not included along with impeachment as one of Congress’s constitutional powers over the Judiciary, it should not be treated as such.130 Separation of powers principles also preclude treating the Impeachment Clauses as relevant to the Court’s recusal powers. The severity and permanence of impeachment and the cumbersome procedures required to effectuate it are carefully designed to allow the Legislature to check judicial power without unduly threatening judicial independence.131 Permitting Congress to remove Justices from a

127. This argument evokes the interpretive canon of expressio unius est exclusio alterius (“to state the one is to exclude the other”).


129. Congress does of course have the authority to create official positions within the federal government that function as at-will employment (individuals who serve at the pleasure of the President) or that expire after a term of years, in the case of many independent agencies. In these cases, however, the separation of powers calculus is different because the officials being removed by means other than impeachment are not constitutionally guaranteed to hold those offices “during good Behaviour.” Id. art. III, § 1.

130. See McElroy & Dorf, supra note 60, at 99 (“[I]mpeachment is the only available remedy for clearly unethical decisions not to recuse . . . .”). That is not to say that other issues about which the Constitution is silent, such as judicial review, are all equally problematic. Unlike judicial power, congressional power is carefully enumerated in eighteen clauses in Section 8 of Article I. It is therefore easier to assume that an omission of an aspect of legislative power from this list of eighteen sources is more significant than the exclusion of a specific aspect of judicial power from Article III, which limits its explanation of the authority of the federal courts to its grant of the “judicial Power” to the Supreme Court and any “inferior courts Congress may from time to time ordain and establish.” U.S. CONST. art. III, § 1; id. art. III, § 2, cl. 2.

131. See U.S. CONST. art. III, § 1 (protecting judicial independence by providing federal judges with life tenure and salary protections).
particular case by a simple legislative mandate, rather than the full Article I impeachment process, threatens the independence of the Judiciary in a way not anticipated by constitutional text or structure. As a result, the Impeachment Clauses should be read in connection with Article III to support the proposition that recusal is constitutionally distinct from impeachment, and as such is an artifact of judicial, rather than legislative, constitutional authority.

The Necessary and Proper Clause grants Congress the power "[t]o make all Laws which shall be necessary and proper for carrying into Execution . . . all other Powers vested by this Constitution in the Government of the United States, or in any Department or Officer thereof." Unlike the provisions previously discussed, the Necessary and Proper Clause appears, at least facially, to grant Congress exceedingly wide latitude to regulate all aspects of government, including the courts. When read against the vesting clause of Article III, however, we see that Congress’s power under the Necessary and Proper Clause must end where the “inherent” judicial power bestowed upon the Court by Article III begins. If there is no realm of exclusively judicial power under the Constitution, then the Necessary and

132. That is not to say that Congress could not constitutionally employ its impeachment powers in an attempt to influence recusal decisions by Supreme Court Justices. See infra Part IV.A.1.

133. Congress may still be able to influence recusal standards in the lower courts. The point of introducing the Impeachment Clauses is simply to point out that by defining the impeachment power and assigning it to Congress, the Constitution’s structure offers insight into the proper constitutional home for judicial recusal. Impeachment is the only method explicitly provided for in the Constitution by which Congress may prevent an individual judge or Justice from performing his or her judicial duties under Article III. We can thus infer that the Constitution does not contemplate any additional congressional authority in the area, especially if those additional methods would expand congressional power to remove judges from cases at the expense of their obligations under Article III.


135. Id.

136. Id. art. III, § 1 (assigning the “judicial Power of the United States” to the Supreme Court). Of course Article III also bestows the “judicial Power” on any “inferior courts that Congress may . . . establish,” but because the lower federal courts are not the subject of this analysis, the present discussion will focus solely on the inherent power of the Supreme Court. Id.
Proper Clause renders Congress supreme among the three Branches of government. Because the absolute supremacy of any one branch is anathema to our constitutional system, there must be some area of judicial power that belongs exclusively to the Justices. The specific issue pertaining to Supreme Court recusal is thus whether the Necessary and Proper Clause empowers Congress to regulate the Court’s recusal practices, or whether recusal is within the Court’s “inherent power” under Article III such that Congress (along with other governmental entities outside the Court) is precluded from doing so.

137. See, e.g., The Federalist No. 47, at 264 (James Madison) (Colonial Press ed., 1901) (“The accumulation of all powers legislative, executive and judicial in the same hands, whether of one, a few or many, and whether hereditary, self-appointed, or elective, may justly be pronounced the very definition of tyranny.”).

138. The Court has confirmed this idea on several occasions throughout its history. See, e.g., Chambers v. NASCO, Inc., 501 U.S. 32, 47 (1991) (“[W]e do not lightly assume that Congress has intended to depart from established principles such as the scope of a court’s inherent power.” (quoting Weinberger v. Romero-Barcelo, 456 U.S. 305, 313 (1982))); Michaelson v. United States ex rel. Chi., St. Paul, Minneapolis & Omaha Ry. Co., 266 U.S. 42, 65–66 (1924) (explaining that the inherent contempt power of the courts cannot be “abrogated nor rendered practically inoperative” by Congress); United States v. Klein, 80 U.S. (13 Wall.) 128, 146–47 (1871) (explaining the importance of maintaining the separation of powers); United States v. Hudson, 11 U.S. (7 Cranch) 32, 34 (1812) (“Certain implied powers must necessarily result to our courts of justice from the nature of their institution.”).

139. The President and lower federal courts have no constitutional basis for asserting authority over Supreme Court recusal. See U.S. Const. art. III, § 1 (describing all federal courts other than the Supreme Court as “inferior” and therefore without power to command the Supreme Court); see also discussion infra note 155 (discussing the President’s lack of constitutional authority over recusal at the Court). The Judicial Conference of the United States Code of Conduct for United States Judges also does not apply to Supreme Court Justices. See Code of Conduct for United States Judges, Introduction, available at http://www.uscourts.gov/rulesandpolicies/codesofconduct/codeconductunitedstatesjudges.aspx (“This Code applies to United States circuit judges, district judges, Court of International Trade judges, Court of Federal Claims judges, bankruptcy judges, and magistrate judges.”). A broader consideration of whether recusal in the federal court system as a whole (including the lower federal courts) was part of the inherent judicial power vested by Article III would also need to consider the Tribunals Clause of Article I, by which Congress is expressly empowered “[t]o constitute Tribunals inferior to the supreme Court.” U.S. Const. art I, § 8, cl. 9. The Tribunals Clause, especially when read in conjunction with Article III’s vesting of the judicial power in “one supreme Court, and in such inferior Courts as the Congress may from time to time ordain
The precise boundaries of the Court’s inherent power remain an open question, but a narrow consensus has emerged among commentators and the courts around the central features of that power. In addition to concurring that some inherent judicial power exists, there is widespread agreement that this power must at least include the ability to independently and completely decide individual cases that fall within the jurisdiction of the federal courts. As a court of last resort, and one that is singled out in the text of Article III as unique among the federal courts as a whole, the Supreme Court provides the strongest constitutional case for retaining the inherent power to decide cases properly before it. Even when viewed in light of this

and establish,” is evidence that lower federal courts’ grant of the judicial power is subject to greater congressional control than that of the Supreme Court. Id. art. III, § 1, cl. 1; see id. art. I, § 8, cl. 9 (the Tribunals Clause). The issue is, however, beyond the scope of this analysis, which is only concerned with recusal at the Supreme Court.

140. As Professor James Liebman and William Ryan explained: [T]he judicial Power means the Article III judge’s authority and obligation, in all matters over which jurisdiction is conferred, independently, finally, and effectually to decide the whole case . . . . By independently, finally, and effectually decide, we mean dispositively to arrange the rights and responsibilities of the parties on the basis of independently developed legal reasons, subject to review only by a superior Article III court. By case, we mean a court action that can be resolved on the basis of enforceable law, and by whole case, we mean not only the construction of applicable provisions of law but also their actual application to the facts to reach a decision.

Liebman & Ryan, supra note 96, at 771 (emphasis removed) (internal quotation marks omitted); see also, e.g., Caminker, supra note 97, at 1519 (“[T]he core of the judicial power . . . is the authority to adjudicate and resolve Article III cases and controversies.”); David E. Engdahl, Intrinsic Limits of Congress’ Power Regarding the Judicial Branch, 1999 BYU L. REV. 75, 84–86 (stating that the vesting clause of Article III is “self-executing” with regard to issues of “judicial potency,” which includes the power to “adjudicate claims”); Frankfurter & Landis, supra note 96, at 1020 (explaining that Congress’s authority does not extend to regulation of fundamental features of courts such as the authority to independently and finally decide cases); Pushaw, supra note 98, at 741 (“[T]he constitutional provisions concerning congressional regulation of the judiciary do not pertain to the courts’ exercise of their essential function of adjudication.”).

141. See U.S. CONST. art. III, § 1 (vesting “[t]he judicial Power . . . in one Supreme Court, and in such inferior courts as the Congress may from time to time ordain and establish”).
(limited) understanding of inherent judicial power, Supreme Court recusal, at least the substantive standards for determining whether recusal is warranted,\(^\text{142}\) falls within the core of inherent judicial power under Article III. Because recusal precludes an individual Justice from participating in a case, the power to mandate recusal is the power to prevent any number of Justices—up to and including the entire Court—from exercising the core judicial function of deciding individual cases. A mandatory recusal statute could thus result in there not being enough Justices to decide a case that is otherwise properly before the Court. This would represent an unconstitutional intrusion into a core judicial power that is possible in every instance in which Congress seeks to provide substantive recusal standards for the Court.\(^\text{143}\) The Necessary and Proper Clause is thus not a valid source of congressional regulation of Supreme Court recusal because it has the significant potential to interfere with the power to decide cases that is inherent and exclusively vested in the Court by Article III.\(^\text{144}\)

\(^{142}\) There is one aspect of recusal regulation that does not fall as neatly within the Court’s inherent power to decide cases—procedural rules. There is a stronger argument to be made that congressional promulgation of procedural requirements for the Supreme Court’s recusal decisions do not interfere with the exceedingly narrow definition of inherent power advanced here. This fact does not, however, change the significance of the present inquiry or its relevance to the question of how a separation of powers analysis can help identify the most effective and constitutionally sound resolution to the current impasse over recusal at the Court. See Virelli, *supra* note 11, at 1223–25 (addressing the merits of procedural reform as a means of resolving the inter-branch impasse over recusal).

\(^{143}\) The most common arguments in favor of congressional power over recusal—that Congress could build in exceptions or “fixes” to the recusal statute, that Congress can regulate recusal because it has control over the size and qualifications of the Court, or that Congress can regulate recusal because it has power over the Court’s appellate jurisdiction all fall short when confronted with this inherent power argument. See Virelli, *supra* note 11, at 1221–22 (identifying the weaknesses in all three of the above arguments defending congressional involvement in Supreme Court recusal).

\(^{144}\) Because at first blush this argument may seem to beg questions about the constitutionality of more significant practices like judicial review, it is worth taking a moment to explain why the instant account of inherent judicial power does not apply equally to judicial interference with any inherent legislative authority of Congress. The answer lies in the fact that judicial review is limited to occasions in which Congress’s exercise of its legislative authority has run
It is not necessary that this constitutional analysis be beyond dispute for the subsequent discussion of Congress’s constitutional role in its current impasse with the Court over recusal. For present purposes, it is sufficient to acknowledge that there are legitimate constitutional questions about Congress’s authority to regulate recusal at the Court, and that constitutional uncertainty contributes significantly to the broader interbranch tensions over recusal. Those tensions include the normative question of how the interbranch conflict over recusal should be resolved within the context of our tripartite government.

C. Why is this Congress’s Problem?

Even if congressional regulation of Supreme Court recusal raises constitutional questions that contribute to the ongoing impasse between Congress and the Court over recusal, the question remains: Why is this Congress’s problem? Why not focus on the Court’s role in alleviating the impasse? The answer, like the question, lies in the separation of powers. The separation of powers depends, at least in part, on the political integrity of the individual Branches and the quality of the relationships among them. Any resolution to the interbranch conflict over recusal should therefore seek to promote a sense of responsibility and comity on behalf of both Congress and the Court. 145 When viewed

afoul of a constitutional provision or mandate. Congress’s power over the Judiciary under the Necessary and Proper Clause is not similarly limited, nor would it be the sole prerogative of the Legislature to establish when the Judiciary has run afoul of the Constitution even in cases in which Congress sought to prevent a constitutional violation; it is, after all, “emphatically the province and duty of the Judicial Department to say what the law is.” Marbury v. Madison, 1 U.S. (Cranch) 137, 177 (1803).

145. Professor Paulsen uses the term “coordinacy” to describe the same concept:

The “coordinacy” of the three branches of the federal government is one of the fundamental political axioms of our federal Constitution. . . . This does not mean that the branches are equal in the quantum of powers assigned to them. . . . Coordinacy is a term of power-relationship, not of power-scope. . . . It is the idea of coordinacy, even more than the cognate concept of separation on which it depends and builds, that fuels the system of “checks and balances” that guards against “a tyrannical concentration of all the powers of government in
in this light, it becomes clear that although the Court possesses the constitutional authority to resolve the recusal conflict, the mechanisms at its disposal are far more potentially damaging to the Court as an institution and to our constitutional system in general than those available to Congress.

The Court has at least three measures at its disposal to resolve its impasse with the Legislative Branch over recusal, none of which are ultimately satisfactory when considered in light of the separation of powers. One is to state openly and explicitly that it will not comply with any statutory mandates regarding recusal because of constitutional concerns, issues of judicial integrity, or some other institutional reason. This approach would bring transparency and accountability to the Court’s position and would alleviate concerns that the Court is eluding statutory standards for some less compelling reason like political gain or mere convenience. This approach has, however, some considerable weaknesses in the context of the separation of powers. For one, without grounding its statement in legally binding authority, the Court could be made to look more obstinate than it would in the absence of a public statement. More importantly, a public statement of noncompliance may exacerbate, rather than help resolve, the conflict. Chief Justice Roberts’s Year-End Report, which is the closest example to date of a public explanation of the Court’s views on the recusal impasse, provoked a largely negative reaction. It prompted a letter from five Senators asking the Court to voluntarily adopt formal recusal standards.146 Professors Yeomans and Schwartz argued that “Chief Justice John Roberts’s response in his year-end report to the increasing controversy over the ethics of Supreme Court Justices served to drive home the need for the high court to adopt reforms immediately,”147 and Professor Ifill characterized his statement as “far short of an adequate

\textsuperscript{146.} See Senator Letter, supra note 4.
\textsuperscript{147.} Yeomans & Schwartz, supra note 4.
response,” primarily because it did not increase the “transparency of [recusal] procedures that go to the very legitimacy of the Court’s decisionmaking.” These critiques highlight the shortcomings of any mere statement by the Court that it will not comply with statutory recusal standards. In the current debate over recusal, an explanation of the status quo—even one based on high-minded principles like the separation of powers or institutional competence—will simply not suffice.

Alternatively, the Court could use its power as the final expositor on the Constitution to rule that the impasse between Congress and the Court over recusal is not legally justiciable. This relatively common approach in separation of powers cases often leads to positive results. The War Powers Resolution, for instance, has long been a source of interbranch controversy. A debate has persisted for decades regarding whether the Resolution is a constitutional limit on the President’s authority to use military force as Commander in Chief under Article II. Although several members of Congress have expressed concern about various presidential decisions in relation to the War Powers Resolution, there have never been any attempts to judicially enforce it against the President, and several

149. See id. (“I found Justice Roberts’ defense of the status quo in the Supreme Court’s recusal practice to be the most unsatisfying aspect of his remarks.”).
151. See U.S. CONST. art. II, § 2.
152. Members of Congress raised concerns under the War Powers Resolution during President Clinton’s bombing of Kosovo in 1999 and President Obama’s bombing of Libya in 2011, and in both instances there was much public and scholarly debate over the constitutional basis for those objections. See, e.g., Geoffrey S. Corn, Clinton, Kosovo, and the Final Destruction of the War Powers Resolution, 42 WM. & MARY L. REV. 1149, 1153 (2001) (discussing the debate over the constitutionality of the War Powers Resolution); Jack Goldsmith, War Power: The President’s Campaign Against Libya is Constitutional, SLATE.COM (Mar. 21, 2011, 6:48 PM), http://www.slate.com/id/2288869/ (last visited Sept. 18, 2012) (same) (on file with the Washington and Lee Law Review). In neither case, however, was there any serious threat of legal action by Congress to enforce the requirements of the Resolution.
153. See CHARLES A. SHANOR, AMERICAN CONSTITUTIONAL LAW: STRUCTURE AND RECONSTRUCTION 208 (4th ed. 2009) (“The requirements of the War Powers Resolution were never tested in court. . . . No judicial decision has ever
Presidents have employed military force without strictly adhering to its requirements or challenging the statute in court. But the fact that the conflict between Congress and the Executive over the War Powers Resolution has never been substantively addressed by the Court does not mean that the Resolution is constitutionally problematic. Whether it refuses to do so for undisclosed policy reasons or because it formally concludes that the dispute is nonjusticiable as a matter of constitutional law, the presence of the Court as a neutral arbiter makes the lack of a legal resolution between the other two Branches more constitutionally palatable. A decision by the Court not to intervene legitimizes the otherwise seemingly unruly process of political interchange and compromise between Congress and the President by making it part of the constitutional design, thus creating political space for the Legislative and Executive Branches to better deal with complex questions about the use of military force on a case-by-case basis.

The conflict over Supreme Court recusal, however, would not similarly benefit from a decision by the Court that it is without constitutional power to resolve the dispute. Unlike with the War Powers Resolution, there is no third-party mediator for the

---

154. As Professor Jeffrey Tulis explained:

In the aftermath of the Vietnam War, Congress passed a constitutionally aggressive statute, the War Powers Resolution, which seemed to display just the sort of institutional turf protection that the Federalist theory describes. Nevertheless, since the passage of the Resolution, the President has violated its terms repeatedly without challenge from Congress.

interbranch conflict over recusal at the Court.\footnote{There is no constitutional basis for the President to intervene in a dispute between Congress and the Court over recusal, and review by any court other than the Supreme Court would run afoul of Article III’s hierarchy of “one Supreme Court” and other “inferior” courts. See Russell Wheeler, Regulating Supreme Court Justices’ Ethics—“Cures Worse than the Disease?”, BROOKING INST. (Mar. 3, 2011), http://www.brookings.edu/opinions/2011/0321_justices_ethics_wheeler.aspx?p=1 (last visited Sept. 18, 2012) (discussing the constitutional concerns surrounding proposals calling for the creation of a court to review recusal decisions) (on file with the Washington and Lee Law Review); see also McElroy & Dorf, supra note 60, at 107–12 (considering the meaning of Article III’s “one supreme Court” requirement in the context of reviewing Supreme Court recusal decisions).} As a result, the same type of political exchange and compromise—such as the Justices citing to the statute in their recusal memoranda despite neither feeling bound by it nor applying it with the same force or rigor as they would other statutes\footnote{See supra notes 111–19 and accompanying text (outlining the Court’s prior practices of analyzing the federal recusal statute differently from other statutes).}—loses the imprimatur of legitimacy that comes from direct or indirect endorsement of the process by a coequal branch of government. Instead, the Justices’ failure to comply makes them appear unprincipled, regardless of the true quality of their decisions. Thus, due to the seemingly inevitable unenforceability of congressional mandates regarding Supreme Court recusal and the absence of a constitutionally recognized third-party arbiter, a constitutional decision by the Court that the dispute is nonjusticiable is not a satisfactory resolution under the separation of powers.

The third way the Court could break the impasse over the Justices’ recusal practices is to use its power of judicial review to invalidate the statutory recusal standards on the basis that they are unconstitutional interferences with the judicial power vested in the Court by Article III.\footnote{For a more detailed treatment of this constitutional issue, see supra Part III.B, and see generally Virelli, supra note 11 (stating that separation of powers principles prevent Congress from directly regulating Supreme Court recusal).} This would amount to essentially deleting the word “justices” from the existing statute and returning it to its pre-1948 status, when it purported to cover all federal judges, but not Supreme Court Justices.\footnote{Compare 28 U.S.C. § 455 (2006) (governing disqualification of Justices,}
the most intuitive response to the problem, as we generally think—at least since Chief Justice Marshall’s famous pronouncement in Marbury—that matters of constitutional interpretation lie finally, and most comfortably, with the Judiciary. Despite its intuitive appeal, this approach is seriously flawed when viewed from a separation of powers perspective.

As a general matter, Supreme Court review and invalidation of federal statutes is an unremarkable and fundamental feature of our constitutional existence. The stakes become a bit higher, however, when the statutes under review regulate not subordinate government actors or the public at large, but the highest levels of a coordinate branch. In those instances, questions of interbranch comity and cooperation become of paramount importance to ensure the continued functioning and legitimacy of the constitutional design. In the context of congressional regulation of Supreme Court recusal, concerns about comity and legitimacy counsel strongly against the Court striking the recusal statute.

Some examples illustrate the effect of comity concerns on the Court. In Myers v. United States, the Court considered a question dividing the other two Branches of government—whether Congress could limit the President’s authority to remove an executive official from office. The conflict had all the makings of an impasse: a postmaster who had been removed from office brought suit against the United States alleging that President Wilson’s decision to fire him ran afoul of a federal statute requiring the President to obtain “the advice and consent of the Senate” before removal. The President did not claim that the

judges, and magistrate judges), with Act of May 8, 1792, ch. 36, § 11, 1 Stat. 275, 278–79 (regulating only federal “judges,” rather than “justices”).

159. Chief Justice Marshall did, after all, famously explain in Marbury that it is “emphatically the province and duty of the judicial department to say what the law is.” Marbury v. Madison, 1 U.S. (Cranch) 137, 177 (1803).


Senate had consented to the postmaster’s removal under the terms of the statute, but simply that Congress was without constitutional authority to interfere with the Executive’s prerogative to manage its workforce. The Supreme Court took up the issue and held that the separation of powers precluded Congress from interfering with the President’s decision to remove executive officers at will. The statute was invalidated, and although not uncontroversial, the decision brought relative clarity to both Branches. Less than a decade later, the Court again addressed the removal issue, and held that Congress could limit the Executive’s removal authority in the context of independent agencies. Again, the Court’s conclusion was adopted as the authoritative solution to the interbranch conflict. Since then, the Court has addressed additional removal issues without any challenges to its institutional authority or capacity to do so. These examples show that the Court can be well-suited to resolve disputes between Congress and the Executive, and in fact may be an important bulwark against undue conflict between those Branches in its role as a neutral constitutional arbiter.

The circumstances change dramatically, however, when the Court becomes involved in constitutional review of statutes that are targeted at the Court itself or the Justices. The most direct example of this phenomenon is in cases that have come before the Court under the Exceptions Clause of Article III. The

163. See Myers, 272 U.S. at 51 (discussing the President’s power to remove executive officers).
164. See Fisher, supra note 154, at 60–64 (discussing Chief Justice Taft’s decision in Myers and the subsequent reactions thereto).
165. See Humphrey’s Executor v. United States, 295 U.S. 602, 629 (1935): The authority of Congress, in creating quasi legislative or quasi judicial agencies, to require them to act in discharge of their duties independently of executive control cannot well be doubted; and that authority includes, as an appropriate incident, power to fix the period during which they shall continue in office, and to forbid their removal except for cause in the meantime.
166. See Free Enter. Fund v. Pub. Co. Accounting Oversight Bd., 130 S. Ct. 3138, 3164 (2010) (holding that the particular arrangement of “good cause” requirements for removal of members of the Public Company Accounting Oversight Board violated separation of powers principles, as the President did not have the authority to oversee the officials within the Executive Branch).
167. The Exceptions Clause is the best example here because it is the only
Exceptions Clause expressly empowers Congress to provide “Exceptions” or “Regulations” regarding the Supreme Court’s appellate jurisdiction. The Clause is the only constitutional provision that is aimed expressly and exclusively at the relationship between Congress and the Supreme Court; it does not empower any other government actor to influence the Court’s appellate jurisdiction, and does not permit Congress to do anything with respect to any other constituency of the Judicial Branch. The Exceptions Clause is thus the purest analog for the impasse between Congress and the Court over recusal, and as such provides useful insights into how the Court views its role as constitutional interpreter differently when the Court’s own institutional interests are at stake.

The Court’s Exceptions Clause jurisprudence is the subject of longstanding debate, but one feature of that jurisprudence—and the most important feature for present purposes—is clear: The Court has never invalidated a statute solely on the basis that it overstepped Congress’s authority under the Exceptions Clause. The Court has upheld congressional authority in all but one case interpreting the Clause, albeit often narrowly and without conceding that the Court’s jurisdiction is entirely precluded. In constitutional provision that gives Congress authority over just the Supreme Court and not the other parts of the Judicial Branch. There are of course other—in fact significantly more common—instances in which the Court engages in judicial review of statutes that raise conflicts between the Legislative and Judicial Branches. See, e.g., CFTC v. Schor, 478 U.S. 833, 835–36 (1986) (resolving whether the statutory authorization of the Commodity Futures Trading Commission to adjudicate certain claims unconstitutionally interfered with the authority of Article III judges). In terms of drawing analogies to the impasse over Supreme Court recusal, it is important to focus on the relationship not between Congress and the federal courts in general, but Congress and the Supreme Court. While there may be reasons of comity and legitimacy that prevent the Court from resolving disputes between the Legislature and the Judiciary as a whole, those problems are magnified significantly when the Court’s decision deals solely with its own authority vis-à-vis Congress. Because the instant analysis focuses solely on recusal practices at the Supreme Court, it is more instructive to draw analogies from equally narrow conflicts. The only constitutional conflict that is as narrow as—and is thus properly analogized to—the conflict between Congress and the Court under the Exceptions Clause.

the lone case in which the Court did strike a statute under the Exceptions Clause, United States v. Klein, the Court went to great lengths to offer alternative explanations for its holding. Klein struck down a statute that required the Court to dismiss for lack of jurisdiction any suit to recover proceeds from property seized during the Civil War in which the claimant’s entitlement to the property was based on a presidential pardon. The Court importantly acknowledged Congress’s power to limit the Court’s appellate jurisdiction. It then declared the statute unconstitutional on the grounds that it impermissibly attempted to direct the outcome of specific cases in violation of the judicial power vested in the Court by Article III, and interfered with the President’s pardoning power under Article II. Taken

statute against an Exceptions Clause challenge); Ex parte Yeager, 75 U.S. 85, 106 (1869) (same); Ex parte McCordle, 74 U.S. 506, 514–15 (1869) (same).


171. See id. at 145 (“If it simply denied the right of appeal in a particular class of cases, there could be no doubt that it must be regarded as an exercise of the power of Congress to make ‘such exceptions from the appellate jurisdiction’ as should seem to it expedient.”).

172. The Klein Court explained that:

[T]he language of the proviso shows plainly that it does not intend to withhold appellate jurisdiction except as a means to an end. Its great and controlling purpose is to deny to pardons granted by the President the effect which this court had adjudged them to have. . . .

It is evident from this statement that the denial of jurisdiction to this court . . . is founded solely on the application of a rule of decision, in causes pending, prescribed by Congress. . . . It seems to us that this is not an exercise of the acknowledged power of Congress to make exceptions and prescribe regulations to the appellate power. The court is required to ascertain the existence of certain facts and thereupon to declare that its jurisdiction on appeal has ceased, by dismissing the bill. What is this but to prescribe a rule for the decision of a cause in a particular way? . . . We must think that Congress has inadvertently passed the limit which separates the legislative from the judicial power.

Id. at 145–47.

173. See U.S. Const. art. II, § 2, cl. 1 (“[The President] shall have Power to grant Reprieves and Pardons for Offenses against the United States.”). The Klein Court went on to explain that:

The rule prescribed is also liable to just exception as impairing the effect of a pardon, and thus infringing the constitutional power of the Executive. . . . To the executive alone is intrusted the power of pardon; and it is granted without limit. . . . Now it is clear that the
together, the Court’s Exceptions Clause jurisprudence demonstrates the Court’s reluctance to assert its own interests over those of Congress. It has generally affirmed Congress’s authority to limit the Court’s appellate jurisdiction, and even when it found it necessary to invalidate a statute it relied heavily on the countervailing interests of the Executive in doing so.

The purpose of the Exceptions Clause example is not to take a position as to the quality of the Court’s reasoning or conclusions in those cases. It is offered rather as support for the narrow proposition that the Court exercises its power of judicial review more cautiously when addressing conflicts between itself and another branch of government, as in the Supreme Court recusal context. While the Exceptions Clause analogy is a strong one, it is limited both by its relatively small number of cases and the fact that there are no corroborating examples, primarily because no other constitutional provisions single out the Court as the sole subject of regulation. Nevertheless, a normative analysis of the Court’s approach to resolving conflicts between itself and the other Branches confirms the impression left by the Court’s Exceptions Clause cases, and counsels strongly against the Court using its power of review to overturn the recusal statute.

As an initial matter, cooperation and comity among the coordinate Branches of government are necessary to the effective functioning of our tripartite constitutional arrangement. The Branches are coequal and necessarily interactive, and legislature cannot change the effect of such a pardon any more than the executive can change a law. Yet this is attempted by the provision under consideration. . . . This certainly impairs the executive authority and directs the court to be instrumental to that end.

Klein, 80 U.S. at 147–48.

174. See Felker, 518 U.S. at 654 (stating that the Act at issue did not violate the Exceptions Clause); Ex parte McCordle, 74 U.S. at 513 (stating that the Constitution gives Congress the power to “mak[e] exceptions to the appellate jurisdiction of the Supreme Court” (citation omitted)).

175. See Klein, 80 U.S. at 147–48 (discussing the potential infringement on the power granted by the Constitution to the Executive).

176. See Paulsen, supra note 93, at 228–29 (describing the need for coordinacy among the three Branches of the federal government); Entin, supra note 145, at 226 (explaining that “[t]he separation of powers assumes a minimum level of interbranch comity”).

collaborate on nearly every issue of national importance such as the national budget, military readiness and deployment, social policy questions, and law enforcement initiatives. Although each branch is equipped with powerful constitutional checks designed to maintain a sense of equilibrium and balance within the federal structure, the use of those checks in a confrontational or unwelcome way may exchange short-term gains for long-term damage to governmental efficiency and productivity. For this reason alone, the Court should be cautious in exercising its powers of judicial review, especially in the context of resolving disputes in favor of one branch of government over another. As Professor Bruce Peabody and John Nugent explained:

[We] do not suggest that the judiciary should never intervene in separation of powers conflicts. But we do think this intervention should be infrequent [and] restrained . . . . The judiciary should, to the best of its ability, resist efforts to become embroiled in interbranch disputes while they are still unfolding . . . . [W]hen the judiciary does intervene in disagreements over the authority or powers of the different divisions of government, it should . . . address how its ruling will affect the various levels at which the separation of powers operate.178

The remaining normative arguments against the Court striking the recusal statute are perfectly in line with the concerns expressed by proponents of recusal reform—protection of the Court’s public reputation and, in turn, its legitimacy as the final expositor of constitutional law. As Justice O’Connor explained, “The Court’s power lies, rather, in its legitimacy, a product of

imposes upon the Branches a degree of overlapping responsibility, a duty of interdependence as well as independence the absence of which ‘would preclude the establishment of a Nation capable of governing itself effectively.’” (citation omitted); id. at 703 (“As Madison explained, separation of powers does not mean that the Branches ‘ought to have no partial agency in, or no control over the acts of each other.’” (quoting THE FEDERALIST No. 47, at 266 (James Madison) (Colonial Press ed., 1901))); Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 635 (1952) (Jackson, J., concurring) (“While the Constitution diffuses power the better to secure liberty, it also contemplates that practice will integrate the dispersed powers into a workable government. It enjoins upon its branches separateness but interdependence, autonomy but reciprocity.”).

substance and perception that shows itself in the people’s acceptance of the Judiciary as fit to determine what the Nation’s law means and to declare what it demands.” 179 She went on to explain that “[t]he underlying substance of this legitimacy . . . is expressed in the Court’s opinions, and our contemporary understanding is such that a decision without principled justification would be no judicial act at all.” 180 Striking the portion of the recusal statute that governs the Justices invites criticism that the Court is aggrandizing power at the expense of at least one of the political Branches. This criticism, especially when combined with parallel critiques that the Court is becoming overly politicized, 181 could undermine public confidence in the institution and its fitness not only to adjudicate, but also to fulfill its role under the separation of powers as a check on the other Branches. The Court’s legitimacy could be similarly imperiled by a decision striking regulations on Supreme Court recusal if such a decision appeared unprincipled. Particularly in an area such as Supreme Court recusal and the separation of powers, in which the constitutional text, history, and judicial precedent are at best sparse, a decision in which the Court favors its own authority over that of another branch—even in the face of a written opinion explaining the decision—could be seen as pretextual and thus democratically illegitimate. Finally, the Court’s Exceptions Clause jurisprudence provides yet another reason why the Court should refrain from using its power of review in the debate over recusal. Whereas application of the recusal statute could harm litigants by precluding them from receiving an otherwise constitutionally-provided level of judicial review in a single case, 182 the application of a jurisdiction-stripping statute is almost certainly more likely to bar review in a wider array of

180. Id.
181. See supra note 8 and accompanying text (noting recent critiques of the Justices’ political associations and interactions).
182. See supra Part III.B (outlining the constitutional difficulties with congressional influence over Supreme Court recusal based at least in part on the fact that a recusal statute threatens to preclude the Court from exercising its inherent authority to decide individual cases).
cases. Nevertheless, despite their potential for harm, the Court has been extremely reluctant to strike jurisdiction-stripping statutes under the Exceptions Clause. To the extent that the harmful impact on litigants of a jurisdiction stripping statute is greater than that of a recusal statute, there is even less reason for the Court to overturn the latter.

Because the Court cannot remedy its impasse with Congress without doing precisely the damage to its institutional reputation and legitimacy that proponents of recusal reform seek to avoid, the task lies with Congress to employ the non-legislative tools at its disposal to alleviate the interbranch tension over recusal.

IV. Congress as an Effective Constitutional Actor

The fact that regulatory approaches to Supreme Court recusal raise constitutional concerns about the bounds of congressional authority does not mean that Congress is without potentially effective methods to remedy its impasse with the Court. By viewing the question of Supreme Court recusal as a matter of constitutional structure, we can more clearly identify ways that Congress may legitimately exercise its non-regulatory authority to address concerns over the Justices’ recusal practices without the countervailing harm to judicial and congressional legitimacy that results from regulatory intervention.

A. Congress’s Indirect Constitutional Tools

This analysis reveals five indirect constitutional approaches for Congress to use in influencing Supreme Court recusal practice within the confines of a legitimate tripartite governmental structure—impeachment, procedural reform, judicial confirmation, appropriations, and investigation.

1. Impeachment

Congress’s impeachment power is perhaps its strongest means of curtailing perceived recusal abuses by the Justices. The benefits of impeachment in the Supreme Court recusal context
are its clear constitutional legitimacy as a legislative check on the Judiciary and its effectiveness, both as a response to past recusal misconduct and as a deterrent against future misconduct by the Justices. First, impeachment is a valuable tool for Congress in the recusal context because it is clear as a constitutional matter. In fact, questions of impeachment are best understood as political questions, dedicated exclusively to the Legislature by Article I.¹⁸³ Unlike direct regulatory limits on Supreme Court recusal, impeachment is not a separation of powers problem, but rather an anticipated and explicitly prescribed potential solution. Impeachment is also attractive because of the nature of the remedy; by removing a Justice from office, any future problems with that Justice’s recusal practices will be alleviated. It likewise stands to reason that the specter of impeachment will work as a deterrent, encouraging Justices to conform their recusal practices to those norms not considered impeachable by Congress.

Impeachment’s primary shortcoming in the recusal context is its lack of constitutionally mandated criteria. The converse of impeachment’s strong claim to legitimacy as a legislative check on the Judiciary as an institution is the potential legitimacy problem arising from a specific exercise of the impeachment power over an individual Justice. Article III guarantees the Justices their “Offices during good Behavior,”¹⁸⁴ but that term is otherwise undefined in the constitutional text, and Article I provides little or no guidance as to how or when impeachment proceedings should be instituted.¹⁸⁵ Impeachment proceedings against a Justice for any reason, including recusal issues, would require significant explanation by Congress to avoid the public perception of overreaching and to avoid the risk of lowering the bar for the impeachment of federal judges, especially Supreme Court Justices, to the detriment of the Judiciary’s ability to function as a coequal branch of government. These dangers, however, while significant, are neither unique to Congress’s


¹⁸⁵. One feature of impeachment proceedings that is clear from the constitutional text is that they must initiate in the House. See id.
exercise of the impeachment power nor fatal to impeachment’s potential usefulness for recusal. Arbitrariness and overuse are no more present in impeachment than in any other largely discretionary exercise of authority under the Constitution. The power to declare war, for example, does not come with any easily cognizable objective legal standards limiting its application; it is a political decision by a political branch that is checked only by electoral and political processes.186 The President’s veto power is a similarly discretionary act that is checked not by constitutional restraints on the exercise of that power, but by political and electoral checks on the President.187 Impeachment enjoys similar political limitations, as well as the additional constraint of requiring two separate and coordinated efforts by both Houses of Congress.188 By dividing the power of removal through impeachment between the two Houses, impeachment is better protected from abuse and overuse than other discretionary constitutional acts. Finally, the lack of demonstrable standards does not make impeachment impossible to employ as a productive and valuable, even if rarely used, tool for influencing Supreme Court recusal. Federal judicial impeachment has rarely been used by past Legislatures.189 There is no reason to believe that future Congresses will have difficulty deciding how impeachment may be used to affect the Justices’ recusal decisions without unduly damaging itself, the Court, or the separation of powers more generally. In fact, the severity of the impeachment remedy may cause it to be used even more sparingly. In short, because impeachment is constitutionally dedicated to Congress and holds substantial promise as an effective (if rarely used) remedy, it remains a viable tool for congressional involvement in the debate over Supreme Court recusal.

186. See id. art I, § 8, cl. 11 (giving Congress the power “[t]o declare [w]ar”).
187. See id. art I, § 7. The possibility of a congressional override is of course a check on the effectiveness of the President’s veto power, but not on his constitutional authority to choose to issue the veto in the first instance.
188. See U.S. Const. art I, § 2, cl.5; id. art. I, § 3, cl. 6.
189. See Report of the National Commission on Judicial Discipline and Removal 29–30 (1993) (noting that since the Founding, only eleven judges have been tried in impeachment proceedings, and only seven have been convicted).
2. Procedural Reform

A second example of congressional authority over the Justices’ recusal practices lies in Congress’s authority to promulgate procedural requirements for the federal courts. The operative word in this category is *procedural* in light of the working assumption that the separation of powers precludes direct congressional regulation of the Justices’ substantive recusal decisions. With that caveat in mind, there are several ways in which Congress could seek to regulate Supreme Court procedure. The first is through the Exceptions Clause of Article III. The Exceptions Clause could be used by Congress to regulate or “strip” appellate jurisdiction from the Court in cases in which, for instance, a certain number of recusals occurred, or conversely in which a certain number of recusal motions were denied. The Exceptions Clause could thereby serve as an incentive to encourage or discourage recusals, as Justices would presumptively be less likely to recuse when doing so would imperil the Court’s jurisdiction over a case, and more likely to do so when it would not. Nevertheless, although theoretically straightforward under Article III, use of the Exceptions Clause is not uncontroversial, especially when, as in the above recusal examples, it could effectively eviscerate the Court’s otherwise valid appellate jurisdiction. The Court’s Exceptions Clause jurisprudence has never upheld a statute that has the effect of closing off Supreme Court review altogether, and there are potential separation of powers problems associated with Congress taking that step. Moreover, stripping jurisdiction in reaction to the denial of motions to recuse could incentivize meritless recusal motions by litigants who were successful below. An additional requirement aimed to remedy this problem, for instance counting

---

190. The precise opposite motivations could also be true, namely when a Justice who is not enamored of a case could base their recusal decision at least in part on how that decision would affect the case’s viability, but I think this is a far less likely scenario, especially when the alternative is to vote on the merits of the case.

191. See, e.g., ERWIN CHERERINSKY, FEDERAL JURISDICTION § 3.2, at 182 (5th ed. 2007) (explaining that congressional attempts to limit the Court’s appellate jurisdiction have “never . . . been interpreted as precluding all Supreme Court review”).
only nonfrivolous motions to recuse in the jurisdictional analysis, highlights another problem with the use of the Exceptions Clause to influence recusal at the Court—the ultimate unenforceability of the standards. As long as the Court remains the final word on the Exceptions Clause, and there is no constitutional basis to conclude that it would not, Congress’s authority to limit the Court’s appellate jurisdiction remains dependent on the Court’s own reading of the Clause. While the Court would not lightly strike such a statute,\textsuperscript{192} it may attempt to protect some measure of its own authority (and litigants’ access to it) through the same statutory interpretation and constitutional line-drawing techniques it has employed in prior Exceptions Clause cases.\textsuperscript{193} Finally, the Exceptions Clause is under-inclusive in the recusal context because it cannot reach cases involving the Court’s original jurisdiction.

One possible counter to the Exceptions Clause’s under-inclusiveness is Congress’s power over the Court’s quorum requirements. Quorum standards are not limited to cases arising under the Court’s appellate jurisdiction. They also do not necessarily pose a separation of powers problem.\textsuperscript{194} Quorum values could be used to either encourage or discourage recusals. A high quorum requirement would likely encourage participation

\textsuperscript{192} See discussion supra Part III.C (discussing the importance of inter-branch comity and cooperation to the constitutional separation of powers).

\textsuperscript{193} See, e.g., \textit{Ex parte McCardle}, 74 U.S. (7 Wall.) 506 (1868) (interpreting a jurisdiction-stripping statute to still permit appellate review on alternative statutory grounds); \textit{Ex parte Yerger}, 75 U.S. (8 Wall.) 85 (1868) (employing the alternate basis for appellate jurisdiction suggested in \textit{McCardle}).

\textsuperscript{194} Quorum standards are less likely to create separation of powers problems when they do not set minimum participation requirements for the Justices that are unattainable, either due to the number of Justices currently sitting on the Court, or the total number of Justices authorized to serve. For example, quorum requirements almost certainly would create a separation of powers problem if they set the minimum number of Justices required to decide a case as higher than the total number of Justices authorized to sit on the Court (envision a quorum requirement of 10 on the present Court). A similar problem would arise if the quorum number was greater than the number of Justices available to serve at any one time, due to an unfilled vacancy or a prolonged absence. In those instances, I would argue that Congress has a constitutional duty to either lower the quorum requirements or to fill the existing vacancies by confirming additional Justices in order to honor the Article III requirement that there be “one supreme Court.” \textit{U.S. Const.} art. III, § 1.
and a lower number could decrease the perceived cost of a Justice’s decision to recuse. While they are limited by the fact that the substantive decision to recuse is ultimately left to the Justices themselves, and as such permits easy circumvention of any quorum requirements, the absence of regulatory recusal standards at minimum makes the Justices accountable for a decision to circumvent quorum in a way they were not under a regulatory recusal regime.

Finally, Congress could require that the Justices follow specific procedural steps in making and issuing their recusal decisions under its traditional, inherent power to regulate the procedures of the federal courts. Procedural reforms, like a requirement that the Justices publish explanations of their recusal decisions, or that those decisions be subject to review by the entire Court, are popular among reformers concerned about the public perception and legitimacy of the Court, and admittedly are less susceptible to constitutional arguments invoking the separation of powers than substantive recusal requirements. This does not, however, mean that procedural requirements are without limitations. The suggested procedural reforms are not only ultimately unenforceable against the Justices, but their anticipated benefits may also be diminished by the absence of parallel substantive reforms. If the Court is

195. This is precisely what happened in North American Co. v. SEC, 327 U.S. 686 (1946). Chief Justice Stone originally recused himself from the case, only to reverse his decision and participate when he realized his recusal could defeat a quorum.

196. This authority could also be grounded in the Necessary and Proper Clause of Article I, Section 8, but, for these purposes, the precise constitutional basis is unimportant. It is adequate for the present discussion to acknowledge that Congress retains constitutional authority to influence—at least to some degree—the procedures of the federal courts, including the Supreme Court.

197. See discussion supra notes 49–56 and accompanying text (discussing the criticism of current Supreme Court recusal practices and the suggestions for various types of reform).

198. See Virelli, supra note 11, at 1223–25 (discussing potential methods of procedural recusal reform).

199. To the extent supporters of statutory procedural recusal reform value the potential symbolic or persuasive effects of those reforms on the Supreme Court, any such benefit is arguably outweighed by the costs of creating a seemingly irresolvable conflict between two Branches of government.
constitutionally protected from congressional interference in deciding whether a Justice will ultimately be recused from a case, the procedural framework in which that decision is made is not likely to shed significantly more light on that decision. For example, requirements that Justices publish their reasons for failing to recuse themselves could, in the absence of defined, binding criteria for recusal decisions, do little to promote the integrity or public perception of the Court because there will be no baseline against which to measure the quality of the Justice’s explanations. It is also difficult to imagine how, in cases where a Justice would not voluntarily choose to publicly explain their decision, an unenforceable reporting requirement would result in anything more than a cursory statement by the Justice. A similar problem arises if the procedural requirement consists of the full Court reviewing an individual decision of one of its members. As evidenced by the fact that only once in its history has the Court experienced a public dispute between its members over recusal, it is unlikely that internal Court review of its members’ recusal decisions would lead to any useful insight into either the decision under review or the Court’s feelings about recusal more broadly. Nevertheless, even if procedural reforms do not drastically change the outcome of individual recusal decisions, they can serve to increase the transparency of the process by either increasing the amount of substantive information provided by the Justices in their recusal decisions, or, at minimum, shifting responsibility for providing that information (and thus blame for not providing it) from Congress to the Justices.

Procedural reform is an example of an indirect constitutional tool that is far greater than the sum of its parts. Despite the limitations of each of the proposed reforms, the impact of procedural reform in general offers potentially significant benefits in terms of transparency and accountability, both of which could enhance the Court’s public perception and legitimacy.

200. See Jewell Ridge Coal Corp. v. Local No. 6167, United Mine Workers, 325 U.S. 897, 897 (1945) (Jackson, J., concurring) ("There is no authority known to me under which a majority of this Court has power under any circumstances to exclude one of its duly commissioned Justices from sitting or voting in any case.").
without the expense of the constitutional impasse created by direct congressional regulation of recusal standards.

3. Judicial Confirmation

The Senate’s power over judicial confirmations is another means by which Congress can influence Supreme Court recusal. Unlike its legislative authority, which is limited by Article I as well as competing constitutional provisions, the Senate’s power to confirm is seemingly unconstrained as a constitutional matter. Moreover, Supreme Court confirmation hearings have become increasingly detailed and substantive. As Professors Lori Ringhand and Paul Collins demonstrated in their exhaustive empirical study of the last seven decades of Supreme Court confirmation hearings, “it is evident there was a steady increase in the amount of dialogue that transpires at the hearings” since 1939, and “substantive issues . . . have long dominated the hearings.”

There is also no reason to believe that Senators’ questions regarding a nominee’s views on recusal would not be answered. Although many of the Senators’ inquiries about specific and controversial areas of the law are met with generic and noncommittal responses by the nominee in order to avoid appearing as if they have prejudged issues that could come before the Court, questions about a potential Justice’s views on judicial recusal would be largely immune from such an objection. Recusal questions are technically not the subject of cases before the Court, as they are committed entirely to an individual Justice’s judgment. They are more akin to questions about judicial philosophy, which is a popular topic at confirmation hearings.

201. See U.S. Const. art. II, § 2, cl. 2 (stating only that the President “shall nominate, with the Advice and Consent of the Senate . . . Judges of the Supreme Court”).


203. See id. at 617–18 (“Judicial philosophy is the third most frequently occurring issue following chatter and civil rights. Comments about judicial philosophy, which include such things as discussions of constitutional interpretation, stare decisis and judicial activism, constitute 12.4% of the
and has not been treated as objectionable by the nominees; in fact, Chief Justice Roberts’s most memorable statement from his own confirmation hearing explained his jurisprudential philosophy by analogizing judges to baseball umpires.\textsuperscript{204} In light of the Chief Justice’s Year-End Report, it is clear that the Court is equivocal about which, if any, specific standards govern their recusal decisions. Without attempting to impose binding legal requirements on those decisions—something I have argued here and elsewhere would violate the separation of powers\textsuperscript{205}—the Senate could exercise its unbounded discretion over confirmation to screen Supreme Court candidates based on their views of Supreme Court ethics and recusal. This approach is admittedly limited, as a candidate’s views on recusal may not be sufficient to deny their confirmation, and questioning a nominee about recusal would not have any legal effect on a Justice post-confirmation. Nevertheless, by focusing at least in part on recusal at a nominee’s confirmation hearing, the Senate could encourage both the nominees and the public to more closely examine the recusal issue in a way that could create pressure on the new Justice to comport with the views expressed at their confirmation hearing.

4. Appropriations

Congress’s power of the purse is yet another potentially useful means of legislative influence over the Court’s recusal practices. Like many of the other indirect methods mentioned, Congress’s appropriations power has the benefit of a clear constitutional pedigree; there is no question that the ultimate authority to provide funding for the coordinate Branches lies squarely and solely with Congress.\textsuperscript{206} Appropriations are also a

\textsuperscript{204} See Confirmation Hearing on the Nomination of John G. Roberts, Jr. To Be Chief Justice of the United States: Hearing Before the S. Comm. on the Judiciary, 109th Cong. 55 (2005) (“Judges are like umpires. Umpires don’t make the rules, they apply them. The role of the umpire and a judge is critical. They make sure everyone plays by the rules, but it is a limited role.”).

\textsuperscript{205} See generally Virelli, supra note 11 (stating that separation of powers principles prevent Congress from directly regulating Supreme Court recusal).

\textsuperscript{206} U.S. Const. art. I, § 9, cl. 7 (“No Money shall be drawn from the
powerful source of leverage over the other Branches. Finally, because appropriations are inherently focused on the Court as an institution, rather than the individual Justices, Congress’s power of the purse could be an effective way to encourage the Court to adopt its own recusal reforms.\footnote{This would be perhaps the best of all possible outcomes, as a decision by the Justices to voluntarily adopt clearer and more transparent recusal practices could promote the legitimacy of the Court without disrupting the balance of powers between the Court and Congress. See infra Part IV.B.2.}

There are, however, limitations on any congressional attempt to influence Supreme Court recusal through appropriations. Although the institutional focus of appropriations may serve as a benefit in attempting to change the recusal practices of the Court as a whole, appropriations are not a good way for Congress to try and address an individual Justice’s recusal practices. Appropriations are also limited by the fact that they do not address the Court’s recusal practices directly; they are a source of pressure designed to incentivize the Justices to change their behavior in exchange for funding that likely has little or nothing to do with that behavior. While this is not a weakness in terms of the relevance or availability of appropriations as a source of legislative influence, it does render the power of the purse inferior to other approaches such as procedural reforms and even investigations that are able to target and potentially change specific recusal practices directly. Moreover, in the event Congress chose to rely heavily on its appropriations power to influence the Justices’ recusal practices, additional problems could arise. A decision to withhold funding in order to affect recusal could have serious consequences for the Court’s ability to perform its constitutionally-assigned judicial function. At the extreme, a deprivation of funding could impair the Court so severely as to threaten the Article III requirement that there be “one supreme Court.”\footnote{U.S. Const. art. III, § 1.} Notwithstanding these limits, Congress’s power of the purse is another constitutionally-recognized tool by which Congress may influence the Justices’ conduct, and as such is a potentially useful feature in the process of reformulating the balance of power between Congress and the Court over recusal.
5. Investigation

Finally, Congress can use its general investigatory power\(^{209}\) to investigate and, more specifically, conduct public hearings on issues of national importance. In light of the constitutional and practical realities of Supreme Court recusal, any lasting, effective institution-wide reform will ultimately be up to the Justices themselves. One way for Congress to instigate such reform is to bring additional public awareness and pressure to bear on the Justices such that they reevaluate their own recusal practices.\(^{210}\) Congress’s investigatory authority is broad, and includes the ability to question members of the Court regarding recusal.\(^{211}\) Congress began to explore this approach in 2011, as Justices were asked to testify before Congress on two separate occasions about

\(^{209}\) This reference to “general investigatory power” is meant to contrast the exercise of congressional power being discussed here with the power to investigate incident to impeachment proceedings. Investigations relating to an impeachment are part of the broader discussion of impeachment and recusal included supra at notes 127–32 and accompanying text.

\(^{210}\) This is not to suggest that Congress should employ the full weight of its investigatory authority—such as its subpoena power—on the Justices, as this would implement questions of interbranch coordination and comity that may themselves run afoul of the separation of powers. The investigatory authority contemplated in this section is akin to an ongoing confirmation hearing, whereby Congress may enhance the public’s awareness and knowledge of the Supreme Court’s recusal practices as a way of encouraging the Justices to remain vigilant in thinking about and evaluating their recusal practices.

\(^{211}\) See, e.g., William P. Marshall, The Limits On Congress’s Authority to Investigate the President, 2004 U. ILL. L. REV. 781, 785 (“Although there is no explicit textual grant of investigative power to Congress in the Constitution, the proposition that a legislative body generally possesses investigative powers is not controversial as a historical matter.”); Todd David Peterson, Congressional Investigations of Federal Judges, 90 IOWA L. REV. 1, 11 (2004)

As with the testimony concerning court administration, testimony by federal judges on policy matters directly affecting the court (as long as it does not morph into an investigation of the judge) seems not only to be non-problematic, but also to be essential to the preservation of good relations between the legislative and judicial branches, and important for the protection of the interests of the federal courts.

See also Fisher, supra note 154, at 160 (“Congress uses its investigative power to satisfy four main purposes: to enact legislation, to oversee the administration of programs, to inform the public, and to protect its integrity, dignity, reputation and privileges.”).
ethical and recusal practices at the Court.\textsuperscript{212} Following those
critical appearances, and in conjunction with public pressure from
academics and the media to consider reform,\textsuperscript{213} Chief Justice
Roberts chose to dedicate nearly all of his Year-End Report to
ethical issues, including recusal. Although it is impossible to
draw a definite causal link between the increased attention to the
Justices’ recusal practices and the Chief Justice’s decision to
publicly address those practices in his annual report, at least
some connection between the two is easy to imagine. It is rational
to assume that the Court would take seriously the feelings of a
coordinate branch, especially when those feelings are
corroborated by public opinion and advocacy organizations.\textsuperscript{214}
This assumption is supported by the fact that the first public
statement by the Court regarding recusal since 1993 occurred in
the wake of just this type of congressional and public pressure.

Prohibiting Congress from directly regulating the Justices’
recusal decisions does not render it constitutionally helpless in
influencing those decisions. Congress may exercise its largely
unfettered constitutional discretion to impeach, fashion
procedural standards, confirm judicial nominees, appropriate
funds, and investigate in order to protect against recusal
decisions that imperil the public’s perception of the Court and its
institutional legitimacy. Although each of these indirect
constitutional approaches is limited in its scope and potential
effectiveness with regard to individual recusal decisions,
collectively these approaches have the opportunity to create some

\textsuperscript{212} Justices Breyer and Kennedy were asked about Supreme Court ethics
and recusal standards as part of their testimony before a House subcommittee
on April 14, 2011. See Malloy, supra note 5, at 2389. Additionally, on October 5,
2011, Justices Breyer and Scalia testified about the same topic before the
Senate Judiciary Committee. See Considering the Role of Judges under the
Constitution of the United States: Hearing Before the S. Judiciary Comm., 112th

\textsuperscript{213} See supra notes 2–10 and accompanying text (outlining the public
media and academic scrutiny brought to bear on the recusal question in 2011).

\textsuperscript{214} The Alliance for Justice recently circulated a request for signatures for
a letter requesting the Court to take up ethical reforms on its own. See Alliance
for Justice, Supreme Court Ethics Reform, http://www.afj.org/connect-with-the-
issues/supreme-court-ethics-reform/judicial-ethics-prof-letter-scoutus.html (last
normative benefits that are potentially lost in the current impasse between Congress and the Court over recusal.

B. Political and Institutional Benefits

There are two broad categories of benefits that accrue from shifting the conversation about Supreme Court recusal to a constitutional framework. The first are systemic benefits, benefits that adhere to the effectiveness of our entire tripartite system by promoting the constitutional legitimacy and reputation of the Court. The second are institutional benefits, which include the substantive benefits to both litigants and the Court that arise from the Justices retaining (at least direct) control over their own recusal decisions.215 Rather than encouraging Congress to engage in unenforceable and unconstitutional direct regulation of the Justices’ recusal practices that has little if any transformative effect on those practices, a constitutional lens highlights the other options available to release the interbranch tension over recusal and to create a new dynamic between the Court and Congress (and ultimately the public) that is both constitutionally sound and effective. The result is more and better opportunities for lasting solutions to the impasse over recusal at the Court.

1. Systemic Benefits

The impasse over Supreme Court recusal reflects badly on the Justices and draws into question the constitutional legitimacy of all of their decisions, including recusal decisions. The ongoing tension facilitated by the Court’s failure to comply with an unenforceable congressional mandate creates a sense of arbitrariness and overreaching on behalf of the Justices, regardless of whether their decisions are in fact legitimate or

215. As discussed in this section as well as in supra Part IV.A, Congress has a series of constitutional mechanisms at its disposal to influence the Court’s recusal decisions. Nothing in this discussion of the benefits associated with the Justices’ control over those decisions is meant to diminish or otherwise overlook Congress’s authority in that regard. The benefits argued for here are understood to exist alongside congressional use of its own constitutional influence.
constitutional. Part of this problem stems from the fact that the Court, as the regulated entity, is also the final word in applying the recusal statute. Short of reading the recusal statute to create a strong presumption in favor of recusal, a position that would raise other constitutional problems, the Court invites criticism that it is acting in a self-serving or aggrandizing manner simply by virtue of its fulfilling its role as the governmental entity ultimately responsible for deciding “what the law is.” Assigning constitutional responsibility for alleviating the impasse over recusal to Congress helps avoid the problems of legitimacy for the Court that come both from an irresolvable dispute with Congress over recusal as well as an attempt by the Court to resolve such a dispute in its own favor.

In addition to enhancing the Supreme Court’s reputation and legitimacy, relief from the interbranch tension over recusal could refocus Congress on more constitutionally defensible and appropriate activities regarding recusal. Congress’s pursuit of indirect constitutional mechanisms represents a more effective and cooperative use of our constitutional structure to resolve difficult problems, and as such offers the public an example of its federal government at work that makes sense within the constitutional framework. The broader range of constitutional tools at Congress’s disposal are well-suited to raise public awareness about recusal and put the Justices on notice that their recusal decisions—and as a result the institutional legitimacy of

216. One could argue that a reciprocal perception problem adheres for Congress; that the ineffectiveness of the recusal statute makes Congress appear ineffectual. While this is true in theory, I would suggest that the appearance of arbitrariness or overreaching is far more damaging to the public perception of a democratic institution than the perception that it is unable to enforce its own positions, especially when the failure to enforce is due to the apparent obstinacy of a coequal branch. As a result, the public perception problem created by the impasse over recusal is depicted here as solely a judicial problem not because there is no countervailing negative repercussions for Congress, but because those repercussions are less severe and in some ways derivative of the problems created for the Court.

217. See discussion supra Part III.B (discussing the constitutional limits on congressional regulation of Supreme Court recusal). See generally Virelli, supra note 11 (stating that separation of powers principles prevent Congress from directly regulating Supreme Court recusal).

the Court in general—will be subject to greater public scrutiny. Confirmation hearings can work to create cultural change about recusal at the Court as well as put individual Justices on specific notice about what Congress and the public expect in terms of fairness and consistency from the Justices. Congressional investigations and hearings can also raise public awareness and communicate clear messages to the Court as a whole about recusal, which in turn may serve as incentives for internal recusal reform. Appropriations offer further incentives for internal recusal reform at the Court and provide a potentially valuable opportunity for Congress and the Court to cooperate in the best spirit of the separation of powers. The specter of impeachment is an unlikely but powerful deterrent against the Justices overreaching to hear cases that Congress and the public consider them unfit to decide, and jurisdictional limitations and procedural reforms are potentially useful ways to maintain some additional checks on the Justices’ conduct. Even if these mechanisms are rarely utilized, a shift by Congress to focus on these measures creates a public image of an orderly and effective system of checks and balances, rather than the unrestrained exercise of personal judgment by unelected Justices.

Recognition of the proper division of constitutional responsibility over recusal also promotes legitimacy by signaling public trust in the integrity and professionalism of the Court. This signal is valuable for at least two reasons. First, it is factually accurate, at least in the majority of cases. Although ethics and recusal at the Court have become a hotly discussed issue, even the strongest critics of the Court’s practices concede that in the overwhelming majority of instances in which recusal decisions are required, the Justices either get it right or make responsible decisions.219

219. See Stempel, supra note 32, at 642

To be fair, the low use and success rate of recusal motions probably stems in large part from the Justices’ ability to stop a lurking conflict of interest and voluntarily remove themselves from questionable cases. . . . Indeed, because of this strong informal tradition of stepping aside where appropriate without being asked, the custom of counsel has been to refrain from seeking recusal by motion . . . .

See also McElroy & Dorf, supra note 60, at 99 (explaining that “Justices would
Second, it facilitates a sense of interbranch comity and respect that is essential to the integrity of a tripartite government.\textsuperscript{220} Under the current arrangement, Congress appears to be calling into question the legitimacy of the Court without either the imprimatur of a binding legal standard to initiate change or a means of addressing the unique institutional concerns associated with Supreme Court recusal. Expressly acknowledging the constitutional reality that the Court may, at minimum, exercise significant discretion in its recusal decisions promotes public understanding of, and confidence in, the separation of powers. This argument is made even more powerful by the fact that the benefit is available virtually free of cost. As it currently stands, Supreme Court recusal decisions are entirely dependent on the Justices' individual judgment and integrity.\textsuperscript{221} Acknowledgment that the Court is the sole arbiter of its own recusal questions (subject only to those constitutional checks clearly assigned to Congress) could serve as a powerful endorsement of the Court's competence and integrity without requiring that Congress relinquish any actual authority.

Finally, judicial procedure generally, and recusal decisions in particular, are squarely within the expertise of the Judiciary. Congressional acknowledgement of the Court's greater expertise further inspires public confidence in the Court and properly draws focus on the Legislature's constitutionally assigned prerogatives for curtailing judicial power.

2. Institutional Benefits

Just as alleviating the impasse over recusal at the Court will have systemic benefits for the overall efficacy and legitimacy of our constitutional democracy, it may also present potential advantages for both individual litigants and the Court. Perhaps

\textsuperscript{220} See Paulsen, \textit{supra} note 93, at 228–29 (describing the importance of “coordinacy” among the Branches).

\textsuperscript{221} See Stempel, \textit{supra} note 32, at 642 (“Supreme Court recusal practice provides an almost unique illustration in American government of substantive law without force when applied to a certain institution.”).
the most obvious of these advantages is that a congressional decision to focus on indirect constitutional influences and formally commit substantive recusal decisions exclusively to the Justices most accurately describes current and historic constitutional practice. This is significant because it eliminates any distortive effects from the current statutory regime. Under the existing statutory standards, Congress provides the Court with unenforceable statutory “cover” for its recusal decisions. This discourages transparency and accountability among the Justices, as they are free to explain their recusal decisions in the context of the statutory standard when there is little reason to believe they feel bound by or accountable to that standard; although they generally cite to the governing statute when discussing recusal, in practice the Justices seem content to rely on extrastatutory sources and arguments to support their conclusions. The statute thus provides a veil of legality over what is, in actuality, a constitutionally assigned judicial policy decision about whether to recuse. By lifting the statutory veil from the Court’s recusal practices, each Justice’s recusal decisions can be properly understood and attributed to them as their own analysis of what is in the best interests of justice in that case. While this may not necessarily lead to more satisfying results in specific cases, it will more accurately focus public attention on the correct target in evaluating the Court’s recusal jurisprudence. This increased transparency and accountability among the individual Justices could in turn promote internal reform by encouraging the Justices to reexamine their views on the proper role of recusal for themselves and the Court as an institution.

In addition to promoting transparency in the Court’s recusal jurisprudence, there are reasons to believe that Congress could promote better substantive recusal practices by accepting its proper constitutional role and foregoing attempts to directly regulate recusal at the Court. First, stepping aside could open up space for the Court to take greater responsibility for the recusal

222. See discussion supra notes 111–20 and accompanying text (describing the recusal analyses proffered by then-Justice Rehnquist in Laird and Justice Scalia in Cheney).
issue and be more creative in seeking its own responses to public and governmental criticism. Although it is unlikely, especially in light of recent events, that the Court would adopt sweeping formal reforms of its recusal practices, heightened public awareness of the issue, coupled with the absence of a statutory standard, may push the Justices toward more frequent explanation of their recusal decisions or more serious consideration of recusal questions in certain circumstances. An incentive to engage in voluntary recusal reform may be the best possible solution to the public discontent over Supreme Court recusal, as it incorporates recusal reform without threatening the balance of power between Congress and the Court.

Another outcome that seeks to achieve this balance and is made more likely by Congress refraining from direct regulation of recusal at the Court is the reintroduction of the Due Process Clause into the discussion of Supreme Court recusal. Although the separation of powers may reserve authority over recusal to the Court alone, it does nothing to limit the scope of the Court’s authority over its own recusal practices or to insulate the Justices from the constitutional requirements of fairness, dignity, and personal liberty embodied in the concept of due process. Moreover, because due process is a constitutionally mandated individual right, it is uncontroversial as a separation of powers matter because it is precisely the sort of standard that the Court is qualified to apply and constitutionally bound to uphold.


224. See 2011 YEAR-END REPORT, supra note 3 (explaining why Supreme Court Justices must exercise their own independent judgment in making recusal decisions); Chief Justice Letter, supra note 13 (“[F]or the reasons explained in my year-end report, the Court does not plan to adopt the Code of Conduct for United States Judges.”).

225. See Poe v. Ullman, 367 U.S. 497, 542 (1961) (Harlan, J., dissenting) (describing the concept of due process as “the balance which our Nation, built upon postulates of respect for the liberty of the individual, has struck between that liberty and the demands of organized society”).
Finally, the elimination of direct congressional influence over the recusal process leaves that process with the Court. By virtue of their experience and training alone, the Justices possess greater institutional competence than Congress in fashioning effective recusal standards.

In sum, treating Supreme Court recusal as a constitutional question governed by the principle of separation of powers offers significant systemic and institutional benefits. It promotes public confidence in the Court as well as transparency and accountability among the Justices by shaping the legal framework to better reflect constitutional practice. It also has potential to improve the Justices’ decision making in the area by incentivizing them to consider their own recusal practices and to incorporate standards that are squarely within the Court’s institutional expertise and responsibility.

V. Conclusion

Recusal has come to dominate current conversations about the Court because it implicates the most foundational features of Supreme Court jurisprudence—fairness, legitimacy, transparency, and the proper scope and exercise of governmental power. Nevertheless, the debate over recusal has so far been framed almost exclusively as a matter of judicial ethics. This Article is the second part of an effort to reexamine the recusal issue through a constitutional lens in order to better illustrate how structural principles like the separation of powers are necessary to help us more fully understand the consequences of the Justices’ recusal decisions and the dynamics of tripartite government more broadly. In an earlier treatment of recusal at the Court, I argued that statutory recusal standards are inconsistent with the principle of separation of powers because they unduly infringe on the Court’s inherent judicial power under Article III of the Constitution. The present analysis takes this constitutionalization of the recusal debate a step further, asking how the separation of powers can help us reach a constitutionally acceptable and effective resolution of the ongoing impasse between Congress and the Court over recusal.
The answer can be found at the intersection of Congress’s status as a constitutional interpreter and all three Branches’ responsibility to promote interbranch coordination and comity in a tripartite constitutional regime. When the Supreme Court is at odds with a coequal branch of government, as in the recusal context, the separation of powers suggests that the Court should not take it upon itself to resolve the dispute, especially if the better constitutional argument favors the Justices. It is precisely in this scenario when Congress must be called on to take up the interpretive mantle and correct the constitutional impasse in a way that best promotes effective government. In the context of recusal, this calls for Congress to cease regulating the Justices’ recusal practices directly, and instead to employ its indirect constitutional tools such as impeachment, procedural reform, judicial confirmation, appropriations, and investigation to influence the Court’s recusal decisions.

Perhaps more importantly, this application of the separation of powers to the interbranch impasse over recusal offers broader insights into the best way to think about constitutional conflicts between the Court and its coordinate Branches going forward. By considering the role of the Court in its conflict with Congress over recusal, we are better able to understand which structural principles can be used to alleviate interbranch tension and diffuse potential crises of institutional legitimacy. The instant analysis also highlights a series of critical questions regarding constitutional structure and dynamics for future research, such as: Should democratic legitimacy always be the guiding principle in resolving interbranch disputes? What is each branch’s institutional responsibility in resolving those disputes? Do those responsibilities change when a branch is a party to the dispute, or are they a function of institutional competency and constitutional authority? Under what circumstances should a coordinate branch feel comfortable asserting its own interests against another branch? Does it matter if those interests are legal or political? What constitutional mechanisms are best suited to provide resolutions? The constitutional lessons learned from Supreme Court recusal serve as a template for addressing these bigger questions about the boundaries of constitutional power and the role of each of the coordinate Branches in facilitating our constitutional democracy.
“Waiving” Goodbye to Arbitration: A Contractual Approach

Paul Bennett IV*

Table of Contents

I. Introduction ................................................................... 1611
II. Arbitration: Objectives and Law .............................. 1617
   A. The Contractual Benefits of Arbitration:  
      Forum and Procedural Freedom ....................... 1617
   B. Arbitration Doctrine: The Supreme Court’s  
      Interpretation of the FAA ............................... 1622
      1. Early Twentieth Century Arbitration: Pre-  
         and Post-FAA ........................................... 1622
      2. Current Arbitration Doctrine: Enforcing  
         Arbitration Agreements as Contracts .......... 1625
      3. Severability: A “Gateway” to Arbitration .... 1627
      4. The Impact of Severability on Parties’  
         Arbitral Rights in the Courts ....................... 1630
      5. The Origins of the “Waiver” Doctrine in  
         the Arbitral Context: Correctly Applied? ........ 1632
   III. The Circuit Split .................................................... 1635
      A. The Majority: Circuits Requiring Prejudice ....... 1636
      1. Strict Enforcement: Circuits Imposing  
         Burdensome Prejudice Requirements .......... 1636
         a. The Second Circuit ................................. 1636
         b. The Fourth Circuit .................................. 1638
         c. The Fifth Circuit .................................... 1640
         d. The Sixth Circuit ..................................... 1642

* Candidate for J.D., Washington and Lee University School of Law, May 2013. I would like to thank Professors Susan D. Franck and George Bermann for their helpful input throughout the Note-writing process. I would also like to thank Professor David Millon for his invaluable advice and mentoring on this Note and throughout my law school career. Finally, I would like to thank my family—Paul, Laura, Shaun, Erin, and Arline—I would not be where I am today without your constant love and support.
e. The Ninth Circuit ......................................................... 1643
f. The Eleventh Circuit .................................................. 1645
g. A Brief Summary: The Strict Enforcement Divergence .......... 1647

2. Circuits Imposing Lenient Prejudice Requirements .......................... 1649
   a. The First Circuit .................................................. 1649
   b. The Eighth Circuit ............................................... 1650
   c. Lenient Prejudice Requirements:
      A Brief Comparison ............................................. 1652

3. The Third and Tenth Circuits: Multifactor Tests .................................. 1653
   a. The Third Circuit ............................................... 1654
   b. The Tenth Circuit ............................................... 1655
   c. The Third and Tenth Circuits Compared .................................. 1657

B. The Minority: Circuits Not Requiring Prejudice .................................. 1658
   1. The Seventh Circuit ............................................. 1659
   2. The D.C. Circuit .................................................. 1660
   3. The Seventh and D.C. Circuits Compared ..................... 1662

IV. “Waiving” the Right to Arbitrate: A Contractual Approach .................... 1663
   A. The Need for a Uniform Standard .................................. 1663
   B. “Waiver” as a Generally Applicable Contract Law Defense .................. 1667

V. The Proposal .............................................................................. 1670
   A. A Contractual Approach ............................................. 1670
      1. The Groundwork: A Duty to Arbitrate Subject to Conditions ............. 1670
      2. The Reasonableness Test: What Constitutes a “Reasonable Time” for Performance? ......... 1672
      4. The Efficacy of the “Reasonableness Test” as Compared to Waiver .................. 1680

VI. Conclusion .............................................................................. 1683
I. Introduction

Arbitration, as reflected through the Federal Arbitration Act (FAA), is “simply a matter of contract between the parties.” This oft-cited “fundamental principle” reflects an indisputable “axiom,” but one which the Circuit Courts of Appeals have failed to apply consistently. Courts diverge as to whether a party resisting arbitration must show prejudice in order to prove that its opponent waived its right to arbitrate. This Note demonstrates not only that prejudice is unnecessary, but that the waiver doctrine itself is conceptually flawed. In its place, this Note proposes a multifactor judicial framework derived from generally applicable contract law principles—a reasonableness test.

The standard conception is that “parties agree to use arbitration—to use private judges rather than public court judges to resolve their disputes—because arbitration is a process that improves upon the court system for dispute resolution.” Indeed, arbitration is a private, alternative adjudicatory forum to which one gains access by contract. Although the “default forum of dispute resolution is litigation,” parties can “override the

8. See Christopher R. Drahozal & Quentin R. Wittrock, Is There a Flight from Arbitration?, 37 Hofstra L. REV. 71, 76 (2008) (noting that parties can be compelled to arbitrate only “if they have agreed to do so . . . by entering into an arbitration agreement”).
litigation default rule” by contracting to do so. This notion, however, begs the question: Why would a party contractually relinquish its right to seek legal recourse in court? The justifications for doing so are ubiquitous, but can be reduced to a veritable cost–benefit analysis. If the projected benefits of arbitration outweigh those of litigation in relation to the costs surrounding each regime, the parties will choose to arbitrate. Herein lies the critical contractual freedom provided by arbitration. Parties attempting to curb the costs and diminish the risks appurtenant to their contractual arrangements can tailor arbitral procedures designed to accomplish those very objectives.

Typically, the arbitration clause—a binding agreement entered into before a dispute arises—is the key which grants access to the arbitral forum and denotes the parties’ arbitral

---

9. See Drahozal, Why Arbitrate, supra note 6, at 165.
10. See, e.g., id. at 163 (“[A]rbitration may be preferred to litigation because it is cheaper and faster; because it enables parties to pick a decision maker (the arbitrator) who is an expert in the field; or because it provides a neutral forum . . . among other reasons.”); George A. Bermann, The “Gateway” Problem in International Commercial Arbitration, 37 YALE J. INT’L L. 1, 2 (2012) (“[A]ll participants . . . have an interest in ensuring that arbitration delivers the various advantages associated with it, notably speed, economy, informality, technical expertise, and avoidance of national fora . . . .”).
11. See Drahozal, Arbitration Costs, supra note 7, at 833 (noting the importance of “[c]omparing the costs of arbitration with the costs of litigation . . . in evaluating the efficiency of the two processes”).
12. See infra Part II (discussing the cost–benefit analysis conducted by parties in determining whether to arbitrate or litigate certain disputes).
13. This Note will not confront the numerous social issues surrounding consumer and employment arbitration clauses in contracts of adhesion. Such issues have been the subject of intense debate among scholars and commentators, and are outside the parameters of this Note. Avoiding this issue, however, has allowed this Note to focus squarely on the doctrinal underpinnings of arbitration in formulating an alternative to the waiver doctrine. Focusing on the doctrine surrounding this issue mirrors the Supreme Court’s approach in addressing other arbitral issues. See, e.g., AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1750 (2011) (finding a state law which held class arbitration waivers unconscionable preempted by the FAA, and stating that although the “rule is limited to adhesion contracts, . . . the times in which consumer contracts were anything other than adhesive are long past”).
14. See Drahozal, Why Arbitrate, supra note 6, at 177 (noting that parties can devise procedures to fit “the type of contract or perhaps even the type of dispute” they foresee as potentially arising from that contract).
Title 9, § 2 of the United States Code governs the validity and enforcement of such agreements. The Supreme Court recently provided a concise articulation of § 2's substantive mandate: “FAA [§ 2] . . . places arbitration agreements on an equal footing with other contracts . . . and requires courts to enforce them according to their terms. Like other contracts, however, they may be invalidated by generally applicable contract defenses, such as fraud, duress, or unconscionability.”

The FAA also establishes procedural mechanisms by which federal courts implement § 2’s substantive mandate. Section 3 requires courts to stay litigation, pending arbitration of any claims falling within the scope of the parties' arbitration agreement. Further, § 4 provides a coercive measure, allowing courts to compel arbitration on behalf of any party “aggrieved” by its opponent’s refusal to arbitrate.

Overall, the FAA embodies “a broad congressional expression of social policy and, in a barebones statute, a delegation of decisionmaking responsibility to the judiciary.” As such, courts play an essential role in “policing” the arbitral process, safeguarding the key benefits that arbitration provides and which parties expect to derive. This often entails ensuring that arbitral proceedings are initiated expeditiously and

---

15. See id. at 165–66 (explaining the difference between “pre-dispute” and “post-dispute” arbitration agreements). Although post-dispute arbitration agreements are possible, “the substantial majority of arbitration proceedings arise out of pre-dispute agreements” and, as such, are the sole focus of this Note. Id. at 165.


18. See id. (describing how, procedurally, FAA §§ 3 and 4 “implement § 2’s substantive rule”).

19. See 9 U.S.C. § 3 (2011) (stating that if a party files suit in court, the court “shall on application of one of the parties stay the trial of the action until arbitration has been had in accordance with the terms of the agreement”).

20. See 9 U.S.C. § 4 (2011) (stating that, upon the application of a party to a valid arbitration agreement, “the court shall make an order directing the parties to proceed to arbitration”).


22. See Bermann, supra note 10, at 2.
efficaciously, as challenges to the process frequently arise at the outset of a dispute.\textsuperscript{23} Unfortunately, the procedural tools provided in FAA §§ 3 and 4 to assist the courts in fulfilling their role are simply that—procedural.\textsuperscript{24} Neither section denotes the appropriate circumstances in which a stay of litigation, or an order compelling arbitration, should be granted.\textsuperscript{25} Thus, devoid of statutory directives, courts have struggled to maintain juridical clarity in confronting complex issues which arise at the outset of disputes—in the grey area between litigation and arbitration.\textsuperscript{26}

One such issue has caused a significant split among the Circuit Courts of Appeals. Particularly, courts have struggled to discern whether a party resisting arbitration must show prejudice in order to prove that its opponent waived its right to arbitrate by engaging in pretrial conduct.\textsuperscript{27} A majority of circuits

\begin{itemize}
\item \textsuperscript{23} See id. (describing how one way in which courts fulfill their policing role is to ensure that “arbitral proceedings are initiated and pursued in a timely and effective manner,” as “courts are commonly asked . . . to intervene at the very outset of a dispute”).
\item \textsuperscript{24} See 9 U.S.C. §§ 3–4 (2011) (governing stays of litigation pending arbitration and motions to compel arbitration).
\item \textsuperscript{25} See Bermann, supra note 10, at 3 n.5 (noting that neither § 3 nor § 4 “addresses the issues that may specifically be raised” in order to secure an order to stay, or compel, arbitration under those sections).
\item \textsuperscript{26} See id. at 4 (stating that, in addressing issues which arise at the outset of disputes, courts have developed “disparate strands of analysis” which “have combined to produce a needlessly confusing case law to the detriment of clarity, coherence, and workability”); Steven H. Reisberg, \textit{The Rules Governing Who Decides Jurisdictional Issues: First Options v. Kaplan Revisited}, 20 AM. REV. INT’L ARB. 159, 159 (2009) (discussing the “significant confusion as to how a court is to decide which forum, the court or the arbitrator, has the jurisdiction to decide [a] threshold issue”).
\item \textsuperscript{27} See \textit{The Federal Arbitration Act—Waiver}, APPELLATE.NET (Feb. 22, 2011), http://www.appellate.net/docketreports/html/2010/docketreport_22Feb11.asp#Case1 (last visited Sept. 24, 2012) [hereinafter \textit{FAA Waiver}] (stating the issue as whether “a party opposing arbitration must demonstrate that it was prejudiced by the other party’s conduct in order to show that that party had waived its right to compel arbitration”) (on file with the Washington and Lee Law Review); Petition for Writ of Certiorari at 1, Stok & Assocs., P.A. v. Citibank, N.A., No. 10-514 (11th Cir. 2010) (framing the issue as whether “a party [should] be required to demonstrate prejudice after the opposing party waived its contractual right to arbitrate by participating in litigation”); \textit{see also} Born, supra note 5 (describing the issue as whether “prejudice on the part of a resisting party is necessary for an opposing party’s right to compel arbitration to be deemed waived”).
\end{itemize}
do require a showing of prejudice, albeit at varying degrees. Conversely, the minority does not require prejudice, but holds a presumption of waiver which may be rebutted in certain circumstances.

The circuit split recently prompted the Supreme Court to grant certiorari in order to resolve the matter in Stock & Associates, P.A. v. Citibank, N.A. Although the parties settled their dispute before the Court ruled on the merits, its decision would have had a momentous impact on the business community and the arbitral process itself. Indeed, the issue is “of great interest to any business that makes use of arbitration agreements,” as the grounds invoked to resist arbitration often include waiver “through preliminary litigation conduct.” The Court’s decision, as described by one practicing arbitration attorney, would have “affect[ed] how quickly (or not) parties must determine whether arbitration clauses apply to their case and when arbitration rights must be asserted.”

28. See, e.g., Lewallen v. Green Tree Servicing, LLC, 487 F.3d 1085, 1090 (8th Cir. 2007) (utilizing a three-pronged test for waiver which, in part, requires prejudice); Rankin v. Allstate Ins. Co., 336 F.3d 8, 12 (1st Cir. 2003) (noting that a “modicum of prejudice” is necessary to find waiver); Sovak v. Chugai Pharm. Co., 280 F.3d 1266, 1270 (9th Cir. 2002) (utilizing a three-pronged test for waiver which, in part, requires a showing of prejudice); Ivax Corp. v. B. Braun of Am., Inc., 286 F.3d 1309, 1315–16 (11th Cir. 2002) (requiring that one party somehow prejudice an opposing party in order to find waiver); Coca-Cola Bottling Co. of N.Y. v. Soft Drink & Brewery Workers Union Local 812, 242 F.3d 52, 57 (2d Cir. 2001) (stating that proof of prejudice is necessary to find waiver); MicroStrategy, Inc. v. Lauricia, 268 F.3d 244, 249 (4th Cir. 2001) (requiring “actual prejudice” to find waiver); Walker v. J.C. Bradford & Co., 938 F.2d 575, 577 (5th Cir. 1991) (requiring an invocation of the judicial process by one party which prejudices the other party).

29. See, e.g., Cabinetree of Wis., Inc. v. Kraftmaid Cabinetry, Inc., 50 F.3d 388, 390 (7th Cir. 1995) (“We have deemed an election to proceed in court a waiver of a contractual right to arbitrate, without insisting on evidence of prejudice . . . .”); Nat’l Found. for Cancer Research v. A.G. Edwards & Sons, Inc., 821 F.2d 772, 777 (D.C. Cir. 1987) (“This circuit has never included prejudice as a separate . . . element of the showing necessary to demonstrate waiver of the right to arbitration.”).


32. FAA Waiver, supra note 27.

33. American Arbitration Association, Supreme Court Agrees to Hear
opinion would have “provide[d] an analytical framework that parties and courts [could] apply to a broad range of arbitration issues that arise.”

Given the importance of this issue, commentators predict that the Court will address the matter in the near future. In the meantime, this Note attempts to fill the doctrinal void recognized by the Supreme Court, and assuage the juridical strife by developing its own analytical framework to supplant the waiver doctrine.

This Note’s analysis of the waiver doctrine proceeds as follows: Part II of this Note discusses the key contractual benefits which attract parties to arbitration. Particularly, it discusses how arbitration agreements, as forum-selection and procedural-mapping devices, stabilize and optimize parties’ contractual relationships. Further, Part II analyzes the Supreme Court’s FAA jurisprudence and concludes that contract law should govern the waiver by conduct analysis. Part III provides an overview of the circuit split surrounding waiver doctrine. Particular attention is paid to the prejudice requirement and the discrepancies it perpetuates among the circuits. Part IV provides the contractual background for the proposal advanced by this Note. It draws the distinction between waiver as defined by contract law, and waiver as applied by the Circuit Courts of Appeals in the arbitral context. It concludes that, in light of the Supreme Court’s recent FAA jurisprudence, a party cannot “waive” its right to arbitrate. Therefore, it posits that an entirely new framework is required.


34. Id.

35. See Born, supra note 5 (“Now that the case has been dismissed, it seems likely that the Supreme Court will grant certiorari the next time it finds itself presented with an appropriate opportunity to address this issue.”).

36. See infra Part II (discussing the contractual benefits provided by arbitration).

37. See infra Part II (discussing how, under the severability doctrine, waiver is a gateway issue which must be resolved in accordance with contract law).

38. See infra Part III (describing the circuit split and discussing the divergence engendered by the prejudice requirement).

39. See infra Part IV (concluding that a party cannot, in accordance with contract law, waive its right to arbitrate).
Finally, Part V proposes a comprehensive contractual solution through a succinct judicial framework in order to discern whether a party has lost its right to arbitrate through pretrial conduct. Part V first proposes that the discharge-of-duty doctrine\textsuperscript{40} is the appropriate contract-law defense for a party responding to excessive pretrial conduct.\textsuperscript{41} Next, Part V proposes that courts should read an implied term into all arbitration agreements. This term would require that parties demand arbitration within a reasonable time after one party files a claim in court.\textsuperscript{42} In order to discern what constitutes a reasonable time, this Note proposes the adoption of a reasonableness test, comprised of the four-factor framework promulgated by the New York courts. These courts consider “the nature and object of the contract, the previous conduct of the parties, the presence or absence of good faith, . . . and the possibility of prejudice or hardship” to either party.\textsuperscript{43} Part V concludes with a hypothetical illustrating the efficacy of the reasonableness test both doctrinally and from a policy perspective.

\textit{II. Arbitration: Objectives and Law}

\textit{A. The Contractual Benefits of Arbitration: Forum and Procedural Freedom}

Before discussing the current legal doctrine surrounding arbitration in the United States, it is crucial to consider what attracts parties to arbitration. What core benefits do parties derive through bargained-for arbitral procedures? In what

\textsuperscript{40} See Restatement (Second) of Contracts § 225(2) (1981) (“Unless it has been excused, the non-occurrence of a condition discharges the duty when the condition can no longer occur.”).

\textsuperscript{41} See Doctor’s Assocs., Inc. v. Casarotto, 517 U.S. 681, 687 (1996) (stating that arbitration agreements may only be invalidated by “generally applicable contract defenses, such as fraud, duress, or unconscionability”).

\textsuperscript{42} Restatement (Second) of Contracts § 204 (1981) (“When the parties to a bargain . . . have not agreed with respect to a term which is essential to a determination of their rights and duties, a term which is reasonable in the circumstances is supplied by the court.” (emphasis added)).

\textsuperscript{43} Zev v. Merman, 533 N.E.2d 669, 669 (N.Y. 1988).
circumstances does arbitration present a more appealing
adjudicatory forum than litigation?

As a starting point, arbitration agreements are “in effect, a
specialized kind of forum-selection clause.” 44 Parties who bargain
for arbitration essentially contract for the choice between two
legal regimes—litigation and arbitration. 45 Comparatively,
parties might choose one forum over the other for any number of
reasons. For example, arbitration may be more expedient, cost
effective, and pose fewer risks of “aberrational” jury awards. 46 In
other circumstances, parties might desire the full panoply of
procedural protections and appellate review processes offered by
the judicial system. 47 Particularly, high stakes disputes or issues
concerning clearly delineated legal principles may prompt
recourse to a judicial, rather than an arbitral forum. 48 In any
event, arbitration provides a judicially accepted 49 alternative
“structure,” or “basic set of parameters within which people are
free to” resolve their disputes. 50

Additionally, parties acting within this alternative structure
have the ability to establish procedures by which their disputes
will be resolved. 51 To be sure, parties could opt to include a

44. Scherk v. Alberto-Culver Co., 417 U.S. 506, 519 (1974); see Christopher
R. Drahozal & Peter B. Rutledge, Contract and Procedure, 94 MARQ. L. REV. 1103, 1105 (2011) (stating that arbitration clauses are essentially “contractual
forum selection devices”).
45. See Keith N. Hylton, Arbitration: Governance Benefits and Enforcement
Costs, 80 NOTRE DAME L. REV. 489, 490 (2005) (stating that “parties who can
choose to submit their disputes to arbitration . . . have a choice between two
legal regimes,” litigation and arbitration).
46. See Christopher R. Drahozal & Stephen J. Ware, Why Do Businesses
Use (or Not Use) Arbitration Clauses?, 25 OHIO ST. J. ON DISP. RESOL. 433, 451
(2010) (stating that “arbitration may be faster and cheaper than litigation, . . .
may lessen the risk of punitive damages awards or aberrational jury
verdicts[,] . . . [and] may decrease exposure to class actions”).
47. See id. at 453 (discussing the circumstances in which parties might
choose to resolve their disputes through litigation rather than arbitration).
48. See id. at 436 (stating that parties may prefer litigation to arbitration in
“high stakes disputes” and “disputes in areas with clear and well developed
law,” as the “industry expertise” of arbitrators is less valuable in such cases).
49. See Drahozal & Rutledge, supra note 44, at 1105 (stating that “[c]ourts
[have] largely accepted” arbitration clauses as forum-selection devices, subject
only “to a narrow range of exceptions”).
50. See Hylton, supra note 45, at 489.
51. See Drahozal & Rutledge, supra note 44, at 1105 (noting that parties to
forum-selection clause in their contract to site potential disputes in a court with favorable procedural rules. Once invoked, however, most rules of civil procedure provide little, if any, opportunity to restrict their application. Stated differently, “a party might pick from among several restaurants but could not control what would be on the menu.”

By contrast, “arbitration clauses have a more profound effect on the procedure by which disputes are resolved.” In some cases, they incorporate the rules of arbitral institutions by reference, which act as “mini-codes of civil procedure.” These codes, however, merely provide default rules which give way in the wake of express terms in the parties’ arbitration agreement. In other instances, parties construct their own arbitral framework, negotiating procedural rules in lieu of those offered by arbitral institutions. In either case, however, parties may select any number of procedural measures, including restrictions on discovery, remedies, and limitations periods, subject only to due arbitration agreements have the ability to bargain “over procedural rights even before a dispute arises”).

52. See id. at 1114 (describing the operative purpose of forum-selection clauses).

53. See id. (noting that once the litigation forum is “fixed contractually . . . most rules of civil procedure limit[] the parties’ ability to contract around its provisions”).

54. Id.

55. Id.

56. See id. at 1106 (stating that arbitration agreements “may incorporate by reference the rules of arbitral institutions”).


58. See American Arbitration Association, Commercial Arbitration Rules and Mediation Procedures 20 (2009) [hereinafter AAA Rules] (“The parties, by written agreement, may vary the procedures set forth in these rules.”); Drahozal & Rutledge, supra note 44, at 1106 (stating that the rules of arbitral institutions “can be overridden by the express terms of the parties’ arbitration agreement”).

59. See Drahozal & Rutledge, supra note 44, at 1106 (noting that procedural rules may be “explicit terms of the parties’ contract, decoupled from the rules of an administering institution”).

60. See id. at 1107 (stating that examples of potential arbitral procedures include “limits on the availability of discovery, contractually imposed limitations periods, . . . [and] limitations on remedies”).
process concerns. Thus, arbitration clauses serve as both forum-selection devices and “procedural contracts” in which parties construct dispute resolution mechanisms that best support their contractual arrangements.

Ultimately, parties spend substantial resources establishing arbitral procedures, in lieu of the no-cost regime provided by the judicial system, to optimize the “processes and outcomes” surrounding their disputes. Indeed, “[a]rbitration provides an alternative forum in which parties can structure rules and enforcement methods so that the difference between governance benefits and enforcement costs is larger than in the default regime represented by ordinary courts.” Governance benefits encompass the costs saved by having rules which limit the level of risk that one party can impose on the other (i.e., a rule against breaching the contract). Conversely, enforcement costs include the costs of both writing the rules and arbitrating to enforce the rules which govern the contractual relationship in question. After engaging in this cost–benefit analysis, rational parties will adopt procedural rules that they expect “will provide them with a better process than litigation,” better “outcomes than litigation, or both.” At its core, then, arbitration provides

61. See 9 U.S.C. § 10 (2011) (providing limited grounds, rooted in due process concerns, upon which a party may seek to modify or vacate an arbitral award); AAA RULES, supra note 58, at 32 (“[E]ach party has the right to be heard and . . . given a fair opportunity to present its case.”).


63. Drahozal & Ware, supra note 46, at 451.

64. Hylton, supra note 45, at 493.

65. See id. at 491 (describing governance benefits as “[a]ny set of rules governing interaction among private parties” which will “provide a benefit for which the parties are willing to pay,” such as rules which “govern the amount of risk that one can impose on others”).

66. See id. (stating that enforcement costs include the costs of both writing and enforcing “the rules governing private interaction”).

67. See id. at 492 (noting that parties “will waive a legal rule whenever the governance benefit from the rule is less than the enforcement cost”); Christopher R. Drahozal, Contracting Out of National Law: An Empirical Look at the New Law Merchant, 80 NOTRE DAME L. REV. 523, 531–32 (2005) [hereinafter Drahozal, Contracting Out] (stating that, in choosing an adjudicatory forum, parties consider whether “the process costs of arbitration” are “higher or lower than the process costs of litigation”).

68. Drahozal & Ware, supra note 46, at 451.
stability. It allows parties to mitigate risks and costs appurtenant to their contractual relationships by extending predictability to the dispute resolution process. Thus, adequately enforced, post hoc arbitral procedures optimize the very contractual arrangements which impel their existence. The key, however, is adequate enforcement. Therefore, the current FAA jurisprudence aims to preserve the stability derived through bargained-for arbitral procedures by protecting parties’ legitimate expectations.

69. See Antony W. Dnes, Franchise Contracts, Opportunism and the Quality of Law, 3 ENTREPRENEURIAL BUS. L.J. 257, 259 (2009) (noting that “private enforcement mechanisms,” such as arbitration, are “a means to stabilize contracts within a foreseeable range of variation of market conditions”); Christopher R. Drahozal & Keith N. Hylton, The Economics of Litigation and Arbitration: An Application to Franchise Contracts, 32 J. LEGAL STUD. 549, 550 (2003) (stating that predispute arbitration agreements are “designed to minimize the costs of [contracting parties’] relationship[s]”); Hylton, supra note 45, at 491 (stating that “[a] rule against breaching contracts, if complied with perfectly, provides an ex ante benefit” to the parties’ contractual relationship); Amy J. Schmitz, Consideration of “Contracting Culture” in Enforcing Arbitration Provisions, 81 ST. JOHN’S L. REV. 123, 153 (2007) (“Prior dealings, personal relationships, and concern for maintaining good business reputations may compel players within a business community to comply with their contracts, or resolve disputes without the aid of the courts . . . . because private dispute resolution often solves far more problems than rigid litigation . . . .”); Thomas J. Stipanowich, Contract and Conflict Management, 2001 WIS. L. REV. 831, 831–32 (stating that, through arbitration, lawyers have the opportunity to “limit or manage problems prospectively” by negotiating and drafting “suitable issue and conflict resolution mechanisms for contractual relationships” which “assur[e] control and reduce uncertainty and risk”).

70. See Drahozal, Contracting Out, supra note 67, at 532–33 (finding that predispute choice of law provisions in contracts, “like the choice between arbitration and litigation,” lend certainty to parties’ contractual relationships which provides for adequate pricing of contract rights and duties and decreases the costs and frequency of litigation); Drahozal & Hylton, supra note 69, at 580, 582 (finding that parties bargain for arbitral procedures when that forum “provides the optimal level of deterrence against undesirable conduct,” thus increasing governance benefits and “the quality of output and level of effort” in parties’ contractual relationships); Hylton, supra note 45, at 500 (finding that because arbitration agreements “enhance governance benefits,” they promote “organizations working more effectively on a day-to-day basis”).

71. See Bermann, supra note 10, at 2 (noting that courts play an “important policing role” in assuring “that arbitration delivers the various advantages associated with it”).

B. Arbitration Doctrine: The Supreme Court’s Interpretation of the FAA

Because the benefits derived from arbitration are, in essence, contractual, it is appropriate that the FAA protects arbitration agreements as contracts. The Supreme Court, however, has not always interpreted the FAA in a manner so conducive to parties’ arbitral rights. Thus, a brief historical note will better frame the importance and implications of the Supreme Court’s current FAA jurisprudence.

1. Early Twentieth Century Arbitration: Pre- and Post-FAA

Prior to the FAA, opportunities to bargain for arbitral procedures were severely constricted. Considered nothing more than tools of oppression, courts generally nullified arbitration agreements and assumed jurisdiction over the matter in question. In 1925, however, Congress moved to quell the “judicial hostility” toward arbitration agreements by enacting the FAA.

The FAA’s stated purpose is to place arbitration agreements “upon the same footing as other contracts.” It “declares simply that . . . agreements for arbitration shall be enforced, and

---

74. See Drahozal & Rutledge, supra note 44, at 1112 (noting how “opportunities to control procedure” through arbitration clauses “were largely non-existent” prior to the FAA).
75. Parsons v. Ambos, 48 S.E. 696, 697 (Ga. 1904) (stating that arbitration agreements are used merely as tools to “oppress the weak”).
76. See W.H. Blodgett Co. v. Bebe Co., 214 P. 38, 40 (Cal. 1923) (stating that courts could simply “disregard [arbitration] agreements, assume jurisdiction, and determine the matters in dispute”); see also Parsons, 48 S.E. at 697 (finding that arbitration clauses “may be revoked by either party at any time before the award”); Cocalis v. Nazlides, 139 N.E. 95, 98 (Ill. 1923) (holding arbitration clauses “void”).
77. AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1745 (2011).
WAIVING GOODBYE TO ARBITRATION

provides a procedure in the federal courts for their enforcement.79 Despite this seemingly clear mandate, however, skepticism surrounding arbitrators' motives and juridical prowess remained.80 As such, federal courts continued to hold certain claims nonarbitrable, including alleged violations of federal securities81 and antitrust laws.82 As a result, claims arising under those laws remained in court, where the ability to bargain for optimal procedures remained limited.83 Gradually, however, federal courts' distrust of arbitration subsided.

Beginning with Prima Paint Corp. v. Flood & Conklin Manufacturing Co.84 in 1967, the Supreme Court became increasingly willing to enforce arbitration agreements.85 Indeed, the Prima Paint Court devised one of the most doctrinally significant concepts in U.S. arbitration law to date—the severability doctrine.86 It essentially “permits courts to entertain challenges specifically applicable to the arbitration agreement, and not to the contract as a whole.”87 The Court devised the

---

79. Id. at 1–2.
80. See Peter B. Rutledge, Whither Arbitration?, 6 GEO J.L. & PUB. POL’Y 549, 553 (2008) (noting that “courts showed greater tolerance” for arbitration agreements after Congress passed the FAA, but that “there were still limits” to that tolerance).
81. See Wilko v. Swan, 346 U.S. 427, 436 (1953) (refusing to allow the arbitration of federal securities claims because arbitral awards “may be made without explanation,” without a “record of their proceedings,” and “arbitrators’ conception of the legal meaning” of certain statutory requirements “cannot be examined”).
82. See Am. Safety Equip. Corp. v. J.P. Maguire & Co., 391 F.2d 821, 827–28 (2d Cir. 1968) (finding that antitrust cases should be resolved in court given their complexity and the fact that antitrust laws regulate the business community in which many arbitrators are a part).
83. See Drahozal & Rutledge, supra note 44, at 1112–13 (discussing how, under the “nonarbitrability doctrine,” claims which remained in federal court were subject to limited “opportunities to influence procedure by contract”).
86. See Bermann, supra note 10, at 24 (finding the severability doctrine “well-entrenched” in U.S. arbitration law).
87. Id. at 23; see Prima Paint, 388 U.S. at 403–04 (holding that if a claim is directed at “the arbitration clause itself...the federal court may proceed to adjudicate it,” but that claims directed at “the contract generally” must be
doctrine in accordance with the FAA's mandate to limit judicial obstruction and facilitate expedient access to the arbitral forum.88

After Prima Paint, parties' rights to bargain for arbitral procedures continued to expand throughout the 1970s and 1980s in two critical ways.89 First, the nonarbitrability doctrine dissipated,90 as the Supreme Court held federal securities claims, RICO claims,91 and antitrust claims arbitrable.92 Second, the Court interpreted FAA § 2 as declaring a national federal policy favoring arbitration, thus withdrawing states' power to declare certain claims nonarbitrable.93 In effect, state authority was supplanted by the "federal substantive law of arbitrability."94 This body of law requires that "doubts concerning the scope of arbitrable issues" be resolved in favor of arbitration, including those regarding "the construction of the contract language itself or . . . allegation[s] of waiver, delay, or a like defense to arbitrability."95 The Supreme Court thus came to embrace freedom of contract principles in the arbitral context, creating greater potential for parties to regulate the forum and procedures by which to resolve their disputes.96

88. See Prima Paint, 388 U.S. at 404 (formulating the severability doctrine in order to "honor . . . the unmistakably clear congressional purpose that" arbitration "be speedy and not subject to delay and obstruction by the courts").

89. See Drahozal & Rutledge, supra note 44, at 1113 (noting that throughout the "1970s and 1980s," the "opportunities to control procedure by contract expanded").

90. See id. (noting that "as the nonarbitrability doctrine crumbled," parties "had an incentive . . . to use their new contractual freedom" to establish appropriate procedures).


93. See Southland Corp. v. Keating, 465 U.S. 1, 10 (1984) ("In enacting § 2 of the [FAA], Congress declared a national policy favoring arbitration and withdrew the power of the states to require a judicial forum for the resolution of claims which . . . contracting parties agreed to resolve by arbitration.").


95. Id. at 24–25.

96. See Drahozal & Rutledge, supra note 44, at 1114 (noting that the Supreme Court's acceptance of "freedom of contract" raised a "newfound
2. Current Arbitration Doctrine: Enforcing Arbitration Agreements as Contracts

Judicial acceptance of arbitration agreements continues to expand today. For example, in Stolt-Nielsen S.A. v. AnimalFeeds International Corp., the Supreme Court found that a party cannot be forced to submit to class arbitration unless "there is a contractual basis for concluding that the party agreed to do so." Similarly, in First Options of Chicago, Inc. v. Kaplan, the Court found that parties can empower tribunals to determine whether they have jurisdiction over certain disputes. Thus, courts will defer to an arbitrator's decision regarding its own jurisdiction if the parties agreed to confer such power. If the parties did not confer such power, then the courts will decide. Indeed, one of the only contractual measures invalidated by the Supreme Court involved an attempt to expand the grounds on which to vacate an arbitral award under the FAA. The Court's aim, however, was to limit post-proceeding contractual freedoms in order "to maintain arbitration's essential virtue of resolving disputes straightaway." Thus, the Court

97. See id. at 1165 (stating that "[t]he greater judicial solicitude [toward arbitration] that took root in the early 1970s . . . fully blossomed by the late 1980s" and continues to grow).
99. Id.
100. See First Options of Chi., Inc. v. Kaplan, 514 U.S. 938, 943 (1995) (finding that parties may empower arbitrators to determine their own jurisdiction over certain disputes).
101. See id.
102. See id. (finding that "court[s] should give considerable leeway to the arbitrator" to determine its own jurisdiction if the parties agreed to confer such power).
103. See id. ("[I]f . . . the parties did not agree to submit the arbitrability question itself to arbitration, then the court should decide that question . . . independently." (emphasis in original)).
105. Id. at 588.
sought to uphold the national policy favoring arbitration by preserving the key benefits that arbitration provides both before and during arbitral proceedings.106

Recently, the FAA’s limits on contractual freedom were tested yet again in AT&T Mobility LLC v. Concepcion.107 In Concepcion, the issue was whether state laws which held class-arbitration waivers per se unconscionable conflicted with the FAA.108 Answering in the affirmative, the Court stated that although the FAA “preserves generally applicable contract defenses,” it does not protect state laws which “stand as an obstacle to the accomplishment of [its] objectives.”109 Thus, although California’s Discover Bank rule sowed its roots in unconscionability, a generally applicable contract defense, it was nonetheless preempted for its “disproportionate” application to arbitration agreements.110

In arriving at its conclusion, the Court declared its latest interpretation of the FAA’s core objectives. It stated that the “overarching purpose of the FAA . . . is to ensure the enforcement of arbitration agreements according to their terms so as to facilitate streamlined proceedings.”111 The two goals inherent within this objective—enforcement of arbitration agreements and facilitation of expedient dispute resolution—should neither conflict nor rank in importance. 112 Rather, the Court alluded to

106. See id. at 583 (viewing FAA §§ 9–11 “as substantiating a national policy favoring arbitration with just the limited review needed to maintain arbitration’s expediency).

107. AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1753 (2011) (finding a state law that held class arbitration waivers unconscionable preempted by the FAA, as it stood “as an obstacle to the accomplishment and execution of the full purposes and objectives of Congress” (internal quotation marks and citations omitted)).

108. See id. at 1744 (considering “whether the FAA prohibits [s]tates from conditioning the enforceability of certain arbitration agreements on the availability of classwide arbitration procedures”).

109. Id. at 1748.

110. Id. at 1747, 1753.

111. Id. at 1748.

112. See id. at 1749 (describing the FAA’s two underlying goals as “enforcement of private agreements and encouragement of efficient and speedy dispute resolution” (internal quotation marks and citations omitted)).
the possibility that an optimal rule would further both objectives in unison.\(^{113}\)

But what role do courts play in enforcing the FAA’s overarching objective in relation to arbitrators? How is it that courts retain jurisdiction to decide certain disputes, but not others? Over which disputes do courts retain jurisdiction? The answer to these inquiries is governed by one of the most important doctrinal concepts in U.S. arbitration law—severability.\(^{114}\)

### 3. Severability: A “Gateway” to Arbitration

The severability doctrine\(^{115}\) has become firmly ensconced in substantive federal arbitration law\(^{116}\) and performs two key functions.\(^{117}\) First, derived from FAA § 4,\(^{118}\) it reflects the fundamental principle that arbitrators derive their authority through party consent.\(^{119}\) In this regard, severability insulates

---

113. See id. (noting explicitly that “[i]n the present case . . . those ‘two goals’ do not conflict—and it is the dissent’s view that would frustrate both of them” (emphasis in original)).

114. See Bermann, supra note 10, at 22 (stating that severability “serves a highly salutary purpose” in arbitration law, in that it upholds “[p]arty expectations”).

115. See supra note 87 and accompanying text (describing the severability doctrine).


117. See Bermann, supra note 10, at 23 (finding that severability in U.S. arbitration law performs “twin functions”).

118. See 9 U.S.C. § 4 (2011) (providing parties the ability to petition a federal court to compel arbitration); Prima Paint, 388 U.S. at 403 (finding that the severability doctrine has its roots in FAA § 4, which requires federal courts to compel arbitration “once it is satisfied that the making of the agreement for arbitration . . . is not in issue” (internal quotation marks and citations omitted)).

119. See AT&T Techs., Inc. v. Comm’ns Workers of Am., 475 U.S. 643, 648–49 (1986) (finding the notion that arbitration requires consent an “axiom” which “recognizes the fact that arbitrators derive their authority to resolve disputes
the arbitral process from challenges directed toward the underlying contract, preserving the forum in which the parties agreed to resolve their disputes. In effect, the doctrine ensures that parties’ disputes are resolved according to their legitimate expectations, thus maintaining arbitration’s stabilizing effect.

Second, severability “permits courts to entertain challenges specifically applicable to the arbitration agreement, [but] not to the contract as a whole.” This function implicates the critical “demarcation between gateway and non-gateway issues.” Indeed, some of the issues most salient to the “tradeoff between [the] efficacy and legitimacy” of the arbitral process occur at the outset of a dispute, before the tribunal is constituted. The critical inquiry at this stage is whether a party is contractually obligated to arbitrate, notwithstanding its objections to the contrary. The answer requires a delicate balancing between arbitration’s consensual foundation and the costs and delays appurtenant to ensuring such consent. Gateway issues thus

only because the parties have agreed in advance to submit such grievances to arbitration”); Gateway Coal Co. v. United Mine Workers of Am., 414 U.S. 368, 374 (1974) (“The law compels a party to submit his grievance to arbitration only if he has contracted to do so.”); United Steelworkers of Am. v. Warrior & Gulf Navigation Co., 363 U.S. 574, 582 (1960) (“For arbitration is a matter of contract and a party cannot be required to submit to arbitration any dispute which he has not agreed so to submit.”).

120. See Bermann, supra note 10, at 22 (One purpose [of severability] . . . is to enable an arbitral tribunal to declare a contract invalid or unenforceable on the merits, without thereby necessarily destroying the basis of its authority to make that very ruling.).

121. See id. (“Party expectations concerning arbitration would clearly be disserved if arbitral tribunals were deemed, by virtue of deciding that a contract is invalid, to deprive themselves of the legal authority to make that very decision.”).

122. See Prima Paint Corp. v. Flood & Conklin Mfg. Co., 388 U.S. 395, 403–04 (1967) (instructing federal courts to “order arbitration to proceed once it is satisfied that ‘the making of the agreement for arbitration . . . is not in issue’”); Berman, supra note 10, at 23.

123. See Bermann, supra note 10, at 22.

124. Id. at 5.

125. See id. (noting that the critical question at the outset of a dispute “is whether a party unwilling to arbitrate is obligated, on the basis of a prior undertaking, to do so”).

126. See id. (“[A] party cannot be bound by an agreement to arbitrate . . . unless it consented to be so bound. On the other hand, arbitration becomes a less effective means of dispute resolution to the extent that, prior to
encompass those “jurisdictional or threshold” issues, grounded in
c consent, which courts will resolve rather than leave for the
tribunal.127 By contrast, non-gateway issues, such as substantive
claims arising from the parties’ underlying contract, are for
tribunals to decide.128 In this regard, severability acts as a
“jurisdiction-allocating device” between challenges directed
toward the underlying contract, and those directed toward the
arbitration clause itself.129 The former are non-gateway issues, as
they encompass the very contractual issues which parties agree
to arbitrate.130 The latter, however, are gateway issues because
they directly implicate “the tribunal’s authority to decide
anything—including the validity and enforceability of the main
contract.”131

Ultimately, the role of severability as a gatekeeper to the
arbitral forum is critical. It remains the guidepost for allocating
jurisdiction over certain issues between courts and arbitrators at
the outset of a dispute.132 But once a gateway issue has been
raised, how does severability impact a court’s decision to uphold,
or deny, parties’ arbitral rights?

127. See id. at 7. Professor Bermann notes that although the term “gateway
issue” has been read broadly by the Supreme Court to encompass a variety of
jurisdictional issues, the Court’s most recent interpretation, and the one
adopted by both Professor Bermann and this Note, subscribe to the narrower
meaning described above. Id. at 7–8. See also Rent-A-Ctr., W., Inc., v. Jackson,
130 S. Ct. 2772, 2777 (2010) (“[P]arties can agree to arbitrate ‘gateway
questions of arbitrability,’ such as whether the parties have agreed to arbitrate
or whether their agreement covers a particular controversy.”).

128. See Bermann, supra note 10, at 8 (describing non-gateway issues as
“those that courts reserve for initial determination, along with the merits, to the
tribunal itself”).

129. Id. at 24.

130. See id. at 23 (“The question whether a contract on which a claim in
arbitration is predicated exists and is valid . . . clearly forms part of the merits
of a case, and as such falls . . . within the arbitrators’ province to resolve.”).

131. Id.

132. See id. at 24 (noting that severability is “the touchstone for determining
whether courts should initially entertain challenges to the enforceability of an
arbitration agreement or refrain from doing so”).
4. The Impact of Severability on Parties’ Arbitral Rights in the Courts

The Supreme Court has solidified the severability doctrine’s place in federal arbitration law in three key cases. First, in *Prima Paint*, the Court applied the doctrine in light of a claim that because the overall contract was fraudulently induced, the arbitration clause within the contract was invalid as well.133 After finding that claims for fraud fell within the scope of the parties’ arbitration agreement, the Court affirmed the trial court’s decision to stay judicial proceedings pending arbitration.134 In doing so, it found that FAA § 4 instructs courts to compel arbitration if a challenge is not directed toward the arbitration clause itself.135 As such, the severability doctrine was conceived.

Almost forty years after *Prima Paint*, the Supreme Court revisited the severability doctrine136 in *Buckeye Check Cashing, Inc. v. Cardegna*.137 In *Buckeye*, however, the Florida Supreme Court refused to enforce an otherwise valid arbitration agreement on grounds that the underlying contract was illegal, rather than fraudulent.138 Rejecting this distinction, the Supreme Court found that state courts must apply the severability doctrine in the same manner as federal courts.139 Further, after *Buckeye*, the Court

---


134. See id. at 399–400, 406 (finding the language “(a)ny controversy or claim arising out of or relating to this Agreement” sufficiently broad to encompass Prima Paint’s claim for fraud).

135. See id. at 403 (finding that a court must “order arbitration to proceed once it is satisfied that the making of the agreement for arbitration . . . is not in issue” (internal quotation marks and citations omitted)).

136. See Steven J. Ware, *Arbitration Law’s Separability Doctrine After Buckeye Check Cashing, Inc. v. Cardegna*, 8  NEV. L.J. 107, 107 (2008) (noting that *Buckeye* “is only the second Supreme Court decision applying the separability doctrine and it comes nearly forty years after” *Prima Paint*).

137. See Buckeye Check Cashing, Inc. v. Cardegna, 546 U.S. 440, 446 (2006) (finding claims against an entire agreement under state usury and consumer protection laws arbitrable under FAA § 2).

138. See id. (noting that the Florida Supreme Court declined “to apply *Prima Paint’s* rule of severability” on grounds that severability cannot apply to contracts found illegal or void).

139. See id. at 449 (“[R]egardless of whether the challenge is brought in federal or state court, a challenge to the validity of the contract as a whole, and
made clear that challenges directed toward the underlying contract are “irrelevant” for severability purposes. Thus, an otherwise severable arbitration agreement will be upheld, even if claims for fraud, illegality, or breach of contract permeate the underlying agreement.

Finally, in Rent-A-Center, West, Inc. v. Jackson, the Supreme Court upheld a provision, as part of a stand-alone arbitration agreement, which delegated all disputes arising out of the parties’ employment contract to arbitration. In doing so, it affirmed the notion established in Buckeye, finding that the “[a]pplication of the severability rule does not depend on the substance of the remainder of the contract.” Thus, even if “the underlying contract is itself an arbitration agreement,” the severability doctrine protects the delegation provision unless challenged directly.

Ultimately, the severability doctrine protects arbitration clauses as binding, autonomous contracts within contracts. It supplements FAA § 2 in ensuring that arbitration agreements are afforded the full spectrum of contract-law protections, regardless of any challenges directed toward the underlying contract.

not specifically to the arbitration clause, must go to the arbitrator.”).

140. Id. at 446.
141. See id. (stating that a valid arbitration agreement will be upheld, regardless of whether claims for “fraud, misrepresentation, breach of contract, breach of fiduciary duty,” or public policy are directed at the overall contract).
143. See id. at 2775 (describing the parties’ employment relationship and arbitration agreement).
144. Id. at 2779.
145. Id.
146. See Bermann, supra note 10, at 22 (noting that the severability doctrine “basically posits that an arbitration agreement constitutes an agreement separate and apart from the main contract”); Ware, supra note 136, at 109 (stating that the severability doctrine treats an “arbitration clause as if it is a separate contract from the contract containing the arbitration clause”).
147. See Jackson, 130 S. Ct. at 2777–78 (finding that an arbitration agreement “is simply an additional, antecedent agreement” which is “valid under [FAA] § 2 ‘save upon such grounds as exist at law or in equity for the revocation of any contract’” (quoting 9 U.S.C. § 2 (2011))); see also AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1745–46 (2011) (stating that FAA § 2 reflects “the fundamental principle that arbitration is a matter of contract”
this regard, the doctrine isolates and magnifies the discrete benefits derived through arbitration. These benefits do not stem from the parties’ underlying agreement, but from the arbitration clause itself—the ex ante stability and predictability which promote parties’ contractual arrangements. As such, gateway issues involving direct challenges to an arbitration agreement must be resolved in accordance with contract law—no more, no less. In this regard, it is next critical to discern whether waiver is a gateway issue to which contract law must apply.

5. The Origins of the “Waiver” Doctrine in the Arbitral Context: Correctly Applied?

In the arbitral context, waiver refers to whether a party, through words or conduct, has somehow lost its right to arbitrate. Federal courts construe the term waiver from FAA § 3, finding the word “default” in that provision analogous “to waiver or laches or estoppel.” In application, the courts distinguish “between contract-based waiver and conduct-based

and, as such, may only be invalidated by “generally applicable contract defenses” (internal quotation marks and citations omitted)).

148. See supra Part II (discussing the benefits that parties derive from arbitration).

149. See Doctor’s Assocs., Inc. v. Casarotto, 517 U.S. 681, 687 (1996) (“[G]enerally applicable contract defenses, such as fraud, duress, or unconscionability, may be applied to invalidate arbitration agreements without contravening [FAA] § 2.”); Perry v. Thomas, 482 U.S. 483, 492 n.9 (1987) (“A court may not, then, in assessing the rights of litigants to enforce an arbitration agreement, construe that agreement in a manner different from that which it otherwise construes nonarbitration agreements under state [contract] law.”).

150. See Prima Paint Corp. v. Flood & Conklin Mfg. Co., 388 U.S. 395, 404 n.12 (1967) (“As the ‘saving clause’ in [FAA] § 2 indicates, the purpose of Congress in 1925 was to make arbitration agreements as enforceable as other contracts, but not more so.”).

151. See supra Part I (explaining that the concept of waiver which this Note addresses is waiver through pretrial conduct, not express waiver).

152. See 9 U.S.C. § 3 (2011) (stating that courts must stay litigation pending arbitration, provided that the party applying for the stay is “not in default in proceeding with such arbitration”).

waiver, holding that the former is for the arbitral tribunal to decide, while the latter may be determined at the threshold by a court.154 Contract-based waiver, for example, occurs when a party waives its right to arbitrate a certain claim arising out of the underlying contract.155 Conduct-based waiver, however, occurs when a party waives its right to invoke the arbitration agreement by, for example, litigating disputes which the parties agreed to arbitrate.156

The Supreme Court has yet to address conduct-based waiver, or to define “default” under FAA § 3. Indeed, the Court’s only reference to waiver occurred in the contract-based context—in discerning the scope of arbitrable claims.157 Critically, however, the Court has found that FAA “§ 3 adds no substantive restriction to § 2’s enforceability mandate.”158 Rather, § 2 creates


155. See Moses H. Cone Mem’l Hosp. v. Mercury Constr. Corp., 460 U.S. 1, 7, 29 (1983) (finding that the question as to whether a party has “lost any right to arbitrat[e]” the “underlying contractual dispute” due to “waiver, laches, estoppel, [or] failure to make a timely demand for arbitration” is for the arbitrator to decide); see also Green Tree Fin. Corp. v. Bazzle, 539 U.S. 444, 452 (2003) (finding disputes “about what the arbitration contract in each case means (i.e., whether it forbids the use of class arbitration procedures) is a dispute relating to this [underlying] contract and the resulting ‘relationships’” and is for “an arbitrator, not a judge” (internal quotation marks and citations omitted)); Howsam v. Dean Witter Reynolds, Inc., 537 U.S. 79, 84 (2002) (finding that “procedural questions which grow out of the dispute and bear on its final disposition’ are presumptively not for the judge, but for an arbitrator, to decide” (emphasis in original) (citations omitted)).

156. See, e.g., Grigsby & Assocs., Inc. v. M Sec. Inv., 664 F.3d 1350, 1353 (11th Cir. 2011) (“It is presumptively for the courts to adjudicate disputes about whether a party, by earlier litigating in court, has waived the right to arbitrate.”); JPD, Inc. v. Chronimed Holdings, Inc., 539 F.3d 388, 393 (6th Cir. 2008) (holding that “the court, not the arbitrator, presumptively evaluates whether a defendant should be barred from seeking a referral to arbitration because it has acted inconsistently with reliance on an arbitration agreement”); Ehleiter v. Grapetree Shores, Inc., 482 F.3d 207, 219 (3d Cir. 2007) (finding that issues surrounding waiver by pretrial conduct are for the “court to decide itself”); Marie v. Allied Home Mortg. Corp., 402 F.3d 1, 14 (1st Cir. 2005) (finding that the Supreme Court “did not intend to disturb the traditional rule that waiver by conduct, at least when due to litigation-related activity, is presumptively an issue for the court”).

157. See Moses H. Cone, 460 U.S. at 24 (“The basic issue presented in Mercury’s federal suit was the arbitrability of the [underlying] dispute between Mercury and the Hospital.”).

“substantive federal law” which incorporates traditional state
contract law principles in governing “the enforceability of
arbitration agreements.”159 Therefore, neither § 3 nor § 4 alter the
“background principles of state contract law” for determining the
validity and enforceability of arbitration agreements.160

Ultimately, the current doctrine illuminates how courts
should approach the conduct-based waiver analysis. First,
conduct-based waiver implicates the “obligation to arbitrate,
rather than the contract’s substantive obligations.”161 Thus,
although courts and commentators have criticized this
approach,162 conduct-based waiver is properly considered a
gateway issue under the severability doctrine.163 Second, FAA § 2
requires that courts enforce arbitration agreements in accordance
with generally applicable contract law principles.164 Finally, as
§ 3 adds no substantive element to § 2’s enforceability mandate,
waiver, as derived from § 3, must meet the definition of waiver
prescribed by contract law.165 Before turning to contract law,

159. Id. at 1901–02.
160. Id. at 1902.
161. Bermann, supra note 10, at 42.
162. See, e.g., Clyde Bergemann, Inc. v. Sullivan, Higgins & Brion, PPE
LLC, No. 08-162-KI, 2008 WL 4279632, at *1, *1 (D. Or. Sept. 18, 2008)
(criticizing the notion, in light of prior Supreme Court jurisprudence, that
waiver of any kind is for the court to decide); Bermann, supra note 10, at 43
(“More general considerations of efficacy and legitimacy suggest that all claims
of waiver should be determined initially by the arbitrators.”). Professor
Bermann explains that the conduct-based, contract-based waiver distinction is
another issue currently arising among federal courts. Id. Namely, the issue
concerns whether conduct-based waiver is, in fact, a gateway issue. This Note
does not address this issue. Rather, it adopts Professor Bermann’s notion that,
from a purely doctrinal perspective, waiver by conduct should be considered a
gateway issue under the severability doctrine and, thus, for the courts to decide.
Id. at 42.

163. See Bermann, supra note 10, at 42 (stating that, for severability
purposes, waiver “should be treated as a gateway issue”).
(“When deciding whether the parties agreed to arbitrate a certain
matter . . . courts generally . . . should apply ordinary state-law principles that
govern the formation of contracts.”); Perry v. Thomas, 482 U.S. 483, 492 n.9
(1987) (stating that a court may not “in assessing the rights of litigants to
enforce an arbitration agreement, construe that agreement in a manner
different from that in which it . . . construes nonarbitration agreements under
state law”).
however, it is first appropriate to address the circuit split surrounding this issue. Doing so will better frame the doctrinal conflict between arbitration and contract law engendered by the circuits’ waiver analyses; a conflict which clearly contravene the FAA’s overarching objective and the Supreme Court’s interpretation of its provisions.\footnote{See AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1748 (2011) (“The overarching purpose of the FAA . . . is to ensure the enforcement of arbitration agreements according to their terms so as to facilitate streamlined proceedings.”).}

\section*{III. The Circuit Split}

The Supreme Court’s gradual acceptance of arbitration agreements as contracts\footnote{See supra Part II (discussing the evolution of the Supreme Court’s FAA jurisprudence over the last half-century).} has caused “several oddities” to develop within the lower courts’ FAA jurisprudence.\footnote{Jeffrey W. Stempel, Reconsidering the Employment Contract Exclusion in Section 1 of the Federal Arbitration Act: Correcting the Judiciary’s Failure of Statutory Vision, 1991 J. DISP. RESOL. 259, 261.} Particularly, a circuit split has formed regarding the following question: Must a party resisting arbitration (nonmovant) show prejudice in order to prove that the party demanding arbitration (movant) waived its right to arbitrate by engaging in pretrial conduct?\footnote{See supra Part I (framing the issues surrounding conduct-based waiver).} The divergence between the circuits denies predictability as to when, and in what forum—litigation or arbitration—parties’ claims will be adjudicated.\footnote{See supra Part I (describing the problems created by the inconsistency among the circuits for parties attempting to bargain for optimal arbitral procedures).} As unpredictability begets instability, the current split denies contracting parties the key benefits derived through bargained-for arbitral procedures.\footnote{See supra Part II (describing the benefits derived through arbitration).} A succinct overview of the circuits’
waiver analyses aptly illustrates how ineffective and doctrinally inaccurate the waiver doctrine truly is.

A. The Majority: Circuits Requiring Prejudice

1. Strict Enforcement: Circuits Imposing Burdensome Prejudice Requirements

An overview of the circuits which impose burdensome prejudice requirements illustrates that the circuit split does not consist of a simple dichotomy—minority versus majority. Rather, it clearly conveys that no circuits’ standards are exactly the same.

a. The Second Circuit

The Second Circuit carries a strong presumption in favor of arbitration, and waiver will not be lightly inferred.\(^{172}\) In order to determine whether a party has waived its right to arbitrate, the court considers the time elapsed in litigation, the total amount of litigation conduct, and the degree of prejudice inflicted by the movant through such conduct.\(^{173}\) To find prejudice, the court considers the amount of discovery conducted which was not available in arbitration, delay, expense,\(^{174}\) and attempts to arbitrate motions that were previously lost on the merits in court.\(^{175}\) Prejudice is the critical component to the Second Circuit’s waiver analysis.\(^{176}\) Thus, courts will find waiver only if

---

173. See id. (noting that its test considers “the time elapsed from the commencement of litigation to the request for arbitration; . . . the amount of litigation (including exchanges of pleadings, any substantive motions, and discovery); and . . . proof of prejudice”).
174. See id. (stating that proof of prejudice includes “taking advantage of pre-trial discovery not available in arbitration, delay, and expense”).
175. See Thyssen, Inc. v. Calypso Shipping Corp. S.A., 310 F.3d 102, 105 (2d Cir. 2002) (“Prejudice can be substantive, such as when a party loses a motion on the merits and then attempts, in effect, to relitigate the issue by invoking arbitration . . . .” (internal quotation marks and citations omitted)).
176. See id. (stating that “[t]he key to the waiver analysis is prejudice”).
the nonmovant proves adequate prejudice. For example, in *Rush v. Oppenheimer & Co.*, the court found that Oppenheimer & Co.’s pretrial conduct did not cause sufficient prejudice to warrant a finding of waiver.

In *Rush*, Rush opened an options trading account with Oppenheimer that required him to sign an arbitration agreement. After disputes arose concerning his account, however, Rush filed claims in federal court. After nearly eight months of pretrial activity, Oppenheimer moved to compel arbitration. During that time, Oppenheimer engaged in extensive discovery, brought a motion to dismiss, and raised numerous affirmative defenses to Rush’s complaint, all without demanding arbitration.

On appeal, the Second Circuit reversed the district court’s finding of waiver for three key reasons. First, it found that expense, delay, and motions to dismiss, standing alone, cannot cause sufficient prejudice to warrant a finding of waiver. Second, it found that Oppenheimer’s demand for arbitration only

---

177. See *Rush v. Oppenheimer & Co.*, 779 F.2d 885, 887 (2d Cir. 1985) (stating that waiver will be found “only when prejudice to the other party is demonstrated”).

178. See id. (summarizing the Second Circuit’s burdensome waiver standard).

179. See id. at 890 (“As indicated herein we think that . . . [movant’s] right to arbitrate has not been waived.”).

180. See id. at 886 (referencing the arbitration agreement as stating “that any controversy between the parties would be settled by arbitration”).

181. See id. (noting that Rush “alleged improper and excessive trading in his account”).

182. See id. at 885 (stating that “[a]fter approximately eight months of pretrial proceedings in the district court,” Oppenheimer moved to compel arbitration).

183. See id. at 887 (stating that Oppenheimer engaged in “rather extensive discovery, [brought] a motion to dismiss, and pos[ed] thirteen affirmative defenses to [Rush’s] amended complaint, all without raising the right to arbitration”).

184. See id. (reversing the district court, stating that its findings were insufficient to establish waiver).

185. See id. at 887–88 (noting that “delay in seeking arbitration” for eight months “is insufficient by itself to constitute” waiver, and that motions to dismiss are to be expected when “a plaintiff files an intricate complaint, setting forth numerous claims . . . partially related to, the arbitrable claims” (citations omitted)).
after the district court allowed Rush’s claim for punitive damages was not prejudicial.\textsuperscript{186} Indeed, the court rejected the notion that such conduct amounted to forum shopping.\textsuperscript{187} Rather, it found that Rush was “no worse off” than if Oppenheimer moved to compel arbitration at the outset of the dispute, as Rush could not claim punitive damages in arbitration at any time.\textsuperscript{188} Third, the court invoked precedent, and declined to repudiate prior Second Circuit decisions which refused to find prejudice in more egregious factual circumstances.\textsuperscript{189} Ultimately, the court found that Oppenheimer did not cause sufficient prejudice to warrant a finding of waiver.\textsuperscript{190} In another case, however, the court found waiver where a party delayed for eight months before demanding arbitration; submitted numerous answers, defenses, and counterclaims, none of which mentioned arbitration; actively pursued discovery; waited until just before trial to compel arbitration; and offered what the court considered a “disingenuous” excuse for delay.\textsuperscript{191}

\textit{b. The Fourth Circuit}

In the Fourth Circuit, waiver will be found if a movant so “substantially utilizes the litigation machinery” that compelling arbitration would prejudice the nonmovant.\textsuperscript{192} The critical inquiry

\textsuperscript{186} See \textit{id.} at 890 (explaining why Oppenheimer’s attempts to change forums was not prejudicial).

\textsuperscript{187} See \textit{id.} (“This is not an instance in which a party sensing an adverse court decision [is, in effect, allowed] a second chance in another forum.” (internal quotation marks and citations omitted)).

\textsuperscript{188} See \textit{id.} (finding that Rush “is no worse off proceeding now to arbitration than had [Oppenheimer] moved for arbitration immediately after being served with the . . . complaint” (internal quotation marks and citations omitted)).

\textsuperscript{189} See \textit{id.} (reasoning that “[s]ince [Rush] would not have been prejudiced by a later motion to arbitrate had [Oppenheimer’s] motion to dismiss been completely denied, [its] motion to compel arbitration following reversal . . . of [its] initial, partial success” could not have prejudiced “Rush either” (citing Sweater Bee By Banff, Ltd. v. Manhattan Indus., Inc., 754 F.2d 457, 466 (2d Cir. 1985))).

\textsuperscript{190} See \textit{id.} (finding that Oppenheimer’s “right to arbitrate has not been waived”).

\textsuperscript{191} See Leadertext, Inc. v. Morganton Dyeing & Finishing Corp., 67 F.3d 20, 26 (2d Cir. 1995).

\textsuperscript{192} MicroStrategy, Inc. v. Lauricia, 268 F.3d 244, 249 (4th Cir. 2001).
is whether the nonmovant has satisfied its “heavy burden” to prove prejudice. To find prejudice, the court considers the totality of the movant’s delay and pretrial conduct. For example, in *MicroStrategy, Inc. v. Lauricia*, the court found that Lauricia failed to prove the degree of prejudice necessary to establish waiver.

In *MicroStrategy*, MicroStrategy responded to Lauricia’s employment discrimination charges by filing three separate claims against her in state and federal court, despite the arbitration clause in her employment contract. Six months after its initial filing, MicroStrategy moved to compel arbitration. On appeal, the Fourth Circuit reversed the district court’s finding of waiver on two key grounds. First, it refused to find prejudice when most of the claims adjudicated in court were unrelated to those subject to arbitration. Second, it found that Lauricia failed to prove whether MicroStrategy used pretrial discovery procedures to obtain information which would have been unavailable in arbitration. Thus, the court concluded that Lauricia failed to prove a sufficient degree of prejudice to establish waiver. Similarly, the Fourth Circuit has refused to

---

193. *Id.* (internal quotation marks and citations omitted).

194. *See id.* (finding that “delay and the extent of the moving party’s trial-oriented activity are material factors in assessing a plea of prejudice”).

195. *See id.* at 254 (reversing the “district court’s conclusion that MicroStrategy waived its right to arbitration”).

196. *See id.* (finding that Lauricia failed to prove “that she suffered the kind of prejudice necessary to support a finding that MicroStrategy waived its right to arbitrate”).

197. *See id.* at 246 (noting that the clause required the parties “to arbitrate any controversy or claim arising out of or relating to [the] . . . employment relationship”).

198. *See id.* at 250 (“[T]he time between the filing of the first action and the arbitration request was . . . six months.”).

199. *See id.* at 248 (reversing the district court’s determination that MicroStrategy waived its right to arbitrate).

200. *See id.* at 251 (refusing to find prejudice where most of the litigation was “directed to claims unrelated to those” subject to arbitration).

201. *See id.* at 254 (refusing to “conclude that Lauricia was prejudiced by the minimal amount of information obtained by MicroStrategy that” was likely obtainable in arbitration).

202. *See id.* (“Because Lauricia has failed to establish that she suffered the kind of prejudice necessary to support a finding that MicroStrategy waived its
find prejudice where a movant delayed its demand for arbitration for eight months; filed “affirmative defenses, engaged in discovery, and responded to motions”; moved for arbitration three months before trial; and inflicted undue costs because of its conduct.203

c. The Fifth Circuit

Similarly, the Fifth Circuit finds waiver “when the party seeking arbitration substantially invokes the judicial process to the detriment or prejudice of the other party.”204 The court holds a presumption against finding waiver, and places a heavy burden on nonmovants to prove that waiver is appropriate.205 Finally, prejudice results when a movant inflicts delay, expense, or forces the nonmovant to arbitrate issues already disputed in court.206 Applying its test, the court in *Walker v. J.C. Bradford & Co.*207 found that Plaintiffs failed to show a sufficient degree of prejudice to warrant a finding of waiver.208

In *Walker*, Plaintiffs filed suit in state court, alleging various state securities law violations, despite the arbitration agreement in their contract with Bradford.209 Thirteen months after Plaintiffs’ initial filing, Bradford removed the case to federal court to insist on arbitration, the district court erred by denying MicroStrategy’s motion to compel arbitration.”).


204. Walker v. J.C. Bradford & Co., 938 F.2d 575, 577 (5th Cir. 1991) (internal quotation marks and citations omitted).

205. See id. (holding “a presumption against finding waiver” and placing “a heavy burden of proof” on nonmovants attempting to establish waiver).

206. See Republic Ins. Co. v. PAICO Receivables, LLC, 383 F.3d 341, 346 (5th Cir. 2004) (finding prejudice when a movant inflicts “delay, expense, or damage to a [nonmovant’s] legal position,” which occurs when the nonmovant is forced to arbitrate an issue which was previously litigated (internal quotation marks and citations omitted)).

207. See *Walker*, 938 F.2d at 578 (finding that Bradford’s “actions in federal court were not so substantial as to mandate that we overcome the legal presumption” against waiver).

208. See id. (finding that Plaintiffs did not present “enough evidence” that Bradford’s conduct “materially prejudiced them”).

209. See id. at 576 (describing the pretrial process in which the parties engaged before Bradford moved to compel arbitration).
court and subsequently moved to compel arbitration. On appeal, the Fifth Circuit reversed the district court’s finding of waiver on three grounds. First, the court cast Plaintiffs’ claims of cost and delay as “generalized protestations,” insufficient to “overcome the strong federal presumption in favor of arbitration.” Second, the court found Bradford’s pretrial conduct—its removal to federal court, preliminary interrogatories, document requests, and answer—insufficient, in light of prior precedent, to establish prejudice. Finally, the court refused to find discovery prejudicial without proof that Bradford obtained information which was unavailable in arbitration. Ultimately, the court found that Plaintiffs failed to satisfy their heavy burden to prove prejudice and, thus, establish waiver. Similarly, the Fifth Circuit has refused to find prejudice where a movant delayed for eight months before demanding arbitration; filed an answer, interrogatories, and document production requests; moved for protective orders; and agreed to a joint motion to extend the discovery period.

210. See id. (“Thirteen months after plaintiffs filed suit, defendant filed a motion to compel arbitration and to stay proceedings.”).

211. See id. at 577 (reversing the district court's finding of waiver).

212. Id. at 578.

213. See id. at 576–77 (discussing Bradford's pretrial conduct and citing prior decisions in which the court refused to find prejudice where parties “invoked the judicial process to approximately the same extent as Bradford” (citing Tenneco Resins, Inc. v. Davy Int’l, A.G., 770 F.2d 416, 420–21 (5th Cir. 1985))).

214. See id. at 578 n.3 (stating that that if Bradford's discovery conduct “revealed items that would not be discoverable in arbitration proceedings,” it would “be more likely to find that plaintiffs were prejudiced”); see also MicroStrategy, Inc. v. Lauricia, 268 F.3d 244, 251 (4th Cir. 2001) (finding that if “the same information could have been obtained in an arbitration proceeding,” then the nonmovant suffers no prejudice).

215. See Walker v. J.C. Bradford & Co., 938 F.2d 575, 578 (5th Cir. 1991) (finding Plaintiffs’ evidence “insufficient to overcome the strong federal presumption in favor of arbitration” (citations omitted)).

The Sixth Circuit addresses arbitral issues “in light of the strong federal policy” favoring arbitration. The court holds a presumption in favor of arbitration under its two-pronged test, and waiver will not be lightly inferred. Indeed, the court will find waiver only where a movant takes actions that are inconsistent with any reliance on the arbitration agreement, and where it delays demanding arbitration to an extent which prejudices the nonmovant. Prejudice may be found when a movant inflicts undue delay and expense. For example, in Hurley v. Deutsche Bank Trust Co. Americas, the court found waiver where Defendants’ conduct satisfied both elements of its test.

In Hurley, the Hurleys filed federal statutory and state law claims in federal court against Defendants, despite the arbitration clause in their mortgage documents. After two years of pretrial activity, Defendants moved to compel arbitration. On appeal, the Sixth Circuit affirmed the district court.
court’s finding of waiver under its two-pronged test. First, it found Defendants’ persistent and active pretrial activity over the course of two years inconsistent with any reliance on the parties’ arbitration agreement. Second, the court found that because the Hurleys conducted substantial discovery, argued numerous summary judgment motions, and changed venue at Defendants’ request, they suffered prejudice as a result of Defendants’ delay. Thus, in light of such conduct, the court found that Defendants waived their right to arbitrate. Similarly, the Sixth Circuit has found waiver where a movant delayed demanding arbitration for one year, during which time it engaged in extensive discovery and filed numerous pretrial motions.

e. The Ninth Circuit

The Ninth Circuit also applies its test in light of the liberal federal policy favoring arbitration agreements. Furthermore, nonmovants bear a heavy burden of proof in establishing each element of its three-pronged test. Particularly, a nonmovant must show that the movant knew of its right to compel arbitration; that the movant acted inconsistently with that right;
and that it suffered prejudice because of the movant’s delayed demand for arbitration. For example, in *Fisher v. A.G. Becker Paribas, Inc.*, the court refused to find waiver when the Fishers failed to prove each element of its test.

In *Fisher*, the Fishers filed claims in federal court alleging violations of various federal securities and state common laws. Three-and-one-half years after the Fishers filed their claims, Becker moved to compel arbitration. During that time, both parties “filed pretrial motions and engaged in extensive discovery.” On appeal, the Ninth Circuit reversed the district court’s finding of waiver on two key grounds. First, it found that Becker did not act inconsistently with its right to arbitrate because, under the intertwining doctrine, the Fishers’ arbitrable claims could not be separated from their nonarbitrable securities claims, making the entire dispute nonarbitrable. The Supreme Court has since rejected both the intertwining doctrine and the notion that federal securities claims are not arbitrable.

---

232. *See id.* (finding waiver where the nonmovant shows that “(1) [the movant] had knowledge of its existing right to compel arbitration; (2) [the movant] acted inconsistently with that existing right; and (3) [the nonmovant] suffered prejudice from [the movant’s] delay in moving to compel arbitration”).

233. *See Fisher v. A.G. Becker Paribas, Inc.*, 791 F.2d 691, 694–98 (9th Cir. 1986) (describing extensive pretrial conduct by a movant that did not amount to a waiver of the right to arbitrate).

234. *See id.* at 698 (“The Fishers have failed to support their contention that Becker acted inconsistently with an existing right to compel arbitration . . . [or] to demonstrate any prejudice resulting from the alleged inconsistent acts.”).

235. *See id.* at 693 (noting that the Fishers alleged violations of “federal securities laws as well as . . . [state] common law claims”).

236. *See id.* (finding that Becker delayed demanding arbitration for three and one-half years).

237. *Id.*

238. *See id.* at 698 (reversing the district court’s finding of waiver and ordering that “any arbitrable claims be submitted to arbitration immediately”).

239. *See id.* at 694–95 (noting that the intertwining doctrine “holds that when it is impractical if not impossible to separate out nonarbitrable from arbitrable claims, a court should deny arbitration in order to preserve its exclusive jurisdiction over federal securities claims” (internal quotation marks and citations omitted)).

the court found that Becker’s failure to raise arbitration as an affirmative defense, the possibility that there may be some duplication of efforts in litigation and arbitration, and the Fishers’ own failure to demand arbitration by filing claims in court, were factors in concluding that the Fishers failed to establish prejudice and, thus, waiver.241

Although Fisher may simply be a product of its time,242 the Ninth Circuit continues to require a strong showing of prejudice. For example, it has found prejudice only where a movant’s actions caused “staleness of [a] claim” and subjected the nonmovant to litigation in state court, including discovery, the costs of litigation, and a judgment on the merits.243 The court also refused to find prejudice when a nonmovant incurred significant costs in the pretrial stages of litigation, but expressed more of a willingness to find prejudice had the case proceeded through discovery and a trial.244

f. The Eleventh Circuit

Finally, the Eleventh Circuit applies its two-pronged test in light of the strong federal policy favoring arbitration.245 It foists a heavy burden upon parties attempting to satisfy its test and invoke waiver.246 To meet its test, a nonmovant must prove that

---

241. See Fisher v. A.G. Becker Paribas, Inc., 791 F.2d 691, 697–98 (9th Cir. 1986) (discussing the grounds on which the court refused to find prejudice).

242. See supra note 240 and accompanying text (describing the Supreme Court’s refutation of the intertwining doctrine and the notion that federal securities claims are not arbitrable).


244. See United Computer Sys., Inc. v. AT&T Corp., 298 F.3d 756, 765 (9th Cir. 2002) (stating that if the “defendants permitted the case to proceed to discovery and to a trial, an argument of prejudice . . . would be much more compelling”); but see Lewallen v. Green Tree Servicing, LLC, 487 F.3d 1085, 1093 (8th Cir. 2007) (finding that a motion to dismiss can cause sufficient prejudice to warrant a finding of waiver).

245. See Citibank, N.A. v. Stok & Assocs., P.A., 387 F. App’x 921, 923 (11th Cir. 2010) (stating that “federal law favors arbitration” (internal quotation marks and citations omitted)).

246. See id. (“[A]ny party arguing waiver of arbitration bears a heavy burden of proof.” (internal quotation marks and citations omitted)).
the movant acted inconsistently with its right to arbitrate, and that it was prejudiced by the movant’s acts. 247 A movant acts inconsistently with its right to arbitrate when its conduct, including pretrial activity, indicates an intent to avoid arbitration. 248 To find prejudice, the court considers the length of delay, expense, and damage to the nonmovant’s legal position incurred through discovery. 249 For example, in Stone v. E.F. Hutton & Co., Inc., 250 the court found Hutton’s delay and pretrial conduct sufficiently prejudicial to warrant a finding of waiver. 251

In Stone, Plaintiff filed claims in federal court despite its arbitration agreement with Hutton. 252 It alleged various federal and Florida securities law violations, as well as common law claims. 253 On appeal, the Eleventh Circuit affirmed the district court’s finding of waiver under its two-pronged test. 254 First, the court found that Hutton’s one-year-and-eight-month delay rendered its motion “untimely.” 255 Second, the court found

---

247. See Ivax Corp. v. B. Braun of Am., Inc., 286 F.3d 1309, 1315–16 (11th Cir. 2002) (articulating its test as a means to determine whether, “under the totality of the circumstances, a party has acted inconsistently with the arbitration right and, second . . . whether, by doing so, that party has in some way prejudiced the other party” (internal quotation marks and citations omitted)).

248. See Stok, 387 F. App’x at 924 (stating that a party acts inconsistently with the right to arbitrate when its conduct, “including [substantial] participation in litigation . . . manifests an intent to avoid or waive arbitration” (citations omitted)).

249. See id. (evaluating “the prejudice prong by considering the length of delay in demanding arbitration,” expenses incurred, and the extent to which pretrial discovery damaged the nonmovant’s legal position (internal quotation marks and citations omitted)).


251. See id. (finding both the “extent of discovery conducted” and the extent of Hutton’s delay sufficient to warrant a finding of waiver).

252. See id. at 1543 (noting that the arbitration agreement required “any controversy arising out of” Plaintiff’s account to be “settled by arbitration”).

253. See id. at 1542 (noting that plaintiff’s complaint alleged a violation of “section 10(b)” of the “Securities Exchange Act of 1934,” Florida securities law violations, and “common law fraud, negligence, and breach of fiduciary obligations”).

254. See id. (affirming the “district court’s order denying [Hutton’s] motion to compel arbitration”).

255. Id. at 1544.
Hutton’s conduct prejudicial because it engaged in discovery typically conducted by parties preparing for trial. Particularly, Hutton deposed Plaintiff twice and responded to document production requests, while Plaintiff submitted four sets of interrogatories, three document production requests, and scheduled numerous depositions. As such, the court concluded that Hutton waived its right to arbitrate. Similarly, the Eleventh Circuit has found waiver where a movant delayed for eight months before demanding arbitration. During that time, the movant deposed five of the nonmovant’s “employees (totaling approximately 430 pages),” and the nonmovant filed both a motion to dismiss and a motion to oppose discovery.

\[ g. \text{ A Brief Summary: The Strict Enforcement Divergence} \]

A comparative analysis of the circuits which impose burdensome prejudice requirements illustrates three significant problems. First, and most importantly, their onerous standards conflict with the FAA’s overarching objective. Each approach sacrifices procedural expediency and efficiency in the name of strict contractual enforcement. The Supreme Court, however,
has expressed a desire to avoid such conflict where possible.\textsuperscript{263} Second, the divergence between each circuit’s analysis detracts from the predictive power and stability that parties expect through bargained-for arbitral procedures.\textsuperscript{264} Compounding this problem is that courts within the same circuit often arrive at discrepant outcomes in similar factual circumstances.\textsuperscript{265}

Finally, a number of circuits which propagate burdensome prejudice requirements have regarded their holdings with disdain. For example, in MicroStrategy, the Fourth Circuit recognized that MicroStrategy took an aggressive “course of litigation for the sole purpose of wearing [Lauricia] out, both emotionally and financially.”\textsuperscript{266} Similarly, in Walker, the Fifth Circuit “sympathized with [P]laintiffs’ exasperation,” conceding that Bradford’s attempts “to switch judicial horses in midstream” wasted both the courts’ and the parties’ time.\textsuperscript{267} Further, the Second Circuit stated that it would refuse to find waiver without a strong showing of prejudice, no matter how “unjustifiable” a movant’s conduct.\textsuperscript{268} Ultimately, burdensome prejudice requirements both flout parties’ legitimate expectations of stability and contravene the FAA’s overarching objective.\textsuperscript{269} A more concrete, contractual approach is needed to unify the circuits’ analyses within the scope of the FAA.

\textsuperscript{263} See Concepcion, 131 S. Ct. at 1749 (discussing how the FAA’s two underlying goals need not, and should not, conflict).

\textsuperscript{264} See supra Part II (describing the benefits that parties expect through arbitration).

\textsuperscript{265} For example, district courts within the Sixth Circuit have applied its waiver standard in an inconsistent manner. Compare U.S. Enrichment Corp. v. Sw. Elec. Co., Inc., No. 5:07CV-36-R, 2008 WL 199881, at *1, *5 (W.D. Ky. Jan. 23, 2008) (refusing to find waiver after two years of negotiations, pretrial activity, and costs because the nonmovant reserved its right to arbitrate at the outset of negotiations and there was no “bad faith”), with Johnson Assocs. Corp. v. HL Operating Corp., No. 3:09-CV-01206, 2010 WL 4942788, at *1, *5 (M.D. Tenn. Nov. 30, 2010) (finding waiver where “the right to arbitrate was not asserted for eight months, during which” time the parties engaged in pretrial conduct).

\textsuperscript{266} MicroStrategy, Inc. v. Lauricia, 268 F.3d 244, 254 (4th Cir. 2001).


\textsuperscript{268} Leadertext, Inc. v. Morganton Dyeing & Finishing Corp., 67 F.3d 20, 26 (2d Cir. 1995).

\textsuperscript{269} See supra Part II (discussing the benefits that parties expect through bargained-for arbitral procedures).
2. Circuits Imposing Lenient Prejudice Requirements

a. The First Circuit

The First Circuit holds that the key factors to finding waiver are undue delay and a “modicum of prejudice to the other side.” To avoid waiver, the First Circuit requires that parties demand arbitration at the earliest opportunity in order to ensure that courts’ and parties’ “resources are not needlessly deployed.” For example, in Rankin v. Allstate Insurance Co., the court found Allstate’s conduct sufficiently prejudicial to warrant a finding of waiver.

In Rankin, the Rankins filed suit against Allstate in federal district court, alleging breach of contract. Nine months after the Rankins filed their claim, Allstate moved to compel arbitration. On appeal, the First Circuit first found Allstate’s nine-month delay unacceptable, as it knew of the coming dispute with the Rankins as early as one month before they filed their claim in court. Second, the court found prejudice “inherent in wasted trial preparation” when a party demands arbitration after months of delay and a short time before trial. Therefore, the

271. Id. at 14 (citations omitted).
272. Id. at 13 (citations omitted).
273. See id. at 14 (finding that Allstate waived its right to arbitrate by inflicting prejudice through undue delay).
274. See id.
275. See id. at 11 (noting that the dispute between the Rankins and Allstate concerned “[w]hether Allstate unreasonably delayed payment of what was due [under the Rankins’ insurance contract], whether it still owe[d] money, and whether it lost its right to invoke the arbitration provision”).
276. See id. at 10. (noting that although litigation began on March 2, 2001, Allstate did not demand arbitration until nine months later on December 19, 2001, less than two months before the trial date set for February 11, 2002).
277. See id. at 12–13 (noting that “by February 2001 the parties were in disagreement” about numerous issues, approximately two months before Allstate filed “its April 2001 answer to the complaint filed in March 2001” by the Rankins).
278. See id. at 14 (finding prejudice where “an arbitration demand is made . . . after many months of delay and only six weeks before . . . trial”).
court concluded that Allstate’s delay, standing alone, was sufficient to fulfill the modicum of prejudice necessary to establish waiver.\textsuperscript{279} Similarly, the First Circuit has found expenses incurred as a result of dilatory behavior sufficient to warrant a finding of prejudice, even when no useful information was acquired through discovery.\textsuperscript{280}

\textbf{b. The Eighth Circuit}

The Eighth Circuit, by contrast, employs a three-pronged test which mirrors the Ninth Circuit’s. Particularly, a party may waive its right to arbitrate if it: “(1) knew of an existing right to arbitrat[e]; (2) acted inconsistently with that right; and (3) prejudiced the other party by these inconsistent acts.”\textsuperscript{281} A party acts inconsistently with its right to arbitrate if it attempts to litigate arbitrable claims, conducts substantial discovery, or delays its demand for arbitration.\textsuperscript{282} The prejudice threshold is not “onerous,”\textsuperscript{283} and may be met when a movant inflicts unnecessary delay, expense, or when compelling arbitration would require a “duplication of efforts” in multiple forums to resolve the dispute.\textsuperscript{284} For example, in \textit{Lewallen v. Green Tree Servicing, LLC}.\textsuperscript{285}

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{279} See \textit{id.} (finding that Allstate waived “its right to arbitration”).
\item \textsuperscript{280} See \textit{Menorah Ins. Co. v. INX Reinsurance Corp.}, 72 F.3d 218, 222 (1st Cir. 1995) (finding prejudice when the nonmovant incurred unnecessary costs through pretrial discovery, even when no “information useful to the resolution of the dispute was . . . procured”).
\item \textsuperscript{281} \textit{Lewallen v. Green Tree Servicing, LLC}, 487 F.3d 1085, 1090 (8th Cir. 2007) (internal quotation marks and citations omitted); \textit{see also} \textit{Sovak v. Chugai Pharm. Co.}, 280 F.3d 1266, 1270 (9th Cir. 2002) (“Sovak must show (1) Cook had knowledge of its existing right to compel arbitration; (2) Cook acted inconsistently with that existing right; and (3) he suffered prejudice from Cook’s delay in moving to compel arbitration.”).
\item \textsuperscript{282} See \textit{Lewallen}, 487 F.3d at 1090 (finding that a party acts inconsistently with its right to arbitrate by “filing a lawsuit on arbitrable claims, engaging in extensive discovery, or failing to move to compel arbitration and stay litigation in a timely manner”).
\item \textsuperscript{283} \textit{Hooper v. Advance Am., Cash Advance Ctrs. of Mo., Inc.}, 589 F.3d 917, 923 (8th Cir. 2009).
\item \textsuperscript{284} \textit{Lewallen}, 487 F.3d at 1093 (internal quotation marks and citations omitted).
\end{itemize}
\end{footnotesize}
“WAIVING” GOODBYE TO ARBITRATION

Servicing, LLC, the court found Green Tree’s delay and pretrial conduct sufficiently prejudicial to warrant a finding of waiver.

In Lewallen, Lewallen filed for bankruptcy due to arrearages on a home loan that was purchased, and serviced, by Green Tree. In response to Green Tree’s proof of claim, Lewallen filed two rounds of discovery requests and counterclaimed for various violations of state and federal lending laws. Sixteen months later, Green Tree moved to compel arbitration pursuant to the arbitration clause in the parties’ loan agreement. Applying its test, the Eighth Circuit affirmed both the bankruptcy court’s and the district courts’ findings of waiver.

First, the Eighth Circuit found that Green Tree’s pretrial conduct—its lengthy delay, interrogatories, document production requests, and motion to dismiss—was inconsistent with its right to arbitrate. The court found Green Tree’s motion to dismiss particularly egregious, noting that it pressed the bankruptcy court to resolve Lewallen’s claims on the merits, while preserving arbitration as an alternative forum in case of an adverse

285. See id. (finding Green Tree’s conduct both inconsistent with its right to compel arbitration and prejudicial).
286. See id. at 1094 (concluding “that Green Tree waived its right to arbitrate Lewallen’s claims,” and denying Green Tree’s motion to compel arbitration).
287. See id. at 1088–89 (noting that Green Tree purchased the right to service Lewallen’s home loan, which was in default at the time of transfer, in 2002 and attempted to foreclose on her home in 2004).
288. See id. at 1089 (stating that Lewallen objected to Green Tree’s proof of claim and “alleged that Green Tree’s conduct violated the Real Estate Settlement Procedures Act, . . . the Fair Debt Collection Practices Act . . . and the Missouri Merchandising Practices Act”).
289. See id. at 1091 (quoting the arbitration agreement which “provided that [Green Tree] retains an option to use judicial or non-judicial relief to enforce a security agreement relating to the collateral secured in a transaction underlying this arbitration agreement, to enforce the monetary obligation or to foreclose on the collateral”).
290. See id. at 1090, 1094 (discussing and affirming both the bankruptcy and the district courts’ findings that Green Tree waived its right to arbitrate).
291. See id. at 1092–93 (finding that Green Tree’s merits-based motion to dismiss, discovery requests, “lengthy interrogatories,” requests for the “production of documents after adversary proceeding[s] commenced,” and motions to extend the “time to respond to Lewallen’s discovery requests” were inconsistent with its right to arbitrate).
ruling. Turning to prejudice, the court found Green Tree’s motion to dismiss and inexplicable delay prejudicial in light of Lewallen’s precarious financial situation. As such, the court concluded that Green Tree’s conduct was sufficiently prejudicial to warrant a finding of waiver. Similarly, the Eighth Circuit has found waiver when a movant delayed for four and one-half months before demanding arbitration; filed a motion to dismiss which required the nonmovants to litigate a number of substantive claims; and the movant would have sought to reargue any adverse rulings in arbitration.

c. Lenient Prejudice Requirements: A Brief Comparison

It is first important to note the obvious doctrinal divergence between the First and Eighth Circuits’ lenient prejudice requirements, and the burdensome requirements imposed by other circuits within the majority. The discrepancy between the Eighth and Ninth Circuits’ prejudice standards is particularly illustrative, given their virtually identical three-pronged waiver tests.

The First and Eighth Circuits’ prejudice standards diverge as well. The Eighth Circuit, conscious of the current circuit split regarding the prejudice requirement, has abandoned prejudice

---

292 See id. at 1092 (finding that Green Tree’s motion to dismiss essentially “urg[ed] the bankruptcy court to dispose of Lewallen’s claims on the merits, reserving arbitration as an alternative avenue to resolve the dispute”).

293 See id. at 1093 (finding the totality of Green Tree’s conduct prejudicial, particularly because any wasted “time and expense” would prejudice Lewallen at a time “when she [could] ill afford to waste resources”—in bankruptcy).

294 See id. at 1094 (concluding that “Green Tree waived its right to arbitrate Lewallen’s claims”).

295 Hooper v. Advance Am., Cash Advance Ctrs. of Mo., Inc., 589 F.3d 917, 922–23 (8th Cir. 2009) (finding waiver when the movant “waited over four-and-a-half months before filing its motion for arbitration”; when the movant’s “motion to dismiss forced Plantiffs to brief fully a number of substantive issues”; and when the movant “would presumably . . . reargue in arbitration issues it lost” in its motion to dismiss).

296 See supra note 281 and accompanying text (describing the virtually identical waiver tests shared by the Eighth and Ninth Circuits).

297 See Erdman Co. v. Phoenix Land & Acquisition, LLC, 650 F.3d 1115, 1118–19 (8th Cir. 2011) (“There is a circuit split over whether the party asserting waiver must show prejudice . . . . [on which] [t]he Supreme Court in
altogether in certain circumstances. By contrast, the First Circuit requires some showing of prejudice in all circumstances to find waiver, although a small degree is sufficient.

To be sure, both circuits’ lenient requirements have prevented manipulation and delay much more effectively than the strict-enforcement regimes. The chaotic state of the current waiver doctrine, however, deprives contracting parties of any jurisprudential predictability. Rather, a uniform, contractual approach would stabilize the current doctrine and realign the circuits’ analyses within the scope of the FAA.

3. The Third and Tenth Circuits: Multifactor Tests

Both the Third and Tenth Circuits require some showing of prejudice to find waiver. A separate analysis, however, provides an illustrative microcosm of the overall inconsistencies caused by the waiver doctrine. Despite both circuits’ facially similar multifactor waiver regimes, each differs significantly in application.

---

the Term just ended granted a petition for a writ of certiorari . . . .” (citations omitted).

298. See id. at 1120 (finding the prejudice requirement unnecessary “[i]n the realm of construction industry disputes”).

299. See Rankin v. Allstate Ins. Co., 336 F.3d 8, 13 (1st Cir. 2003) (“[T]he components of waiver of an arbitration clause are undue delay and a modicum of prejudice to the other side.”).

300. See, e.g., Menorah Ins. Co. v. INX Reinsurance Corp., 72 F.3d 218, 222–23 (1st Cir. 1995) (refusing to allow the use of arbitration agreements for purposes of “manipulation and mischief”).

301. See, e.g., Erdman, 650 F.3d at 1120 (finding that “it makes little sense to litigate endlessly over the details of prejudice” in certain circumstances); Rankin, 336 F.3d at 13 (stating that its particular concern when evaluating prejudice is “when a timely demand for arbitration must be made” in order to facilitate “efficient planning by the court” and adequate protection of “the opponent”).

302. See supra notes 266–68 and accompanying text (discussing the Second, Fourth, and Fifth Circuits’ awareness of the delay and manipulation facilitated by their burdensome prejudice requirements).
a. The Third Circuit

The Third Circuit’s “nonexclusive list of factors relevant to the prejudice inquiry” includes the timeliness of the motion to compel arbitration; the extent to which the movant has contested the merits of the nonmovant’s claims; whether the movant has informed the nonmovant of its intent to seek arbitration; the number of “nonmerits” motions submitted by the movant; the movant’s assent to the court’s pretrial orders; and the extent to which both parties conducted discovery. The court applies its test contextually, and not all of the factors need be present to establish a finding of prejudice. For example, in *Nino v. Jewelry Exchange, Inc.*, the court found waiver where only four of the six factors weighed in favor of finding prejudice.

In *Nino*, Nino brought an employment discrimination suit in federal court against DI, his former employer. After fifteen months, DI moved to compel arbitration pursuant to the arbitration clause in Nino’s employment contract. Applying its test, the Third Circuit reversed the district court’s refusal to find waiver. Particularly, it found DI’s delay; the “substantial amounts of time, effort, and money” Nino spent prosecuting the action; DI’s participation in numerous pretrial conferences; and the significant amount of discovery conducted by the parties sufficient to cause four of the six factors to weigh in favor of prejudice. As such, the court found DI’s conduct sufficiently

---

303. See *Nino v. Jewelry Exch., Inc.*, 609 F.3d 191, 208–09 (3d Cir. 2010) (internal quotation marks and citations omitted).

304. See id. at 209 (characterizing its factors as comprising “a nonexclusive list” which need not all be present to justify a finding of waiver, and which must be applied “based on the circumstances and the context of the particular case” (internal quotation marks and citations omitted)).

305. See id. at 209–13 (applying and discussing each of its six factors).

306. See id. at 213–14 (finding waiver where four of the six factors heavily indicated prejudice).

307. See id. at 196 (noting that Nino alleged “he was discriminated against on account of his gender and national origin”).

308. See id. (“After litigating the matter before the District Court for fifteen months, the employer invoked an arbitration provision . . . and moved . . . to compel the parties to arbitrate their dispute.”).

309. See id. (reversing “the District Court’s order compelling the parties to arbitrate”).

310. See id. at 213–14 (analyzing DI’s pretrial conduct and concluding that
prejudicial to warrant a finding of waiver.\textsuperscript{311} Similarly, the Third Circuit has found waiver where a movant’s ten-month delay and extensive depositions, discovery requests, and document production requests caused four of the six factors to weigh in favor of finding prejudice.\textsuperscript{312}

\textit{b. The Tenth Circuit}

The Tenth Circuit’s waiver test encompasses six factors which largely mirror the Third Circuit’s, including whether the movant: acted inconsistently with its right to arbitrate; substantially invoked the litigation machinery before notifying the nonmovant of its intent to arbitrate; either demanded arbitration close to trial, or delayed its demand for a long period of time; filed a counterclaim without demanding arbitration therein; took advantage of pretrial procedures which are unavailable in arbitration; and whether its delay “affected, misled, or prejudiced ” the nonmovant.\textsuperscript{313} The court applies these factors to supplement three considerations regarding the movant’s conduct. First, the court considers the extent to which the nonmovant was prejudiced by the movant’s conduct.\textsuperscript{314} Prejudice may be shown if the movant inflicted expense, delay, or injury to the nonmovant’s legal position.\textsuperscript{315} Second, the court

\textsuperscript{311} See id. at 214 (refusing to compel arbitration where DI’s “demand . . . came long after the suit commenced and when both parties had engaged in extensive discovery” (internal quotation marks and citations omitted)).

\textsuperscript{312} See Gray Holdco, Inc. v. Cassady, 654 F.3d 444, 454, 458–61 (3d Cir. 2011) (considering the movant’s “litigation conduct as a whole” and finding that its “motion for a preliminary injunction,” ten-month delay, engagement in “three pre-trial conferences,” court-ordered mediation, and numerous “discovery reports” was sufficient to warrant a finding of waiver).

\textsuperscript{313} See Hill v. Ricoh Ams. Corp., 603 F.3d 766, 772–73 (10th Cir. 2010) (internal quotation marks and citations omitted).

\textsuperscript{314} See id. at 775.

\textsuperscript{315} See id. (stating that the relevant considerations for finding prejudice include the “delay and costs” incurred by the nonmovant, and the degree of prejudice to the nonmovant’s legal position which “may be inferred from the extent of discovery conducted in the case” (internal quotation marks and
considers whether the movant is attempting to manipulate the judicial process. Finally, the court considers whether the movant is attempting to hinder the “combined efficiency of the public and private dispute-resolution systems.” For example, in *Hill v. Ricoh Americas Corp.*, the court refused to find waiver because Ricoh’s pretrial conduct implicated none of these three concerns.

In *Hill*, Hill sued Ricoh Americas Corp. in federal court, alleging that he was wrongfully terminated. After four months, Ricoh moved to compel arbitration pursuant to the arbitration clause in Hill’s employment contract. Applying its factors, the Tenth Circuit reversed the district court’s finding of waiver on three key grounds. First, it found that Ricoh’s delay alone was insufficient to establish waiver. Second, the court found that because the trial was five months away and minimal litigation activity occurred, granting Ricoh’s motion would neither cause inefficiency nor facilitate manipulation. Finally, the court found that Hill failed to prove prejudice in light of Ricoh’s minor delay and pretrial conduct. As a result, the court refused to find

citations omitted)).

316. See id. at 773 (“An important consideration in assessing waiver is whether the party now seeking arbitration is improperly manipulating the judicial process.”).

317. Id. at 774.

318. See id. at 775–76 (applying its six-factor waiver test).

319. See id. at 776 (refusing to find waiver because Ricoh’s “minimal litigation activity” did not cause prejudice, inefficiency, or “improper manipulation of the judicial process”).

320. See id. at 769 (noting that Hill filed a claim alleging that he was “terminated from his position at Ricoh in violation of the Sarbanes-Oxley Act”).

321. See id. at 769 n.2 (noting that the arbitration clause required “the parties [to] voluntarily agree to settle the dispute by binding arbitration”).

322. See id. at 776 (“[T]he circumstances of this case, particularly in light of the federal policy favoring arbitration, convince us that the district court should not have found waiver . . . .”).

323. See id. at 775 (“[L]ength of time in itself does not establish waiver.”).

324. See id. (noting that the only important pretrial activities shown on the record were “the magistrate judge’s setting the schedule for litigation,” which was not to begin for eleven months, and the only discovery conducted was “Hill’s request for production of documents”).

325. See id. at 775–76 (finding that Hill “failed to show any substantial prejudice” in light of Ricoh’s minimal pretrial conduct).
waiver.\textsuperscript{326} In another case, however, the Tenth Circuit found a movant’s one year delay; participation in hearings, motions, pleadings, and depositions; and failure to demand arbitration immediately after the claim was filed sufficiently prejudicial to warrant a finding of waiver.\textsuperscript{327}

c. The Third and Tenth Circuits Compared

The Third and Tenth Circuits’ waiver analyses encompass analogous sets of factors which supplement divergent considerations. The Third Circuit’s test amounts to a multifactor “prejudice inquiry.”\textsuperscript{328} Its unitary focus on prejudice could potentially aid parties in predicting what degree of pretrial conduct might trigger a waiver of arbitral rights. The test’s nonexclusive, contextual nature, however, undercuts any such potential.\textsuperscript{329} Thus, although the court recognizes that “arbitration is meant to streamline the proceedings, lower costs, and conserve private and judicial resources,” it has refused to adopt any consistent, doctrinally accurate method to accomplish these objectives.\textsuperscript{330}

\begin{itemize}
\item \textsuperscript{326} See id. at 776 (reversing the district court’s finding of waiver).
\item \textsuperscript{327} See Reid Burton Constr., Inc. v. Carpenters Dist. Council of S. Colo., 614 F.2d 698, 703 (10th Cir. 1980) (finding the one-year delay at trial; the movant’s “participation in numerous hearings, pretrial conferences, motions and other pleadings, and the deposing of witnesses”; and the movant’s failure to insist upon arbitration immediately sufficient to warrant a finding of waiver).
\item \textsuperscript{328} Nino v. Jewelry Exch., Inc., 609 F.3d 191, 208 (3d Cir. 2010) (internal quotation marks and citations omitted).
\item \textsuperscript{329} See id. at 209 (noting that its list of factors is nonexclusive, “not all the factors need be present to justify a finding of waiver, and [t]he waiver determination . . . [is] based on the circumstances and context of [a] particular case” (internal quotation marks and citations omitted)); see also Gray Holdeo, Inc. v. Cassady, 654 F.3d 444, 452 (3d Cir. 2011) (affirming the district court’s judgment, but disagreeing with its finding as to how many factors indicated prejudice); Quilllon v. Tenet Healthstystem Phila., Inc., 763 F. Supp. 2d 707, 721–22 (E.D. Pa. 2011) (refusing to find waiver when three factors weighed in favor of finding prejudice and three weighed against finding prejudice, but when those in favor were not implicated as strongly); Opalinski v. Robert Half Int’l, Inc., No. 10-2069, 2011 WL 4729009, at *1, *7 (D.N.J. Oct. 6, 2011) (refusing to find waiver when two factors weighed in favor, one only “slightly” in favor, and three against finding prejudice).
\item \textsuperscript{330} Nino, 609 F.3d at 209.
\end{itemize}
By contrast, the Tenth Circuit’s test encompasses three disparate considerations. Furthermore, it openly eschews any “mechanical” balancing or exhaustive application of its factors. As a result, the Tenth Circuit’s scattershot approach only compounds the divergent outcomes engendered by the Third Circuit’s regime. A uniform, contractual approach, however, would instill consistency among the circuits, stability for contracting parties, and doctrinal accuracy in accordance with the FAA.

B. The Minority: Circuits Not Requiring Prejudice

The Seventh and D.C. Circuits comprise a small minority which eschews prejudice as a necessary element of waiver. Its dissent from the majority, however, has been influential.

331. See supra notes 314–17 and accompanying text (describing the various considerations inherent in the Tenth Circuit’s waiver test).

332. See Hill v. Ricoh Ams. Corp., 603 F.3d 766, 773 (10th Cir. 2010) (“Of course, our listing these factors . . . was not intended to suggest a mechanical process in which each factor is assessed and the side with the greater number of factors prevails. Nor were we even suggesting that the list . . . is exclusive.”).

333. Compare Lamkin v. Morinda Props. Weight Parcel, LLC, 440 F. App’x 604, 609 (10th Cir. 2011) (finding that all six factors weighed against finding prejudice, but failing to consider whether the movant’s conduct caused manipulation of the judicial process or inefficient maintenance of disputes), with GVL Pipe & Demolition, Inc. v. Adams Cole & Dalton Rail Serv., LLC, No. CIV-10-846-M, 2010 WL 4806900, at *1, *3 (W.D. Okla. Nov. 18, 2010) (refusing to find waiver when one factor indicated that waiver was appropriate, but conflicting with Hill by first engaging in a mechanical balancing of factors before considering how those factors indicated prejudice, manipulation, and efficiency). There is not a significant amount of case law which addresses waiver by conduct subsequent to Hill. However, the opinions cited above, coupled with the inconsistencies caused by Third Circuit’s unitary approach, strongly indicate that the Tenth Circuit’s multifaceted regime will not yield consistent results.

334. See AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1748 (2011) (“The overarching purpose of the FAA . . . is to ensure the enforcement of arbitration agreements according to their terms so as to facilitate streamlined proceedings.”).

335. See, e.g., Zuckerman Spaeder, LLP v. Auffenberg, 646 F.3d 919, 923 (D.C. Cir. 2011) (holding that a party “presumptively forfeit[s]” its right to arbitrate if it fails to invoke that right at the “earliest opportunity on the record”); Cabinetree of Wis., Inc. v. Kraftmaid Cabinetry, Inc., 50 F.3d 388, 390 (7th Cir. 1995) (finding that merely electing to file a claim in court “is a presumptive waiver of the right to arbitrate”).
Particularly, Judge Posner’s opinion in *Cabinetree of Wisconsin, Inc. v. Kraftmaid Cabinetry, Inc.*[^336] is often cited by the majority circuits as either the counterpoint to a burdensome prejudice requirement[^337] or by those justifying a lenient prejudice standard.[^338]

### 1. The Seventh Circuit

The Seventh Circuit interprets the liberal federal policy favoring arbitration agreements as “merely a policy of treating such clauses no less hospitably” than other contracts.[^339] A party need not show prejudice to establish waiver.[^340] Rather, simply electing to litigate operates as a “presumptive waiver” of the right to arbitrate, which may be rebutted only in “extraordinary circumstances.”[^341] For example, in *Cabinetree*, the court found that a movant waived its right to arbitrate by proceeding judicially rather than demanding arbitration at the outset of the dispute.[^342]

In *Cabinetree*, Cabinetree filed suit against Kraftmaid in state court, alleging breach of contract.[^343] Shortly thereafter, Kraftmaid removed the case to federal court and, six months...
later, moved to compel arbitration.\textsuperscript{344} On appeal, the Seventh Circuit affirmed the district court’s finding of waiver, reasoning that Kraftmaid failed to rebut the presumption that it waived its right to arbitrate.\textsuperscript{345} The court found Kraftmaid’s explanation for its removal and subsequent pretrial activity—that it needed “time to weigh its options”—the “worst possible reason for delay.”\textsuperscript{346} Rather, it found that parties must select a forum in which to resolve their disputes at the earliest opportunity.\textsuperscript{347} Therefore, merely electing to proceed judicially is considered “powerful evidence” that the parties agreed to forego arbitration.\textsuperscript{348} In a subsequent case, the Seventh Circuit found that the movant did not trigger the presumption of waiver when it removed the dispute to federal court six weeks after the complaint was filed; demanded arbitration thirty days after removal; and refrained from engaging in pretrial conduct before demanding arbitration.\textsuperscript{349}

2. The D.C. Circuit

Similarly, the D.C. Circuit recently adopted a bright-line approach to determine whether a party has lost its right to arbitrate through pretrial conduct.\textsuperscript{350} In \textit{Zuckerman Spaeder, LLP v. Auffenberg},\textsuperscript{351} the court found that a movant who fails to

\begin{footnotesize}
\begin{enumerate}
\item See \textit{id.} (noting that six months after removal, Kraftmaid "moved the district court under 9 U.S.C. § 3 to stay further proceedings pending arbitration of the parties' dispute").
\item See \textit{id.} at 391 ("The presumption that an election to proceed judicially constitutes a waiver of the right to arbitrate has not been rebutted.").
\item Id.
\item See \textit{id.} ("Selection of a forum in which to resolve a legal dispute should be made at the earliest possible opportunity . . . .").
\item Id.
\item See Halim v. Great Gatsby's Auction Gallery, Inc., 516 F.3d 557, 562 (7th Cir. 2008) (discussing the movant's conduct and affirming the district court's refusal to find waiver).
\item Zuckerman Spaeder, LLP v. Auffenberg, 646 F.3d 919, 922 (D.C. Cir. 2011) (stating that, in considering whether a movant has lost its right to arbitrate through pretrial conduct, the court has "established a few bright-line rules").
\item See \textit{id.} at 922–23 (adopting a new standard for determining whether a party has lost its right to arbitrate through pretrial conduct).
\end{enumerate}
\end{footnotesize}
invoke arbitration “on the record at the first available opportunity... presumptively forfeit[s]” its right to arbitrate.\(^{352}\) Thus, the court replaced the waiver doctrine with forfeiture, reasoning that intent is irrelevant for determining whether a party has lost its right to arbitrate through pretrial conduct.\(^{353}\) Rather, simply failing to demand arbitration in a timely manner will trigger forfeiture.\(^{354}\) To be timely, the movant’s demand must come in its “first responsive pleading or motion to dismiss.”\(^{355}\) If the movant fails this requirement, it may still access the arbitral forum if it proves that its delay did not prejudice the opponent or the court.\(^{356}\) Only a minimal amount of undue expense or delay need be shown to establish prejudice.\(^{357}\) Applying this standard, the \textit{Auffenberg} court found that Auffenberg forfeited his right to arbitrate.\(^{358}\)

In \textit{Auffenberg}, Zuckerman filed claims against Auffenberg in the D.C. Superior Court for unpaid attorneys’ fees.\(^{359}\) Auffenberg subsequently removed the case to federal court, answered Zuckerman’s complaint, and filed counterclaims.\(^{360}\) After several months of engaging in pretrial motions and court-ordered mediation, Auffenberg moved to compel arbitration.\(^{361}\) The D.C.

\(^{352}\) \textit{Id.} at 922.

\(^{353}\) See \textit{id.} (abandoning waiver, which “refers to a party’s intentional relinquishment... of a known right,” and adopting forfeiture, which simply refers to a “failure to make a timely assertion of a right” (internal quotation marks and citations omitted)).

\(^{354}\) See \textit{id.} (finding that “forfeiture, not waiver, is the appropriate standard”).

\(^{355}\) \textit{Id.}

\(^{356}\) See \textit{id.} at 923 (“A defendant who delays seeking a stay pending arbitration until after his first available opportunity might still prevail on a later stay motion provided his delay did not prejudice his opponent or the court.”).

\(^{357}\) See \textit{id.} (allowing access to arbitration only if a movant’s delay inflicts “no or little cost upon opposing counsel and the courts”).

\(^{358}\) See \textit{id.} (affirming the district court and finding that Auffenberg forfeited its right to arbitrate).

\(^{359}\) See \textit{id.} at 920 (noting that “Zuckerman sued Auffenberg in the District of Columbia Superior Court to recover the [attorneys’] fees plus interest”).

\(^{360}\) See \textit{id.} (“Auffenberg removed the case to federal court, answered the complaint, and counterclaimed for legal malpractice.”).

\(^{361}\) See \textit{id.} at 921 (noting that “a client may invoke mandatory arbitration of any fee dispute” with its lawyer under “D.C. Bar Rule XIII”).
Circuit affirmed the district court’s refusal to grant Auffenberg’s motion on two key grounds. First, it found Auffenberg’s failure to demand arbitration in his original answer sufficient to invoke the presumption of forfeiture. Second, the court found that Auffenberg’s pretrial conduct “imposed substantial costs upon both Zuckerman and the district court.” Particularly, Zuckerman was forced to conduct internal investigations, engage in discovery, and prepare depositions. Auffenberg’s conduct also consumed “inherently limited” judicial resources, including the court’s time. Ultimately, because the court found Auffenberg’s conduct prejudicial, it concluded that he failed to rebut the presumption of forfeiture.

3. The Seventh and D.C. Circuits Compared

Notably, both the Seventh and the D.C. Circuits have broken with the majority to facilitate doctrinal accuracy. The Seventh Circuit refuses to require prejudice on contractual grounds, noting that prejudice is not required to find waiver under contract law. Similarly, the D.C. Circuit repudiated the waiver doctrine in favor of forfeiture in order to “realign litigants’ incentives . . . with the FAA.” Finally, both circuits seek to prevent dilatory, manipulative strategies that impose unnecessary costs on both nonmovants and the judicial system.
Yet, although their analyses have moved toward alignment with the FAA, the efficacy of each circuit’s presumptive regime is questionable. Particularly, the Seventh Circuit has failed to uphold its presumption consistently since Cabinetree, allowing movants to affect rebuttal in divergent and even manipulative factual circumstances. This Note, however, proposes a uniform framework which would instill consistency and prevent unscrupulous pretrial conduct.

IV. “Waiving” the Right to Arbitrate: A Contractual Approach

A. The Need for a Uniform Standard

The divergent, ill-defined waiver standards propagated by the Circuit Courts of Appeals present three significant problems. First, their juridical inconsistencies deny contracting parties the

both litigants and the district court[s]); Cabinetree, 50 F.3d at 391 (introducing a presumption of waiver in order to “economize on the resources, both public and private, consumed in dispute resolution” by stifling dilatory “heads I win, tails you lose” strategies).

371. Compare Sharif v. Wellness Int’l Network, Ltd., 376 F.3d 720, 722–23 (7th Cir. 2004) (refusing to find waiver when the movant delayed for nine months before moving to compel arbitration, and filed motions to dismiss on the merits and for improper venue, which were “briefed fully” in the district court), and Benjamin-Coleman v. Praxair, Inc., 216 F. Supp. 2d 750, 752, 754 (N.D. Ill. 2002) (refusing to find waiver where the movant delayed for four months, during which time it filed three merits-based motions to dismiss), with Grumhaus v. Comerica Sec., Inc., 223 F.3d 648, 651 (7th Cir. 2000) (finding waiver after a movant delayed demanding arbitration for one year after its complaint was filed and, subsequently, dismissed), and Pa. Chiropractic Ass’n v. Blue Cross Blue Shield Ass’n, No. 09 C 5619, 2011 WL 210805, at *1, *1 (N.D. Ill. Jan. 20, 2011) (finding waiver where the movants delayed for eight months and filed three motions to dismiss).

372. See N. Cent. Constr., Inc. v. Siouxland Energy & Livestock Coop., 232 F. Supp. 2d 959, 968 (N.D. Iowa 2002) (refusing to find waiver, yet acknowledging that the movant may have engaged in a “limited form of forum shopping” by delaying its demand for arbitration); Benjamin-Coleman, 216 F. Supp. 2d at 753 (refusing to find waiver although the movant engaged in a “limited form of forum shopping, which the waiver doctrine is designed to prevent”).

373. See AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1748 (2011) (“The overarching purpose of the FAA . . . is to ensure the enforcement of arbitration agreements according to their terms so as to facilitate streamlined proceedings.”).
predictability that they expect through bargained-for arbitral procedures, thus destabilizing the arbitral process. For example, parties subjected to onerous prejudice requirements often spend months consuming judicial and personal resources pretrial, only to be forced to arbitrate according to procedures that have lost their initial value. The minority's presumptive regime proves similarly unavailing, as it merely shifts the unpredictability post hoc towards rebuttal.

Second, the waiver doctrine contravenes the FAA's overarching objective: to "ensure the enforcement of arbitration agreements according to their terms so as to facilitate streamlined proceedings." Indeed, the two goals inherent within this objective—contractual enforcement and facilitation of streamlined proceedings—should coincide. As to the former, the FAA reflects the fundamental notion that arbitration agreements, as contracts, should be enforced in accordance with parties' legitimate expectations. The circuits, however, often subordinate these expectations in the name of the oft-cited federal policy favoring arbitration agreements. Unfortunately, this reasoning misconstrues the policy's true objective—to ensure the "enforcement of private contractual arrangements" according

374. See supra Part III (discussing the inconsistencies among the circuits regarding proper waiver standard).
375. See supra Part III (describing a number of circuits which impose burdensome prejudice requirements in their waiver analyses).
376. See Zuckerman Spaeder, LLP v. Auffenberg, 646 F.3d 919, 922 (D.C. Cir. 2011) (describing the costs imposed on both litigants and the courts in the absence of a clear standard); Rankin v. Allstate Ins. Co., 336 F.3d 8, 13 (1st Cir. 2003) (noting the importance of choosing between arbitration and litigation early in the dispute resolution process to avoid the needless deployment of both individual and judicial resources).
377. See supra note 371 and accompanying text (describing the doctrinal inconsistencies permeating the Seventh Circuit since Cabinetree).
379. See id. at 1749 (alluding to the notion that, where possible, the FAA's underlying goals should not conflict).
380. See supra note 72 and accompanying text (discussing the FAA's contractual goal of enforcing arbitration agreements in accordance with parties' legitimate expectations).
381. See Moses H. Cone Mem'l Hosp. v. Mercury Constr. Corp., 460 U.S. 1, 24 (1983); supra Part III (discussing numerous cases in which courts justify their respective waiver analyses by citing the liberal federal policy favoring arbitration agreements).
to their terms.382 Indeed, it merely reflects FAA § 2’s substantive mandate,383 which “requires courts to honor parties’ expectations.”384 Thus, the ill-defined waiver and rebuttal standards imposed by the circuits clearly fail this requirement because both undercut the key benefits that parties expect to derive through bargained-for arbitral procedures—predictability and stability.385

Further, requiring parties to prove either prejudice or rebuttal inhibits procedural efficiency—the FAA’s second underlying goal. Indeed, the Supreme Court has eschewed burdensome procedural obstacles that impede a clear path to the arbitral forum. For example, the Court has refused to require parties to obtain an order compelling arbitration in federal court after securing a stay of judicial proceedings in state court.386 Similarly, the Court has invalidated state laws requiring administrative exhaustion before granting access to arbitration.387 To be sure, strong prejudice requirements ensure that parties’ arbitral rights will be enforced much deeper into the pretrial process. Long delays, however, contravene “Congress’s intent to move the parties . . . out of court and into arbitration as quickly and easily as possible.”388 Indeed, requiring endless


383. See Moses H. Cone, 460 U.S. at 24 (stating that FAA § 2 is “a congressional declaration of a liberal federal policy favoring arbitration agreements”).


385. See supra Part II (discussing the benefits derived through arbitration).

386. See Moses H. Cone, 460 U.S. at 27 (noting that a movant who first obtained a stay in state court “would have no sure way to proceed with its claims [in arbitration] except to return to federal court to obtain a § 4 order—a pointless and wasteful burden on the supposedly summary and speedy procedures prescribed by the Arbitration Act”).

387. See Preston v. Ferrer, 552 U.S. 346, 347 (2008) (finding that state-mandated administrative exhaustion conflicts with the FAA’s mandate to facilitate expedient access to the arbitral forum); Gilmer v. Interstate/Johnson Lane Corp., 500 U.S. 20, 28–29 (1991) (“The mere involvement of an administrative agency in the enforcement of a statute is not sufficient to preclude arbitration.”).

388. Preston, 552 U.S. at 327 (internal quotation marks and citations
litigation over the details of prejudice, or rebuttal flouts the judiciary’s responsibility to facilitate expedient access to the arbitral forum.

Finally, parties face the distinct possibility that their arbitral rights will be turned against them to inflict undue delay and expense. Indeed, arbitration agreements designed to protect parties’ interests may turn from a shield to a sword in the hands of recalcitrant parties whose contractual relationships have gone awry. Such conduct inflicts unnecessary costs on both parties and the judicial system. Furthermore, it strips the arbitral process of any predictability, destabilizing parties’ contractually established dispute resolution mechanisms. The circuits

omitted).

389. See Erdman Co. v. Phoenix Land & Acquisition, LLC, 650 F.3d 1115, 1120 (8th Cir. 2011) (finding that requiring parties “to litigate endlessly over the details of prejudice” makes little sense in certain circumstances).


392. See, e.g., Leadertex, Inc. v. Morganton Dyeing & Finishing Corp., 67 F.3d 20, 26 (2d Cir. 1995) (noting that no matter how “unjustifiable” a movant’s conduct, it would not find waiver without first finding prejudice); Menorah Ins. Co., Ltd. v. INX Reinsurance Corp., 72 F.3d 218, 222–23 (1st Cir. 1995) (noting the tendency of parties to use arbitration clauses to manipulate the dispute resolution process); Walker v. J.C. Bradford & Co., 938 F.2d 575, 577 (5th Cir. 1991) (recognizing the movant’s delay and attempts to forum shop, but refusing to find waiver).

393. See, e.g., Cabinetree of Wis., Inc. v. Kraftmaid Cabinetry, Inc., 50 F.3d 388, 391 (7th Cir. 1995) (refusing to find waiver where a party sought to “play heads I win, tails you lose” by seeing “how the case was going in federal district court before deciding whether it would be better off there or in arbitration”).

394. See id. (“Selection of a forum in which to resolve a legal dispute should be made at the earliest possible opportunity in order to economize on the resources, both public and private, consumed in dispute resolution.”); see also Rankin v. Allstate Ins. Co., 336 F.3d 8, 13 (1st Cir. 2003) (requiring that arbitration be “invoked at the earliest opportunity” so that courts’ and parties’ “resources are not needlessly deployed”).

395. See supra Part II (describing the benefits parties seek through arbitration).
reluctantly condone such conduct, citing the liberal federal policy favoring arbitration as if their hands are tied.396 This Note proposes, however, that they are not so bound. A contractual approach would prevent unscrupulous strategies employed by obstinate parties in accordance with the FAA’s text and overarching objective.397 Before discussing this approach, however, it is important to note the fundamental doctrinal inaccuracies of applying the waiver doctrine, as prescribed by contract law, in the context of parties’ arbitral rights.

B. “Waiver” as a Generally Applicable Contract Law Defense

Recall that waiver by conduct implicates the obligation to arbitrate.398 Therefore, it is a gateway issue that must be resolved by the courts in accordance with contract law.399 In this regard, the waiver doctrine’s legitimacy should be tested against the Restatement (Second) of Contracts. Applying the Restatement would comply with the FAA,400 as it sets the general conceptual

396. See, e.g., MicroStrategy, Inc. v. Lauricia, 268 F.3d 244, 254 (4th Cir. 2001) (refusing to find waiver under the liberal federal policy favoring arbitration, but recognizing that the movant’s “aggressive” course of litigation was seemingly taken to wear the nonmovant out “both emotionally and financially”); Rush v. Oppenheimer & Co., 779 F.2d 885, 889–91 (2d Cir. 1985) (citing the liberal federal policy favoring arbitration agreements in refusing to find waiver where a movant demanded arbitration only after it lost a merits-based motion to dismiss in court); North Cent. Constr., Inc. v. Siouxland Energy & Livestock Coop., 232 F. Supp. 2d 959, 968 (N.D. Iowa 2002) (citing the liberal federal policy favoring arbitration agreements and refusing to find waiver, even after acknowledging that the movant’s conduct constituted “a limited form of forum shopping”); Benjamin-Coleman v. Praxair, Inc., 216 F. Supp. 2d 750, 753 (N.D. Ill. 2002) (refusing to find waiver due to the liberal federal policy favoring arbitration agreements, even after acknowledging that the movant engaged in forum shopping as exhibited by its pretrial activity).

397. See AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1748 (2011) (“The overarching purpose of the FAA . . . is to ensure the enforcement of arbitration agreements according to their terms so as to facilitate streamlined proceedings.”).

398. See supra Part II (discussing the connection between conduct-based waiver, severability, and contract law).

399. See supra Part II (discussing how gateway issues must be decided by the courts in accordance with contract law).

400. See, e.g., Doctor’s Assocs., Inc. v. Casarotto, 517 U.S. 681, 687 (1996) (stating that, under FAA § 2, arbitration agreements may by validated only by “generally applicable contract defenses”).
foundation upon which states have constructed their contract law.401

Under the Restatement, waiver occurs when a party promises to perform a conditional duty under an antecedent contract, despite the condition’s nonoccurrence.402 A condition is “an event, not certain to occur, which must occur . . . before performance [of one’s duty] under a contract becomes due.”403 Thus, a party that waives a condition precedent to its duty to perform a contractual obligation must still perform that duty, even if the condition has not occurred.

A critical corollary to the waiver doctrine, however, is that a condition’s occurrence must not be “a material part of the agreed exchange for the performance of the duty.”404 Thus, the waiver doctrine presumes that a condition is being waived, not a contract, and that the condition was not a material part of the underlying agreement. For example, if Promisor contracts to sell her car to Promisee in exchange for $500, Promisor need not tender the car until Promisee tenders the money.405 Promisor cannot waive the purchase price, however, because she cannot waive a condition that materially affects the value to be received under the contract—payment in exchange for her car.406

401. See E. Allan Farnsworth, Ingredients in the Redaction of the Restatement (Second) of Contracts, 81 COLUM. L. REV. 1, 5 (1981) (noting that the drafters of the Restatement felt “obliged in [their] deliberations to give weight to all of the considerations that the courts, under a proper view of the judicial function, deem it right to weigh in theirs” (internal quotation marks and citations omitted)); Herbert Wechsler, The Course of Restatements, 55 A.B.A. J. 147, 147 (1969) (stating that restatements generally have “enormous influence on the development of our law”).

402. See Restatement (Second) of Contracts § 84(1) (1981) (“[A] promise to perform all or part of a conditional duty under an antecedent contract in spite of the non-occurrence of the condition is binding, whether the promise is made before or after the time for the condition to occur . . . .”).

403. Id. § 224.

404. Id. § 84(1)(a).

405. See id. § 84 cmt. c (noting that in a contractual arrangement to “sell a horse for $500,” a “waiver of the price of a horse is not within this Section” and, thus, is not waiver).

406. See id. (“[W]here a promise to disregard the nonoccurrence of the condition materially affects the value received by the promisor . . . the promise [to waive the condition] is not binding . . . .”).
Drawing from the Restatement, it is clear that the circuits’ analyses fundamentally conflict with contract law and, as such, the FAA. First, under the severability doctrine, an arbitration clause is neither immaterial, nor a condition precedent to performance of the underlying contract. Rather, when pretrial conduct implicates the obligation to arbitrate, the severability doctrine separates the arbitration clause from the underlying agreement. This severance results in a wholly autonomous arbitration clause, fully enforceable in accordance with generally applicable contract law principles. It therefore becomes clear, under the Restatement, that a party can neither waive an independent arbitration agreement nor the sole material benefit derived from that agreement—access to the arbitral forum. Furthermore, neither prejudice nor rebuttal is necessary to establish waiver under contract law.

Ultimately, waiver cannot be the proper standard for determining whether a party has lost its right to arbitrate through pretrial conduct. The Supreme Court has interpreted FAA § 3, from which the waiver doctrine was derived, as purely procedural. It does not alter “background principles of state contract law” in enforcing obligations to arbitrate under § 2. Thus, as both the majority and the minority’s standards alter the waiver doctrine as prescribed by contract law, they clearly contravene the FAA. In this regard, the D.C. Circuit’s forfeiture standard proves equally unavailing. The Restatement limits forfeiture to circumstances in which the “occurrence of the

407. See supra note 161 and accompanying text (discussing how issues concerning waiver, as a gateway issue, must be considered under the auspices of contract law).

408. See supra Part II (discussing the severability doctrine).

409. See supra Part II (describing how FAA § 2 requires the enforcement of arbitration agreements as contracts).

410. See supra notes 404–06 and accompanying text (discussing how, under contract law, a party cannot waive a contract or the material value for which it bargained).

411. See Restatement (Second) of Contracts § 84 cmt. b (1981) (stating when “waiver is reinforced by [detrimental] reliance, enforcement is often said to rest on estoppel,” not waiver).

412. See supra note 153 and accompanying text (discussing the origins of the waiver doctrine).

condition] not a \textit{material} part of the agreed exchange.\textsuperscript{414}

Therefore, forfeiture falls prey to the same conceptual barriers as waiver. Indeed, an entirely new contractual framework must be devised in order to realign the circuits’ analyses with one another, and with the FAA.

\section*{V. The Proposal}

\subsection*{A. A Contractual Approach}

This Note proposes a comprehensive contractual solution through a succinct judicial framework to discern whether a party has lost its right to arbitrate through pretrial conduct. The reasonableness test comprises the core of this solution and requires courts to discern what constitutes a reasonable time to demand arbitration.\textsuperscript{415} First, this Note identifies the key contractual components which drive the operative function of arbitration agreements and, through those components, sets the proper standard. Next, this Note proposes a multifactor reasonableness test to determine whether a party demanded arbitration within a reasonable time after engaging in pretrial conduct. Finally, this Note employs a useful hypothetical, based on the Fourth Circuit’s decision in \textit{MicroStrategy, Inc. v. Lauricia}, to illustrate the efficacy of this approach.\textsuperscript{416}

\subsection*{1. The Groundwork: A Duty to Arbitrate Subject to Conditions}

The reasonableness test begins with the premise that an arbitration clause is an autonomous contract, the performance of which hinges upon the occurrence of “constructive (or ‘implied in law’)” conditions.\textsuperscript{417} The FAA does not impose an affirmative duty

\footnotesize
\textsuperscript{414} \textsc{Restatement (Second) of Contracts} § 229 cmt. c (1981) (emphasis added).


\textsuperscript{416} See \textit{supra} Part II (discussing the Fourth Circuit’s decision in \textit{MicroStrategy}, which was decided under a burdensome prejudice requirement).

\textsuperscript{417} \textsc{Restatement (Second) of Contracts} § 226 cmt. c (1981).
to demand arbitration under a private agreement once a dispute arises. Rather, given parties’ ability to stay judicial proceedings under § 3, the FAA provides parties the opportunity to resolve their disputes through different mediums before resorting to arbitration (i.e., mediation, negotiation, or litigation). Therefore, each party’s respective duty to arbitrate is conditioned upon the other’s demand. If neither party invokes its right to arbitrate, then neither party’s duty to arbitrate becomes due. In contract law, this nonoccurrence eventually discharges each party’s duty to perform its contractual obligation, i.e., to arbitrate, “when the condition can no longer occur.” Thus, the discharge principle provides a conceptually accurate contract law defense for parties attempting to avoid arbitration after litigation has commenced.

Providing a defense, however, does not end the inquiry. The presence of FAA § 3 and the pervasiveness of the liberal federal policy favoring arbitration beg the question: At what point in the pretrial process does a party’s duty to arbitrate become discharged? Under contract law, it is first essential to consider the language of the parties’ arbitration agreement. If the parties specified time limits on their right to demand arbitration, that should certainly end the inquiry. The Supreme Court, however, has recently found that procedural issues, such as time

418. See 9 U.S.C. § 3 (2011) (providing parties to an otherwise enforceable arbitration agreement the opportunity to stay judicial proceedings pending arbitration).


420. See RESTATEMENT (SECOND) OF CONTRACTS § 225(1) (1981) (“Performance of a duty subject to a condition cannot become due unless the condition occurs . . . .”).

421. Id. § 225(2).

limits, are presumptively for arbitrators to decide.\textsuperscript{423} By contrast, disputes regarding “whether the parties are bound by a given arbitration clause” are for the courts to decide.\textsuperscript{424} The circuits have largely interpreted this decision as leaving conduct-based waiver within the courts’ purview.\textsuperscript{425}

Consequently, however, judges will not have recourse to concrete time limits in parties’ arbitration agreements for determining discharge. Furthermore, it is highly unlikely that parties would bargain for procedural contingencies in case of litigation while attempting to establish proper arbitral procedures. Such uncertainty presents a significant quandary for courts because one party must demand arbitration at some point in order to trigger its opponent’s duty to arbitrate. The Restatement, however, provides that in the absence of a “fixed term” in the parties’ agreement setting “[t]he time within which the condition can occur,”\textsuperscript{426} a term “which is reasonable in the circumstances is supplied by the court.”\textsuperscript{427} Thus, to facilitate the arbitral process, “a term calling for performance within a reasonable time” should be “supplied.”\textsuperscript{428}

2. The Reasonableness Test: What Constitutes a “Reasonable Time” for Performance?

An inherently fact-based analysis, standards for determining what constitutes a reasonable time for performance vary not only from state to state, but also case by case.\textsuperscript{429} New York courts,

\textsuperscript{423} See Howsam v. Dean Witter Reynolds, Inc., 537 U.S. 79, 85 (2002) (“We find that the applicability of the . . . time limit rule is a matter presumptively for the arbitrator, not the judge.”).
\textsuperscript{424} Id. at 84.
\textsuperscript{425} See supra note 156 and accompanying text (citing and describing the numerous circuit court decisions which hold waiver by conduct an issue for the courts to decide).
\textsuperscript{426} Restatement (Second) of Contracts § 225 cmt. a (1981).
\textsuperscript{427} Id. § 204.
\textsuperscript{428} Id. § 204 cmt. d.
\textsuperscript{429} See Zev v. Merman, 533 N.E.2d 669, 669 (N.Y. 1988) (“What constitutes a reasonable time for performance depends upon the facts and circumstances of the particular case.”); see also 1700 Rhinehart LLC v. Advance Am., 51 So.3d 535, 540 (Fla. Dist. Ct. App. 2010) (“[W]hen a contract fails to specify a particular period, the law implies a reasonable time under the circumstances.”);
however, have compiled a comprehensive list of factors in order to provide a uniform, predictable approach to this inquiry.\textsuperscript{430} Particularly, the courts consider “the nature and object of the contract, the previous conduct of the parties, the presence or absence of good faith,” and the “possibility of prejudice or hardship” to either party.\textsuperscript{431}

First, the courts contemplate the nature and the object of the contract in question by looking to its operative purpose, including “all of the rules and procedures thereunder.”\textsuperscript{432} For example, the delayed provision of “ministerial” services which bear little significance on a party’s operations\textsuperscript{433} will be found much more reasonable than delayed payments of property taxes required under an option contract to purchase real property.\textsuperscript{434} Second, courts will look to the parties’ prior conduct in order to determine whether they were fulfilling their “contractual duties as intended.”\textsuperscript{435} For example, one court found a two-week closing period, set abruptly by a seller of real property after it delayed closing for five years, unreasonable.\textsuperscript{436} Critical to its finding was

\begin{footnotesize}
\begin{enumerate}
\item \textit{See} Zev, 533 N.E.2d at 669.
\item \textit{Id.}
\item \textit{Smith Barney}, 866 F. Supp. at 117.
\item \textit{See} Parker v. Booker, 822 N.Y.S.2d 156, 158 (N.Y. App. Div. 2006) (“Since real property taxes by their nature are due on a particular day, the reasonable date on which they were required to be paid is the date on which they were due.”).
\item \textit{Smith Barney}, 866 F. Supp. at 118.
\item \textit{See} Knight v. McClean, 566 N.Y.S.2d 952, 954 (N.Y. App. Div. 1991) (describing the seller’s conduct which made the time that it set for performance unreasonable).
\end{enumerate}
\end{footnotesize}
that the seller’s actions exhibited nothing more than a “desire to avoid” performance of the contract as the parties intended.437

Third, courts consider the extent to which each party performed its contractual obligations in good faith; that is, whether each is acting in “consideration” of the other’s interests.438 Indeed, a “covenant of good faith and fair dealing is implied in all contracts,”439 and its meaning varies with the circumstances.440 Generally, however, “good faith performance . . . emphasizes faithfulness to an agreed common purpose and consistency with the justified expectations of the other party.”441 Although particular acts of bad faith vary, two notable instances are evading the “spirit” of the bargain and willfully rendering “imperfect performance.”442 In the arbitral context, for example, one court ordered a labor union to arbitrate with an employer pursuant to an arbitration clause in the parties’ collective bargaining agreement.443 When the union attempted to contest the tribunal’s jurisdiction, the court stated that parties to an arbitration agreement “must live up to the spirit of their agreement.”444 They “will not be permitted to take refuge in subtle and adroit evasions in order to defeat the purposes of the agreement.”445

Finally, the reasonableness inquiry will be affected by the extent to which the delay in question prejudiced each party.446 For example, when a finding of unreasonable delay would cause

437. Id.
440. See RESTATEMENT (SECOND) OF CONTRACTS § 205 cmt. a (1981) (“The phrase ‘good faith’ is used in a variety of contexts, and its meaning varies somewhat with the context.”).
441. Id.
442. Id. § 205 cmt. d.
443. Publishers’ Ass’n of N.Y. City v. N.Y. Newspaper Printing Pressmen’s Union No. Two, 145 N.Y.S.2d 93, 96 (N.Y. Sup. Ct. 1955) (noting that the union contested the arbitrators’ jurisdiction “because the [employee’s] grievance . . . was ambiguous”).
444. Id. at 96.
445. Id. at 97.
one party to suffer substantial losses while benefitting the other, a court will be much less likely to find the delay unreasonable.447

Ultimately, New York’s reasonableness analysis provides a stable, yet flexible framework to determine whether a party has demanded arbitration within a reasonable time. Not all of the factors need weigh in favor of unreasonableness to warrant such a finding, and a strong implication of one factor often implicates another.448 Furthermore, the test’s flexibility renders it applicable to a myriad of contractual arrangements and circumstances.449 Therefore, this Note proposes that federal courts abandon the waiver–forfeiture doctrine, and incorporate instead, through FAA § 2, New York courts’ multifactor approach as a reasonableness test.450 Doing so will establish a uniform, predictable, doctrinally accurate method to determine whether a party has lost its right to arbitrate through pretrial conduct.

447. See Miller v. Almquist, 671 N.Y.S.2d 746, 750 (N.Y. App. Div. 1998) (refusing to find unreasonable delay where the buyers would have lost “their opportunity to purchase” an apartment and their “$54,000 deposit,” while the sellers would “have received the all-cash deal they had bargained for”).

448. See, e.g., Malley v. Malley, 861 N.Y.S.2d 149, 152 (N.Y. App. Div. 2008) (finding that the defendant was given an unreasonable time in which to perform where a finding of reasonableness would have left the plaintiff’s “financial position . . . the same,” but prejudiced the defendant severely); Parker v. Booker, 822 N.Y.S.2d 156, 158 (N.Y. App. Div. 2006) (finding unreasonable delay but no bad faith); Knight v. McLean, 566 N.Y.S.2d 852, 854 (N.Y. App. Div. 1991) (finding that buyers were presented with an unreasonable time to perform where “the previous conduct of the parties” was “[o]f significant importance”).


3. A “Reasonable Time” to Demand Arbitration: MicroStrategy, Inc. v. Lauricia

The Fourth Circuit’s decision in MicroStrategy, Inc. v. Lauricia presents a useful paradigm to illustrate the efficacy of the reasonableness test. Indeed, the impact of this approach is best exemplified when compared against the strong prejudice requirements propounded by the majority. Recall that in MicroStrategy, Lauricia was terminated from her position as the head of MicroStrategy’s Human Resources Department. Within six months of her firing, MicroStrategy filed three separate claims against Lauricia before demanding arbitration—one in state court and two in federal court. Finally, although we are told only that the parties “agreed” to the arbitral procedures in question, let us assume that Lauricia, as a high level employee, negotiated these procedures. At trial, the district court held that MicroStrategy waived its right to arbitrate, finding that its “remarkably aggressive” pretrial conduct and extensive discovery prejudiced Lauricia.

On appeal, let us assume the Fourth Circuit has replaced the waiver doctrine with the reasonableness test. Thus, the question before the court is whether MicroStrategy’s six-month delay in demanding arbitration was unreasonable. If so, the court will consider Lauricia’s duty to arbitrate discharged and continue litigation; if not, then the court will stay further proceedings and compel arbitration.

First, the relevant period in which to consider delay should begin when a party files a claim in court because any delay should coincide with pretrial conduct. Once that point is identified, the court should begin its analysis by considering the

---

452. See id. at 247–50 (discussing the claims and the time frame within which MicroStrategy filed its claims against Lauricia).
453. Id. at 251.
455. See supra Part II (discussing how conduct-based waiver has been the focus of the circuits and the issue upon which the Supreme Court granted certiorari).
nature and object of the parties’ arbitration agreement. It should do so, however, in light of the discrete benefits derived through arbitration, and the inherent destabilizing effect of judicial involvement.\textsuperscript{456} Further, the court should consider the nature of the arbitral procedures for which the parties contracted. For example, if the parties’ arbitration agreement provides for expedited procedures, whether ad hoc or institutional,\textsuperscript{457} the court should be far less willing to tolerate delay. In this hypothetical, the arbitration clause requires only that the parties arbitrate “any controversy or claim arising out of . . . th[e] Employee Handbook.”\textsuperscript{458} Thus, although expedited arbitral review is not in question, the inherent destabilization caused by the judicial forum is important to keep in mind moving forward in the analysis.

Next, the court should consider whether the parties’ prior conduct manifests a desire to carry out the arbitration agreement as intended. As the period for consideration begins when one party files a claim, the court should consider which party first filed the claim, the extent of the parties’ litigation activity, and the nature of the parties’ relationship (i.e., whether they were attempting to ameliorate their differences throughout the pretrial process). First, it is particularly significant that MicroStrategy filed three separate complaints after it learned of Lauricia’s employment discrimination charges.\textsuperscript{459} Such conduct shows a clear intent to avoid the arbitral forum, or perhaps that more suspect motives are in play. Second, MicroStrategy engaged in “remarkably aggressive” litigation conduct, as it deposed Lauricia, seized a number of documents, obtained numerous responses to interrogatories, and obtained Lauricia's personal information, all before demanding arbitration.\textsuperscript{460} Third, after MicroStrategy filed

\textsuperscript{456} See supra Part II (describing the benefits of arbitration).

\textsuperscript{457} For a useful example of expedited arbitral procedures, see AAA RULES, supra note 58, at 42.

\textsuperscript{458} MicroStrategy, Inc. v. Lauricia, 268 F.3d 244, 246 (4th Cir. 2001) (internal quotation marks omitted).

\textsuperscript{459} See id. at 246–48 (describing the complaints filed by MicroStrategy).

\textsuperscript{460} See id. at 254 (noting that “MicroStrategy deposed Lauricia, successfully sought the seizure of documents, . . . received responses from Lauricia to interrogatories and requests to produce, and obtained Lauricia’s employment records from her former employers”).
its first two claims, the Equal Employment Opportunity Commission invited the parties to participate in a “conciliation process.”461 Lauricia, however, declined the invitation and requested that she be issued her “right-to-sue letter” immediately.462 Thus, it is apparent that the parties’ relationship had completely deteriorated by the time MicroStrategy demanded arbitration. Ultimately, the above considerations show that MicroStrategy had no desire to exercise the arbitration agreement as intended—that is, to curb the costs and delays appurtenant to litigation through mutually beneficial arbitral procedures.463 Rather, it sought to exhaust Lauricia’s emotional and economic reserves by manipulating the arbitral process.464 As it is unlikely that Lauricia intended, or expected, this result, MicroStrategy’s conduct drives the analysis toward unreasonableness.

Third, the court should consider the extent to which Lauricia would suffer prejudice if MicroStrategy’s delay were found reasonable. Conceptually, it is appropriate to maintain the circuits’ notions of prejudice (i.e., costs, delay, and damage to a party’s legal position through pretrial activity), as the other factors in the reasonableness test inform these considerations.465 Critically, however, the court should consider expense and delay alone, the very harms sought to be avoided through arbitration,466 sufficient to warrant a finding of prejudice.467 Therefore, it is much more likely that Lauricia would be found prejudiced under the reasonableness test. Although MicroStrategy’s six-month delay was not particularly egregious, Lauricia was forced to engage in substantial pretrial activity at great personal

461. Id. at 247.
462. Id.
463. See supra Part II (discussing the benefits derived, and expected, through bargained-for arbitral procedures).
465. See supra Part III (discussing the various considerations encompassed within each circuit’s prejudice analysis).
466. See supra Part II (discussing the benefits of arbitration).
467. See supra notes 279–80 and accompanying text (discussing the First Circuit’s rationale for finding undue delay and costs, in and of themselves, sufficient to warrant a finding of prejudice).
expense.\textsuperscript{468} Despite the presence of prejudice under the reasonableness test, however, let us assume that the court refused to find prejudice for illustrative purposes.

The final step in the reasonableness test is to consider whether a movant is attempting to exercise the arbitration agreement in bad faith. In doing so, the court should keep in mind that FAA § 3 allows recourse to the judicial forum.\textsuperscript{469} As such, merely filing a claim should not trigger a finding of bad faith. Rather, the court should deny only those demands which contradict the “spirit,” or “defeat the purposes,” of the parties’ arbitration agreement.\textsuperscript{470} Other factors, such as the parties’ prior conduct and the degree of prejudice suffered by the nonmovant, should be informative. In this hypothetical, MicroStrategy’s “remarkably aggressive” pretrial conduct is strongly indicative of bad faith.\textsuperscript{471} It moved to compel arbitration only after inflicting significant costs.\textsuperscript{472} Further, its aggressive use of pretrial discovery evinced a willful desire to evade the arbitral framework in favor of one more propitious to its cause.\textsuperscript{473} Therefore, the court should find that MicroStrategy willfully evaded the spirit of the parties’ agreement and, thus, exercised its arbitral rights in bad faith. Indeed, such a finding should be strongly indicative, if not dispositive, of unreasonable delay.\textsuperscript{474}

\textsuperscript{468} See Lauricia, 268 F.3d at 250–51 (finding “no doubt” that MicroStrategy’s pretrial conduct “involved the expenditure of substantial sums of money by all involved”).

\textsuperscript{469} See 9 U.S.C. § 3 (2011) (permitting parties to stay judicial proceedings pending arbitration, implying that some recourse to a judicial forum before exercising arbitral rights is permissible).


\textsuperscript{471} MicroStrategy, Inc. v. Lauricia, 268 F.3d 244, 248 (4th Cir. 2001) (internal quotation marks and citations omitted); see \textit{Restatement (Second) of Contracts} § 205 cmt. d (1981) (citing “evasion of the spirit of the bargain” and “willful rendering of imperfect performance” as particular examples of bad faith).

\textsuperscript{472} See Lauricia, 268 F.3d at 250–51 (finding that Lauricia was forced to pay “substantial sums of money” engaging in pretrial conduct).

\textsuperscript{473} See id. at 250, 254 (noting that, “[u]nder the rules by which the parties agreed to arbitrate,” discovery was available but “under standards different from those governing discovery in federal court,” particularly because the arbitrator had more discretion to determine what materials were discoverable).

Overall, although there was neither any prejudice nor significant temporal delay, the court should find MicroStrategy’s delayed performance unreasonable. Focusing squarely on the parties’ pretrial conduct, it was clear that MicroStrategy did not attempt to exercise its arbitral rights as intended. Further, its conduct rose to the level of bad faith performance, as it willfully turned cost-effective, stabilizing procedures into weapons of financial malaise and delay. Therefore, the court should hold Lauricia’s duty to arbitrate discharged under the reasonableness test.

4. The Efficacy of the “Reasonableness Test” as Compared to Waiver

The above paradigm illustrates how the reasonableness test would improve upon the waiver doctrine in three key respects. First, the reasonableness test would unify and standardize the method for determining at what point a party, through pretrial conduct, has lost its right to arbitrate. The test is fundamentally aimed toward protecting parties’ legitimate, contractual expectations. This unitary focus, coupled with the severability doctrine’s magnifying effect, would center the analysis on preserving the discrete benefits derived through arbitral procedures—stability and predictability. The resulting approach thus provides an analytic framework capable of uniform federal application to which parties can look for

475. See supra Part II (describing how arbitration benefits parties’ underlying contractual arrangements).

476. See supra note 34 and accompanying test (describing how businesses which use arbitration clauses in their contractual arrangements desire a stable, “analytical framework” in order to facilitate both dispute resolution planning and establishing arbitral procedures).

certainty when establishing arbitral procedures. Further, both the prejudice and the rebuttal standards instill a complete lack of focus, hinging parties’ arbitral rights on undefined standards of pretrial conduct. Conversely, the reasonableness test subordinates prejudice, making it a mere factor in a larger analysis geared toward preserving parties’ expectations. Because parties expect stability, courts would be far less likely to tolerate dilatory “heads I win, tails you lose” pretrial strategies. If courts effectively police such strategies, parties would be far less likely to attempt them from the outset. Thus, by preventing dilatory pretrial conduct, the reasonableness test would further stabilize the arbitral process.

Second, the reasonableness test would realign the standard for assessing parties’ pretrial conduct with FAA’s overarching objective. First, its purely contractual methodology comports with the notion that arbitration is simply a matter of contract. Both the discharge and the reasonableness concepts stem from generally applicable contract law principles. Furthermore, the reasonableness test is aimed directly toward upholding parties’ legitimate expectations, a “fundamental principal” of contract law. To be sure, fewer disputes in all may be referred to arbitration, at least in the short term. This would not, however, contravene the liberal federal policy favoring arbitration agreements. Rather, as the policy merely reflects FAA § 2’s substantive mandate—enforcement of arbitration agreements as contracts—the reasonableness test falls directly within its confines.

478. See supra Part III (describing the circuit split and each circuit’s ill-defined prejudice requirement).
479. See Cabinetree of Wis., Inc. v. Kraftmaid Cabinetry, Inc., 50 F.3d 388, 391 (7th Cir. 1995).
480. See AT&T Mobility LLC v. Concepcion, 131 S. Ct. 1740, 1748 (2011) (“The overarching purpose of the FAA . . . is to ensure the enforcement of arbitration agreements according to their terms so as to facilitate streamlined proceedings.”).
481. See id. at 1752 (“Arbitration is a matter of contract . . . .”).
482. See supra notes 421, 427–28 and accompanying text (discussing the discharge and reasonableness principles under contract law).
484. See Concepcion, 131 S. Ct. at 1746 (“The FAA . . . places arbitration
Indeed, parties can neither waive nor forfeit an autonomous contract, the very regard in which arbitration agreements are held under the severability doctrine. Thus, both standards propound approaches clearly rejected by contract law and, as a result, the FAA.485

Furthermore, endless litigation over the details of prejudice,486 or rebuttal, impugns the FAA’s second underlying goal—to “facilitate streamlined proceedings.”487 Under the reasonableness test, however, disputes deemed referable to arbitration would be both contractually legitimate (i.e., tested against bad faith) and consigned expeditiously under a predictable framework. Thus, the reasonableness test would unite the FAA’s two goals, rather than frustrate both of them—a result clearly favored by the Supreme Court.488 Ultimately, the reasonableness framework accomplishes the FAA’s overarching objective to a greater extent than waiver both doctrinally and efficaciously.

Third and finally, the reasonableness test would prevent bad faith manipulation of the arbitral process. Indeed, bargained-for arbitral procedures confer significant benefits upon parties’ contractual relationships.489 Ensuring that parties exercise their arbitral rights in good faith, a requirement of all contractual arrangements,490 will ensure that they receive the benefit of their

agreements on an equal footing with other contracts . . . ” (citations omitted)); Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc., 473 U.S. 614, 625 (1985) (“The liberal federal policy favoring arbitration agreements . . . is at bottom a policy guaranteeing the enforcement of private contractual arrangements . . .”).

485. See supra Part IV (discussing the doctrinal inaccuracies of applying the waiver doctrine and forfeiture in the context of analyzing a party’s arbitral rights).

486. See Erdman Co. v. Phoenix Land & Acquisition, LLC, 650 F.3d 1115, 1120 (8th Cir. 2011) (finding that requiring parties to “litigate endlessly over the details of prejudice” makes little sense in certain circumstances).


488. See id. at 1749 (stating explicitly that its holding causes the FAA’s two goals to coincide, but eschewing the dissent’s approach in that it “would frustrate both of them” (emphasis in original)).

489. See supra Part II (discussing the benefits derived through established arbitral procedures).

490. See supra note 439 and accompanying text (noting that a covenant of good faith and fair dealing is implied in all contractual arrangements).
bargain in accordance with contract law. 491 Thus, the reasonableness test provides not only a doctrinally accurate approach, but an equitable approach.

VI. Conclusion

Arbitration is, quite simply, a matter of contract. 492 It is “a private process to which the parties have agreed, and the courts’ only obligation is to uphold that agreement pursuant to established arbitration and contract law.” 493 Despite this obligation’s intrinsic simplicity, the Circuit Courts of Appeals have consistently failed to uphold arbitration agreements as contracts, often to the detriment of parties’ arbitral rights. They have yanked the liberal federal policy favoring arbitration from its contractual roots, and use it to justify doctrinally inaccurate, manipulative results. 496

The Supreme Court has recently stated that the FAA’s overarching objective is “to enforce arbitration agreements according to their terms so as to facilitate streamlined proceedings.” 497 The reasonableness test proposed by this Note is precisely what the Supreme Court requires—a purely contractual approach which accomplishes the FAA’s objective. Further, it provides a comprehensive framework which would unify the circuit courts’ analyses and stabilize the arbitral process in

491. See Burton, supra note 483, at 506 (stating that ensuring parties’ legitimate expectations is a fundamental aspect of contract law).
494. See Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc., 473 U.S. 614, 625 (1985) (noting that the liberal federal policy favoring arbitration agreements “is at bottom a policy guaranteeing the enforcement of private contractual arrangements”).
495. See supra Part IV (describing why waiver, under contract law, is the incorrect approach to determine when a party has, through pretrial conduct, lost its right to arbitrate).
496. See supra notes 266–68 and accompanying text (describing how the circuits which impose burdensome prejudice requirements consciously acknowledge the manipulative strategies perpetuated by those requirements).
accordance with parties’ legitimate expectations.498 Under this approach, arbitration will remain, simply, a matter of contract.

498. See supra Part II (discussing how the stabilizing effect provided by bargained-for arbitral procedures is lost without a predictable legal framework).
FATCA: Toward a Multilateral Automatic Information Reporting Regime

Joanna Heiberg*

Table of Contents

I. Introduction ................................................................. 1686
II. International Tax Enforcement ...................................... 1688
   A. Offshore Tax Evasion .............................................. 1690
   B. Chapter 3 Withholding: The Starting Point for International Tax Enforcement ......................... 1691
   C. The Importance of Information Reporting .............. 1692
      1. Tax Treaties and Information Exchange Agreements ................................................................. 1693
      2. Qualified Intermediaries .................................... 1694
      3. Voluntary Disclosure ......................................... 1695
      4. Tax Whistleblowers ........................................... 1696
      5. John Doe Summons ........................................... 1696
      6. Title 31 Subpeonas ............................................ 1697

III. FATCA ........................................................................... 1698
   A. FATCA’s Goals and Mission .................................... 1698
   B. FATCA’s Design ...................................................... 1698
      1. Self-Reporting .................................................... 1699
      2. Third-Party Reporting ........................................ 1700

IV. FATCA Issues and Concerns ......................................... 1701
   A. Citizens Living Abroad ............................................ 1702
   B. Conflict of Laws: Tax Treaties and Banking Secrecy ..................................................................... 1703
   C. Costs and Administrative Burden on FFIs .......... 1704

* Candidate for J.D., Washington and Lee University School of Law, May 2013. I would like to thank my advisor, Professor Michelle Drumbl, for her guidance. I would also like to thank my family and friends for their support and encouragement.
I. Introduction

The Tax Justice Network estimates $11.5 trillion in global assets are hidden in offshore havens. Offshore tax evasion was the primary issue in the recent case of United States v. UBS AG, a dispute between the United States and Switzerland’s largest bank, United Bank of Switzerland (UBS). In 2007, former UBS banker and American citizen, Bradley Birkenfeld, revealed that UBS was actively involved in helping its U.S. clients evade taxes. Birkenfeld revealed that UBS managed assets for 19,000 U.S. clients totaling approximately $19 billion. He estimated that 90% of his own clients were trying to evade taxes. He neglected, however, to disclose his dealings with his biggest client, California real estate developer Igor Olenicoff, whom Birkenhoff helped hide $200 million. Birkenfeld later pled guilty to conspiracy to defraud the IRS and was sentenced to forty months in prison.
penalties, interest, and restitution. Subsequently, the United States sued UBS in an attempt to force disclosure of nearly 52,000 secret accounts. As a result, the Swiss government struck an unprecedented deal with the United States to provide client names on 4,450 UBS accounts held by Americans despite its previous argument that handing over such information violated Swiss bank-secrecy laws. Although the case was arguably a success for American tax enforcement officials, it brought to light many of the current inadequacies of U.S. international tax enforcement.

In response to the UBS case and the call for heightened international tax enforcement, Congress enacted the Foreign Account Tax Compliance Act (FATCA). By enhancing


7. If Switzerland Can, supra note 1.

8. See Hearing on Banking Secrecy Practices and Wealthy American Taxpayers Before the H. Comm. on Ways and Means, 111th Cong. 4–5 (2009) [hereinafter Hearing on Banking Secrecy] (“The ongoing events surrounding UBS AG and its admitted criminal role in helping a number of wealthy U.S. individuals evade U.S. taxes have brought a spotlight to bear on international tax enforcement and the tools that we have at our disposal to help ensure compliance.”); see also infra Part II.C (discussing existing international enforcement mechanisms).

information reporting, increasing withholding taxes for foreign financial institutions that do not engage in information reporting, and strengthening penalties for taxpayers who do not adequately report their income, FATCA makes it more difficult for U.S. persons to engage in offshore tax evasion.\footnote{Foreign Bank Account Reporting and Tax Compliance: Hearing Before the Subcomm. on Select Revenue Measures of the H. Comm. on Ways and Means, 111th Cong. 7 (2009) [hereinafter FATCA Hearing] (statement of Stephen E. Shay, Deputy Assistant Sec’y of the Treasury).} Despite FATCA’s worthwhile goals of increasing tax enforcement and tracking down tax evaders, its enactment raises several significant concerns.\footnote{See infra Part IV (discussing primary areas of concern).}

This Note will argue that international cooperation is essential for successful FATCA implementation. Part II will provide background information on offshore tax evasion and existing U.S. mechanisms for international tax enforcement. Part III will explain key FATCA provisions, and Part IV will discuss concerns regarding FATCA as originally enacted. Finally, Part V will introduce the proposed intergovernmental approach to FATCA and argue that international cooperation and development of standardized requirements will mitigate FATCA concerns and facilitate its implementation. Part V also argues that abandonment of the U.S. policy of citizenship-based taxation is necessary to achieve an efficient multilateral FATCA regime.

\section*{II. International Tax Enforcement}

The U.S. federal income tax system is one of “voluntary compliance,”\footnote{See Flora v. United States, 362 U.S. 145, 176 (1960) (“Our system of taxation is based upon voluntary assessment and payment, not upon distraint.”).} meaning it is initially up to the taxpayer, rather than the government, to determine and pay the appropriate taxes.\footnote{See Rev. Rul. 2007-20, 2007-14 I.R.B. 863–64 (“References to a ‘voluntary’ tax system . . . mean a system that allows taxpayers to determine, in the first instance, the correct amount of their tax and report their liability on appropriate returns, rather than having the government make the determinations for them.”).} The United States currently has one of the world’s
highest compliance rates for tax collections. Despite a relatively high compliance rate, the United States suffers from a significant tax gap. The tax gap is the difference between the amount of tax imposed by law and the amount voluntarily and timely paid by taxpayers for a given year. Nonfiling, underreporting, and underpayment of taxes are all forms of noncompliance that contribute to the tax gap.

The two primary means of tax enforcement are withholding and information reporting. Withholding taxes at the source eliminates the possibility of nonpayment. Information reporting, on the other hand, ensures the government has another source of information to compare against the taxpayer's filing. Thus, a large portion of the tax gap results from income that is subject to neither withholding nor information reporting.


15. The most recent tax gap estimates from the IRS are based on data from tax year 2006 and report a net tax gap of $385 billion for that year, including revenue collected from late payments and IRS compliance and enforcement efforts. IR-2012-4.

16. Nina E. Olson, Minding the Gap: A Ten-Step Program for Better Tax Compliance, 20 STAN. L. & POL’Y REV. 7, 8 (2009); see also Hearing on Banking Secrecy, supra note 8, at 68 (noting that the tax gap refers to “a difference between the taxes [the IRS] collected and taxes it should have collected under existing law”).


18. Hearing on Banking Secrecy, supra note 8, at 68; see also IR-2012-4 (“Compliance is highest where there is third-party information reporting and/or withholding.”).


20. Id.; see also James Andreoni, Brian Erard & Jonathan Reinstein, Tax Compliance, 36 J. ECON. LITERATURE 818, 821 (1998) (noting that “[i]nformation reporting severely limits the scope for tax evasion on many significant income and deduction items” and “reduces the potential for unintentional reporting errors by clarifying for the taxpayer the amount that legally should be reported”).

21. Hearing on Banking Secrecy, supra note 8, at 68.
A. Offshore Tax Evasion

Experts estimate that offshore tax evasion costs the U.S. Treasury approximately $100 billion a year.22 Because domestic tax laws vary from country to country, international taxation23 provides unique opportunities for noncompliance and complicates enforcement efforts.24 This is particularly true of tax haven countries. Although there is no single definition, the Organization for Economic Cooperation and Development (OECD) identifies four primary characteristics of tax haven countries: (1) absence of or nominal amount of taxes imposed; (2) lack of transparency about the application of tax laws and underlying documentation; (3) laws or administrative practices that prevent the effective exchange of information with other countries for tax purposes; and (4) absence of a requirement that the taxpayer’s activity within their jurisdiction be substantial.25 Access to these low-tax, secretive jurisdictions provides an opportunity for U.S. persons to effectively avoid taxation by moving assets and investments offshore.26

22. Staff of S. Comm. on Homeland Sec. and Gov’t Affairs Subcomm. on Investigations, 110th Cong., Rep. on Tax Haven Banks and U.S. Tax Compliance 1 (Comm. Print 2008). But see Hearing on Banking Secrecy, supra note 8, at 95 (recognizing that there are a range of estimates and it is difficult to get a precise number).

23. For the purposes of this Note, the term “international taxation” refers to the U.S. taxation of international transactions, persons, and investments. Taxpayers subject to the international provisions of the U.S. tax rules and reporting requirements can be grouped into four categories: (1) U.S. individuals working, living, or holding assets abroad; (2) U.S. entities doing business abroad; (3) Foreign individuals working or doing business in the United States; and (4) Foreign entities doing business in the United States. TAXPAYER ADVOCATE SERV., 2011 ANNUAL REPORT TO CONGRESS 129 (2011).

24. Dizdarevic, supra note 19, at 2972.


26. See, e.g., Hearing on Banking Secrecy, supra note 8, at 67 (explaining the extent to which U.S. residents can move assets offshore to tax haven jurisdictions and avoid paying U.S. taxes).
As a preliminary matter, the United States uses a withholding system for U.S. source payments to foreign persons.27 A withholding agent must withhold 30% of any payment of fixed or determinable annual or periodical (FDAP) income made to a payee that is a foreign person, unless it has documentation associating the payee with either a U.S. person or a foreign “beneficial owner” entitled to a reduced withholding rate.30 A withholding agent making a payment to a foreign person need not withhold when the foreign person assumes withholding responsibility as a qualified intermediary (QI), a U.S. branch of a foreign person, a withholding foreign partnership, or an authorized foreign agent.32 Therefore, by withholding at the source, U.S. tax authorities ensure collection of the appropriate amount of tax on certain U.S. source income of foreign persons.33 Because withholding does not apply to foreign assets of U.S. persons, however, additional enforcement mechanisms are needed to address offshore tax evasion.

29. “The term beneficial owner means the person who is the owner of income for tax purposes and who beneficially owns that income.” Treas. Reg. § 1.1441-1(c)(6).
30. Id. § 1.1441–1(b). A foreign person entitled to a treaty reduction of the 30% withholding rate must provide the withholding agent with a W–8BEN form prior to the time of payment in order to receive the treaty benefits. Id. § 1.1441–6(b)(1).
31. See infra Part II.C.3 (explaining the QI program).
33. See supra notes 18–21 and accompanying text (noting that withholding at the source increases the voluntary compliance rate).
C. The Importance of Information Reporting

A core problem in tax enforcement is information asymmetry. A taxpayer has knowledge of, or at least ready access to, the information necessary to determine his or her tax liability. The government, on the other hand, must rely on the information it obtains directly from the taxpayer or from third parties. This is true in both the domestic and the international context. For example, U.S. persons are legally obligated to disclose foreign accounts in excess of $10,000 on an annual Report of Foreign Banks and Financial Accounts (FBAR) form. Without third-party reporting, however, the government has no way of knowing if a taxpayer failed to disclose or underreported his or her foreign assets. Thus, “third-party information reporting assists taxpayers in correctly computing and reporting their tax liabilities, increases compliance with tax obligations, reduces the incidence of and opportunities for tax evasion, and . . . helps to maintain the fairness of the U.S. federal income tax system.”

The United States uses a variety of tools to collect information regarding U.S. persons with foreign assets. Current information gathering methods include both cooperative and unilateral measures.
1. Tax Treaties and Information Exchange Agreements

There are two principal forms of bilateral agreements used by U.S. tax authorities for the exchange of information with other countries: (1) articles in income tax treaties governing the exchange of information and (2) Tax Information Exchange Agreements (TIEAs). Article 26 of the U.S. Model Income Tax Treaty requires the contracting countries to exchange tax information as necessary for carrying out provisions of the treaty or domestic laws of the parties. Information received under Article 26 is treated as confidential in the same manner as information obtained under the domestic laws of that country. The agreement is limited to information that is legally obtainable in the normal course of administration of the requested country and does not include trade secrets. Due to this limitation, banking secrecy rules may excuse production of information.

Similarly, TIEAs are bilateral agreements between countries establishing policies and procedures regarding the exchange of information. TIEAs are generally used when the parties do not have a comprehensive bilateral income tax treaty containing an information exchange provision. The agreements do not modify the substantive rules of taxation in either country, but instead provide for cooperation in eliciting information and other assistance for each country in the enforcement of its tax laws. Since the United States

IRS to obtain all the evidence needed for a successful investigation and prosecution. Id.

40. Hearing on Banking Secrecy, supra note 8, at 32 (statement of Peter H. Blessing, Partner, Shearman & Sterling).


43. Id. para. 2.

44. Id. para. 3.

45. Hearing on Banking Secrecy, supra note 8, at 33 (statement of Peter H. Blessing, Partner, Shearman & Sterling).

46. Id.

and Barbados entered into the first TIEA in 1984, the United States has extensively used TIEAs to obtain information needed for tax enforcement.48 Banking secrecy laws may also inhibit the United States’ ability to gather information pursuant to a TIEA.49

2. Qualified Intermediaries

In addition to income tax treaties and information exchange agreements with other countries, the United States uses a QI program to incentivize cooperation from foreign entities.50 The QI program began in 2001 and is primarily directed at U.S.-source income received by foreigners.51 Pursuant to a QI agreement,52 a foreign institution provides the IRS with certain information regarding its U.S. customers in exchange for simplified rules, including non-customer-specific reporting and the ability to claim more easily applicable exemptions or lower withholding taxes.53 Although the QI program is a valuable tool in international tax enforcement, it has material shortcomings and, as evidenced by the

48. Id.
49. Supra note 45 and accompanying text.
50. The term “qualified intermediary” refers to an entity that is a party to a withholding agreement with the IRS and may be a foreign financial institution, a foreign clearing house, a foreign branch of a U.S. financial institution or clearing organization, a foreign corporation for purposes of presenting claims of benefits under an income tax treaty on behalf of its shareholders, or any other person acceptable to the United States. Treas. Reg. § 1.1441-1(e)(5)(ii). The IRS currently has over 5,000 QI Agreements in force. Hearing on Banking Secrecy, supra note 8, at 40 (statement of Peter H. Blessing, Partner, Shearman & Sterling).
53. Treas. Reg. § 1.1441-1(e)(5); see also FATCA Hearing, supra note 10, at 7.
UBS case,\textsuperscript{54} considerable potential for abuse with respect to U.S. accountholders.\textsuperscript{55}

3. Voluntary Disclosure

On several occasions, the IRS has offered amnesty for voluntary disclosure by allowing U.S. taxpayers to disclose foreign assets without the threat of criminal punishment.\textsuperscript{56} Although taxpayers escape criminal liability under the offshore voluntary disclosure initiatives (OVDI), financial penalties still apply.\textsuperscript{57} The 2009 and 2011 OVDIs resulted in a combined 33,000 voluntary disclosures and the collection of more than $4.4 billion.\textsuperscript{58} The repeated use of voluntary disclosure initiatives, however, may result in diminishing returns and reduce incentives to comply due to a belief that an opportunity will exist to come forward under future amnesties.\textsuperscript{59}

\textsuperscript{54} See FATCA Hearing, supra note 10, at 2 ("The recent UBS case revealed problems with the QI program that permitted tax evasion by U.S. persons."); see also Hearing on Banking Secrecy, supra note 8, at 40–41 (statement of Peter H. Blessing, Partner, Shearman & Sterling) (assessing the QI program and its current shortcomings).

\textsuperscript{55} FATCA Hearing, supra note 10, at 9–10 ("While [the QI] regime has been effective in improving compliance with U.S. tax laws in relation to collecting withholding taxes on foreign persons, its provisions relating to U.S. accountholders have been subject to abuse by some foreign banks.").

\textsuperscript{56} See I.R.S. News Release IR-2011-94 (Sept. 15, 2011) ("The programs gave U.S. taxpayers with undisclosed assets or income offshore a second chance to get compliant with the U.S. tax system, pay their fair share and avoid criminal charges."); see also I.R.S. News Release IR-2012-5 (Jan. 9, 2012) (announcing the reopening of the IRS voluntary disclosure program).

\textsuperscript{57} See IR-2012-5 (reporting that the 2012 program imposes a penalty of 27% of the highest aggregate balance in foreign accounts or value of foreign assets during the eight tax years prior to the disclosure).

\textsuperscript{58} Id.

\textsuperscript{59} See Leandra Lederman, The Use of Voluntary Disclosure Initiatives in the Battle of Offshore Tax Evasion, 55 Vill. L. Rev. (forthcoming 2012) (arguing that repeated use of voluntary disclosure initiatives may have diminishing returns unless the government continues to make well-publicized criminal prosecutions).
4. Tax Whistleblowers

Another tool used for information gathering is whistleblower incentives. Under U.S. tax laws, a whistleblower may be entitled to up to 30% of the amount collected by the government from a noncompliant taxpayer.\textsuperscript{60} Despite the high potential monetary rewards, whistleblowers may be reluctant to come forward due to the infrequency of whistleblower rewards awarded\textsuperscript{61} and the possibility of criminal punishment.\textsuperscript{62}

5. John Doe Summons

When the IRS suspects a federal tax violation has occurred by unknown persons, it may issue a “John Doe” summons for the relevant financial information.\textsuperscript{63} The IRS may summon information when it can establish that (1) the summons relates to a particular person or ascertainable class; (2) a reasonable basis exists for issuing the summons; and (3) no adequate alternative for acquiring the information exists.\textsuperscript{64} To enforce a summons in U.S. courts, the IRS must prove that the investigation serves a legitimate purpose, the information might be relevant to that

\textsuperscript{60} I.R.C. § 7623(b)(1) (2006). Whistleblowers may be entitled to 15–30% of the amount collected, including penalties and interest, as a result of administrative or judicial action or settlement. \textit{Id.} The amount of the award depends upon the extent to which the individual substantially contributed to the action or settlement. \textit{Id.}


\textsuperscript{62} UBS whistleblower Bradley Birkenfeld was sentenced to 40 months in prison and received a $30,000 fine despite offering the U.S. Government information crucial to its UBS investigation. Erik Larson & Carlyn Kolker, \textit{UBS Tax Fraud Case Whistleblower Gets 40-Month Prison Sentence}, \textbf{BLOOMBERG} (Aug. 21, 2009, 3:11 PM), http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aqRUmD2LzH.E (last visited Sept. 21, 2012) (on file with the Washington and Lee Law Review); see also supra note 4 (detailing Birkenfeld’s involvement in the UBS investigation).

\textsuperscript{63} I.R.C. § 7609(f) (2006).

\textsuperscript{64} \textit{Id.}
purpose, the IRS did not already possess the information, and the initial requirements of Section 7609(f) were met. Because a John Doe summons is a unilateral mandate, enforcement issues arise when a foreign entity holds the information.

6. Title 31 Subpoenas

Most recently, the IRS has used Title 31 subpoenas to compel U.S. taxpayers suspected of holding offshore accounts to turn over bank account details. Under Title 31 of the U.S. Code, the U.S. government may compel production of records and documents for purposes of an investigation. When subpoenaed, the taxpayer faces a choice of disclosing potentially self-incriminating evidence or being found in contempt of court and subject to civil or criminal penalties, including jail time. Approximately a dozen Title 31 subpoenas were issued in 2011, and it remains to be seen whether the IRS will increase its use of this new enforcement technique.

69. 31 U.S.C. § 3804(a) (2006). Title 31 subpoenas are more commonly used against drug smugglers and money launderers. Browning, supra note 68.
70. Id. § 3804(c). If an individual refuses to obey a subpoena, the U.S. District Courts have jurisdiction to enforce the subpoena. Id. Failure to obey a court order is punishable by the court as contempt. Id.
71. Id. Recently, a wealthy California taxpayer unsuccessfully challenged a Title 31 subpoena on Fifth Amendment grounds. M.H. v United States, 648 F.3d 1067 (9th Cir. 2011).
III. FATCA

Congress enacted FATCA in response to the gaps in existing tax enforcement mechanisms.\footnote{See supra Part II.C (discussing existing enforcement mechanisms).} Although the ability to use offshore tax havens to evade income taxes has increased in recent decades, the U.S. government’s tools to combat evasion have not changed significantly.\footnote{See Hearing on Banking Secrecy, supra note 8, at 68 (statement of Reuven S. Avi-Yonah, Professor of Law, Univ. of Mich. Law School) (noting that “since about 1980 there has been a dramatic lowering of both legal and technological barriers to the movement of capital, goods and services”); Grinberg, supra note 51, at 2 (“The ability to make, hold, and manage investments through offshore financial institutions has increased dramatically in recent years, while the cost of such services has plummeted.” (citations omitted)).} FATCA provides the IRS with the “enhanced tools it needs to continue its expansion of international tax enforcement.”\footnote{FATCA Hearing, supra note 10, at 17 (statement of William J. Wilkins, Chief Counsel, Internal Revenue Serv.).}

A. FATCA’s Goals and Mission

FATCA’s primary goal is to raise revenue by tracking down tax evaders.\footnote{Micah Bloomfield & Dmitriy Shamrakov, The Thirty Percent Solution?: FATCA Provisions of the HIRE Act, STROOCK SPECIAL BULLETIN, Apr. 21, 2010, at 1.} More specifically, it is designed to address the “deliberate and illegal hiding of assets and income from the IRS by U.S. citizens and residents.”\footnote{FATCA Hearing, supra note 10, at 13.} Because an estimated $100 billion is lost annually as a result of offshore tax abuses,\footnote{156 Cong. Rec. S1745 (daily ed. Mar. 18, 2010); see also supra note 22 and accompanying text.} a significant amount of revenue is at stake.

B. FATCA’s Design

FATCA is designed to increase the government’s access to information regarding U.S. citizens and residents with foreign assets, in hopes of detecting and deterring offshore tax evasion. It
adopts a two-prong approach to enhancing the Government’s access to information, one focusing directly on taxpayers and the other on foreign financial institutions (FFIs). 78

1. Self-Reporting

FATCA requires individual U.S. taxpayers with foreign accounts and assets exceeding $50,000 on the last day of the tax year, or $150,000 at any time during the tax year, to report them on an information return. 79 Most American citizens living abroad will likely have over $50,000 in foreign financial assets and, as a result, meet the threshold for Section 6038 reporting. 80 Failure to disclose results in an initial $10,000 penalty. 81 Additionally, the new provisions extend the statute of limitations for failure to disclose from three years to six. 82 These reporting requirements operate in addition to FBAR filing requirements 83 and may be enforced using traditional IRS enforcement mechanisms. 84

78. Hiring Incentives to Restore Employment Act, Pub. L. No. 111–147, § 501, 511, 124 Stat. 71 (2010). FATCA also addresses several other areas of foreign asset reporting that are beyond the scope of this Note. See id. at § 521–41 (addressing passive foreign investment companies, foreign trusts, and dividend equivalent payments received by foreign persons).


82. Id. § 513(a), 124 Stat. at 111 (to be codified at 26 U.S.C. § 6501).

83. See IR-2011-117 (“The new Form 8938 requirement does not replace or otherwise affect a taxpayer’s obligation to file an FBAR.”).

84. See FATCA Hearing, supra note 10, at 15 (noting that FATCA creates reporting obligations and a penalty regime separate from FBAR and allows the IRS to enforce the new penalties using traditional IRS enforcement tools). Conversely, the IRS cannot enforce FBAR penalties because they are imposed through the Bank Secrecy Act and not the Code. Id. at 14. An FBAR penalty must instead be referred to the Justice Department for separate prosecution and collection. Id. at 15.
2. Third-Party Reporting

Perhaps the most controversial aspect of FATCA is its application to FFIs.\(^{85}\) FATCA gives FFIs a choice between disclosure of U.S. account holders and a 30% withholding tax on U.S. source income.\(^{86}\) In doing so, FATCA creates a powerful incentive for FFIs to disclose information regarding U.S. accounts.\(^{87}\)

To meet the information reporting requirements, an FFI must agree to collect information necessary to identify its U.S. accounts.\(^{88}\) For each U.S. account, the FFI must report the name, address, and Tax Identification Number (TIN) of each account holder; the account number; the account balance; and the gross receipts and withdrawals from the account.\(^{89}\) Alternatively, an FFI may elect to be subject to the same reporting requirements as U.S. financial institutions.\(^{90}\) These requirements operate in addition to any existing reporting obligations under a QI Agreement.\(^{91}\)

If an FFI chooses not to comply with FATCA’s reporting requirements, a 30% withholding tax is imposed on all “withholdable payments.”\(^{92}\) There are several notable distinctions

---

\(^{85}\) “FFIs are defined in such a manner as to include foreign banks, foreign brokerage firms, foreign trust companies, foreign mutual funds, foreign hedge funds, foreign private equity funds, and other foreign funds engaged primarily in investing or trading in U.S. or foreign securities.” Lederman & Hirsh, supra note 80, at 1143 (citing I.R.C. § 1471(d)(4), (5)). A subsequent IRS notice clarified this definition and excepted from FFI characterization certain foreign companies. See I.R.S. Notice 2010–60.

\(^{86}\) I.R.C. § 1471(a).

\(^{87}\) See 156 Cong. Rec. S1746 (daily ed. Mar. 18, 2010) (“The legislative intent behind [this choice] is to force foreign financial institutions to disclose their U.S. account holders.”); see also FATCA Hearing, supra note 10, at 10 (statement of Stephen E. Shay, Deputy Assistant Sec’y of the Treasury) (“[FATCA] will improve information reporting with respect to U.S. accountholders by creating a powerful incentive for [FFIs] to provide the IRS with the information it needs to identify persons seeking to evade U.S. tax.”).

\(^{88}\) I.R.C. § 1471(b) (2006).

\(^{89}\) Id.

\(^{90}\) Id.

\(^{91}\) Id. § 1471(c)(3).

\(^{92}\) Id. § 1471(a). The term “withholdable payment” includes “any payment of interest . . . dividends, rents, salaries, wages, premiums, annuities, compensations, remunerations, emoluments, and any other fixed or
between FATCA withholding and existing withholding requirements under Chapter 3 of the Code. First, FATCA imposes withholding on gross proceeds from the sale or disposition of income-producing property from a U.S. source and fixed or determinable annual or periodical (FDAP) income. Second, Chapter 3 withholding only applies to payments to nonresident aliens and foreign corporations, whereas FATCA withholding applies to all U.S. source payments. Finally, failure to meet FATCA information disclosure requirements may result in a QI being withheld upon. As a result, FATCA creates a more expansive withholding scheme than previously existed.

IV. FATCA Issues and Concerns

Although many agree that increased enforcement of offshore tax evasion is desirable, critics have raised numerous issues associated with FATCA’s approach.
A. Citizens Living Abroad

FATCA’s enactment caused significant outcry from U.S. citizens living abroad.\textsuperscript{100} Due to the financial and administrative burdens of holding U.S. accounts under a FATCA regime,\textsuperscript{101} some FFIs are severing ties with U.S. accountholders.\textsuperscript{102} As a result, U.S. citizens may be unable to establish and maintain bank accounts in foreign countries.\textsuperscript{103} Depending on eventual FATCA exclusions, U.S. citizens may also face difficulty in purchasing certain foreign insurance policies and pension funds.\textsuperscript{104}

Additionally, the burdensome individual reporting requirements and harsher penalties are causing Americans to re-evaluate, and in some cases terminate, their U.S. citizenship.\textsuperscript{105}

---


\textsuperscript{101} See infra Part IV.C (discussing the financial and administrative burden on FFIs).


\textsuperscript{103} See FATCA Hearing, supra note 10, at 70 (“The [FATCA] legislation and reinforced QI regulations will make it all the more difficult for overseas Americans to maintain a bank account where they reside.”).

\textsuperscript{104} See Lederman & Hirsh, supra note 80, at 1147 (discussing obstacles that citizens living abroad may face under FATCA).

Renunciation of citizenship, however, comes at a price. Individuals are not relieved of their existing tax liabilities and are required to file back taxes and pay any penalties owed. Furthermore, an exit tax is imposed on individuals with an annual income of approximately $150,000 or a net worth of at least $2 million. Thus, renunciation only serves as a way for individuals to avoid U.S. reporting requirements and related penalties after the date of renunciation.

B. Conflict of Laws: Tax Treaties and Banking Secrecy

A second issue raised by FATCA is that it overrides any conflicting provisions contained in current income tax treaties. The United States’ policy concerning treaty override provides that “the treaty is superior if it is implemented after a law, but the law is superior if it is implemented after the treaty.” FATCA requires greater information reporting than current tax treaties. If an FFI does not comply, FATCA imposes a 30% withholding tax regardless of reduced withholding rates under a governing treaty. To obtain preferential rates under a treaty, taxpayers must rely on a refund mechanism. Despite the

Law Review).

106. The Department of State currently imposes a $450 fee for renunciation. Id.

107. I.R.C. § 877(a)(2) (2006). All property of an individual subject to these provisions is treated as sold on the day before expatriation for its fair market value and any gain or loss on such sale is recognized for the taxable year. Id. § 877A(a).

108. See FATCA Hearing, supra note 10, at 62 (“Treaty overrides adversely affect the treaty-making process and historically have been avoided unless essential to the ends sought by the legislation.”).

109. Reuven S. Avi-Yonah, Double Tax Treaties: An Introduction, in The Effect of Treaties on Foreign Direct Investment 105 (Karl P. Sauvant & Lisa E. Sachs eds., 2009). This is generally referred to as the “later in time” rule. See Whitney v. Robertson, 124 U.S. 190, 194 (1888) (stating that if a treaty and congressional law conflict, the treaty dated more recently controls).

110. Compare supra note 89 and accompanying text (detailing FATCA reporting requirements), with supra notes 42–46 and accompanying text (explaining the U.S. Model Treaty provisions regarding information exchange).


112. This practice is largely consistent with existing refund mechanisms
possibility of a refund, the taxpayer is not necessarily held
harmless, as no interest will be paid on the refund amount
where an FFI is the beneficial owner of the payment.113

Bank secrecy laws raise another issue of FATCA
enforcement.114 Because many countries forbid banks or
companies to transfer client information directly to a foreign
government, FFIs in those countries that serve U.S. citizens
will have no choice but to be withheld upon.115 Some view
FATCA’s conflict with bilateral agreements and foreign laws as U.S.
legislative overreach and an imposition on national
sovereignty.116

C. Costs and Administrative Burden on FFIs

Another concern with FATCA’s requirements is the resulting
burden on FFIs. First, banks face potentially high compliance
costs related to new technology and staffing.117 The Institute of
International Bankers estimates that major global banks may

under Section 1445(c)(1)(c). See id. § 1474(b)(1) (stating that, subject to certain
exceptions, determination of whether a withholding resulted in overpayment
shall be made as if the tax had been deducted and withheld under Chapter 3 of
the Code). It is unclear whether FATCA withholding may be reduced to the level
specified in a particular treaty if the withholding agent receives a beneficial
owner withholding certificate (W-8BEN), as allowed under preexisting law.


114. See David Jolly & Brian Knowlton, Law to Find Tax Evaders
Denounced, N.Y. TIMES, Dec. 27, 2011, at 1 (“Enforcement of the law will be
tricky, as many countries . . . forbid banks or companies to transfer such
information directly to a foreign government.”) (quoting Jeffrey Owens, tax
expert at the Organization for Economic Cooperation and Development).

115. Id.

116. See Scott D. Michel & H. David Rosenbloom, FATCA and Foreign Bank
Accounts: Has the U.S. Overreached?, TAX ANALYSTS 709, 711 (2011), available
at http://www.capdale.com/files/4178_FATCA%20Article.pdf (“In an era of
delicately negotiated tax treaties and information exchange agreements
between the United States and other nations, FATCA is seen as overreaching.”).

117. John Greenwood, TD Resists U.S. Plan to Catch Tax Cheats, FINANCIAL
opposes-u-s-plan-to-catch-tax-cheats/ (last visited Sept. 21, 2012) (noting that
Toronto-Dominion Bank is resisting the regulation because it would cause the
bank to incur $100 million in compliance costs) (on file with the Washington and
Lee Law Review).
spend over $250 million to comply with the regulation, while some businesses fear the annual costs will be in the billions. Additionally, the reporting requirements create an administrative burden and raise efficiency concerns. Ultimately, FATCA forces foreign institutions to bear the costs of tracking down U.S. tax evaders.

**D. Detriment to U.S. Investments**

Due to FATCA’s strategic design, the only way a foreign entity can avoid the reach of these provisions is by not investing in the United States. Simply refusing to accept U.S. persons as account holders will not relieve a foreign entity from being subject to the FATCA provisions because it is the payment of U.S. source income that triggers its application. Thus, the desire to avoid costs associated with compliance may discourage U.S. investments. Alternatively, FFIs that choose to comply might shift compliance costs to U.S. investors.

118. Id. But see A Temperature Check: Who’s Ready for FATCA?, RBC Dexia Investor Services 8 (2011) (reporting that 85% of respondents estimate costs at $1 million or less, with the majority (54%) expecting to spend less than $100,000).

119. See Jolly & Knowlton, supra note 114 (“FATCA . . . is now causing alarm among businesses outside the United States that fear they will have to spend billions of dollars a year to meet the greatly increased reporting burdens.”).

120. See Hearing on Banking Secrecy, supra note 8, at 31 (statement of Peter H. Blessing, Partner, Shearman & Sterling) (noting that “there is a real economic efficiency cost to reporting requirements”).

121. See Jolly & Knowlton, supra, note 114 (“[Foreign] entities are being asked, in effect, to pay for the cost of tracking down American tax evaders.”).

122. Tello, supra note 9, at 3.

123. Id.

124. See Comment: FATCA Could Well Cause Managers to Turn Their Back on the US, Hedgeweek (May 9, 2011), http://www.hedgeweek.com/2011/09/05/130115/comment-fatca-could-well-cause-managers-turn-their-backs-us (last visited Sept. 21, 2012) (predicting foreign managers and investors will choose not to maintain U.S. assets) (on file with the Washington and Lee Law Review); see also FATCA Hearing, supra note 10, at 30 (citing concerns that the real cost of U.S. investment will increase significantly for non-U.S. investors); id. at 70 (noting that, as a result of FATCA, several FFIs are refusing to invest in American securities).

125. Bloomfield & Shamrakov, supra note 75, at 6.
E. Lack of Adequate Taxpayer Services

Finally, FATCA introduces additional complexity into an already highly complex system of international taxation. Although the IRS claimed improvements to taxpayer service as its top strategic goal in 2008, there has been a shift in focus from taxpayer service to international law enforcement. In her 2011 annual report to Congress, National Taxpayer Advocate Nina Olson expressed concern that “[t]he lack of efficient IRS-wide coordination of international taxpayer service may undermine the international enforcement initiatives and discourage future compliance by taxpayers dealing with the complexity and procedural burden of the international tax rules.”

V. International Collaboration Is Essential to FATCA Implementation

A. International Issues Require an International Solution

Because FATCA specifically targets foreign sources of information, international collaboration and compromise is essential to its successful implementation. The development of a multilateral agreement could mitigate the potential unintended consequences of FATCA in several key ways. First, international agreement would reduce the practical effect of treaty override and conflict with foreign law. Further, the existence of a multilateral information exchange regime may

126. See TAXPAYER ADVOCATE SERV., supra note 23, at 32–33 (explaining that the “complexity and administrative detail of the international reporting requirements are overwhelming”).
127. See id. at 177 (detailing the IRS's organizational shift away from taxpayer service and toward enforcement).
128. Id. at 176.
129. See FATCA Hearing, supra note 10, at 10 (noting that international cooperation and coordination is key to FATCA's success) (statement of William J. Wilkins); id. at 59 (“A multi-lateral agreement on the sharing of taxpayer financial information would better serve the enforcement objectives of FATCA without [the] unintended consequences.”) (statement of Dirk Suringa, Covington & Burlington LLP).
130. Id. at 62.
131. Id.
justifying placing higher regulatory burdens on FFIs, as it would benefit more than one country. Similarly, developing international standards for cross-border information reporting would improve efficiency and lower compliance burdens on FFIs.\footnote{Lack of international standards presents the possibility that FFIs will be subject to a variety of conflicting regulations from different countries.} International agreement would also prevent governments from adopting a reciprocal withholding tax or conflicting reporting measures.\footnote{See FATCA Hearing, supra note 10, at 61.} Finally, the more jurisdictions that participate in a multilateral automatic information exchange agreement, the less of an incentive FFIs and foreign investors would have to divest from the United States.\footnote{Id.}

\textbf{B. Other Countries Are Willing to Collaborate}

Because tax evasion is a global problem,\footnote{Doug Saunders, We Need a Global Army of Tax Collectors, GLOBE AND MAIL (Oct. 15, 2011, 2:00 AM), http://www.theglobeandmail.com/news/opinions/opinion/we-need-a-global-army-of-tax-collectors/article2201647/ (last visited Sept. 21, 2012) (“The non-payment of tax has become a chronic problem around the world.”) (on file with the Washington and Lee Law Review); Jolly & Knowlton, supra note 114.} information exchange is of great interest to foreign tax authorities.\footnote{Hearing on Banking Secrecy, supra note 8, at 38 (statement of Peter H. Blessing, Partner, Shearman & Sterling) (“The information exchange process is of great interest to U.S. tax authorities, but for similar reasons it is of great interest to many foreign tax authorities.”).} “Since April 2009, a growing number of governments and NGOs have called for the automatic exchange of tax information.”\footnote{Grinberg, supra note 51, at 3 (citations omitted).} In recent years both the EU and the OECD developed proposals for multilateral automatic information reporting.\footnote{See id. at 17–23 (describing the current OECD and EU approaches to cross-border information reporting).} This trend evidences an acknowledgement that collaboration is essential in the fight against offshore tax evasion\footnote{Hearing on Banking Secrecy, supra note 8, at 9 (statement of Doug Shulman, Comm’r, Internal Revenue Serv.).} and a willingness to participate in a multilateral automatic information reporting regime.
C. Toward an Intergovernmental Approach

The United States took the first step toward international cooperation in February 2012.\(^{140}\) In addition to releasing the proposed FATCA regulations,\(^{141}\) the Treasury Department issued a Joint Statement\(^{142}\) with France, Germany, Italy, Spain, and the United Kingdom announcing the desire to cooperate in an intergovernmental approach to improving international tax compliance and implementing FATCA.\(^{143}\) Notably, the United States expressed its intent to reciprocate in collecting and exchanging information on accounts held in U.S. financial institutions by residents of FATCA partner countries.\(^{144}\) The intergovernmental approach would address “legal impediments to compliance, simplify practical implementation, and reduce FFI costs.”\(^{145}\) International collaboration, therefore, “would enhance compliance and facilitate enforcement to the benefit of all parties.”\(^{146}\)

The joint statement proposes a bifurcated routing approach,\(^{147}\) one for FATCA partners and another for non-partner nations. FFIs established in a FATCA partner nation would report information to that country.\(^{148}\) This system ensures that

\(^{140}\) Regulations Relating to Information Reporting by Financial Institutions and Withholding on Certain Payments to Foreign Financial Institutions and Other Foreign Entities, 77 Fed. Reg. 9,022, 9,023 (proposed Feb. 15, 2012) (“[C]onsistent with the policies underlying [FATCA], the Treasury Department and the IRS remain committed to working cooperatively with foreign jurisdictions on multilateral efforts to improve transparency and information exchange on a global basis.”).

\(^{141}\) Id.

\(^{142}\) U.S. Treas. Dep’t Press Release, Joint Statement from the United States, France, Germany, Italy, Spain and the United Kingdom Regarding an Intergovernmental Approach to Improving International Tax Compliance and Implementing FATCA (Feb. 8, 2012) [hereinafter Joint Statement].

\(^{143}\) Id.

\(^{144}\) Id. at A.5.

\(^{145}\) Id. at A.3.

\(^{146}\) Id.

\(^{147}\) Routing addresses how FFIs must route information to residence country governments. Grinberg, supra note 51, at 17.

\(^{148}\) Joint Statement, supra note 142, at B.2.b. This routing system is similar the EU Savings Directive approach. Grinberg, supra note 51, at 18–21 (describing the EU approach for automatic information reporting).
FFIs in cooperative jurisdictions need only send information to one government, under whose laws they already operate, avoiding the possibility of FFIs attempting to comply with different reporting obligations to dozens of governments.149 “Reporting by financial institutions to the government of the jurisdiction in which they reside, followed by government-to-government exchange . . . conforms most closely to current global understandings regarding first-instance sovereign access to banking information.”150 This system avoids the conflict of laws issues associated with financial institutions reporting directly to foreign sovereigns.151

Compliant FFIs in non-partner countries, however, are required to report directly to the IRS under the original FATCA model.152 This routing method allows financial institutions that wish to participate to do so regardless of their government’s policy decisions.153 It also pressures non-participating jurisdictions to cooperate.154 Thus, the proposed intergovernmental approach’s bifurcated routing method incentivizes cooperation, reduces reporting burdens for FFIs, and mitigates sovereignty concerns.

The Joint Statement also expresses a commitment “to working with other FATCA partners, the OECD, and where appropriate the EU, on . . . the development of reporting and due diligence standards.” Development of uniform standards for reporting and due diligence would mitigate the burden on FFIs and streamline the process of automatic information reporting.155

The proposed intergovernmental approach is a step in the right direction, as it addresses many of the concerns associated

149. Grinberg, supra note 51, at 57.
150. Id.
151. Id.; see also supra Part IV.B (discussing conflict of laws issues associated with FATCA’s unilateral approach).
152. See supra Part III.B.2 (discussing FATCA reporting requirements).
153. Grinberg, supra note 51, at 57.
154. Id.
155. See FATCA Hearing, supra note 10, at 17 (“It is fundamentally important to achieve consistent standards of transparency that support compliance without overly burdening the efficiency of cross border portfolio investment flows.”) (statement of William J. Wilkins, Chief Counsel, Internal Revenue Serv.).
with FATCA while preserving the goal of the legislation. In addition to the five countries represented in the Joint Statement, others have expressed interest in participating.\textsuperscript{156} Broader participation would improve effectiveness of a multilateral automatic information reporting system and increase pressure on resisting jurisdictions.\textsuperscript{157} Eventually, noncompliance may become unsustainable.\textsuperscript{158}

\textbf{D. Room for Compromise: Citizenship-Based Taxation}

Despite its commitment to a collaborative FATCA regime, the proposed intergovernmental approach ignores a significant underlying inconsistency: basis for taxation. Whereas most countries impose taxes on resident and source income, the United States also taxes nonresident citizens. This inconsistency provides an opportunity to reevaluate the policy of taxation based solely on citizenship\textsuperscript{159} and, this Note argues, to terminate it.

First, strong non-FATCA based arguments exist in favor of eliminating citizenship-based taxation.\textsuperscript{160} The United States is the only country in the world to base worldwide taxation solely on citizenship.\textsuperscript{161} This policy dates back to the Civil War,\textsuperscript{162} and is

\begin{itemize}
\item \textsuperscript{157} Grinberg, \textit{supra} note 51, at 61.
\item \textsuperscript{158} Id.
\item \textsuperscript{159} See \textit{id.} at 58 n.205 (“An eventual multilateral system would be unlikely to retain FATCA’s concern with citizenship in addition to residence.”).
\item \textsuperscript{161} Id. at 1. “Eritrea is sometimes mentioned as another one, but it is unclear that it actually taxes nonresident citizens.” \textit{Id.}
\item \textsuperscript{162} Michael S. Kirsch, \textit{Taxing Citizens in a Global Economy}, 82 N.Y.U. L.
A MULTILATERAL AUTOMATIC REPORTING REGIME  1711

protected via a “saving clause” in U.S. income tax treaties.163 Historic U.S. justifications for taxing nonresident citizens, including deterring draft-dodging and flight of wealthy Americans, no longer apply.164 Similarly, arguments in favor of taxing these individuals are weak,165 especially in light of existing alternative bases for taxation.166

Second, FATCA imposes substantial burdens on U.S. citizens living abroad in the form of complex reporting requirements167 and, in some circumstances, barriers to obtaining a foreign bank account, insurance, or pension.168 Under FATCA, FBAR, and other existing reporting requirements, inadvertent noncompliance may result in steep civil and criminal penalties that are often disproportionately high in comparison to the amount of tax involved.169 Abandoning taxation of nonresident citizens could lead to significant simplification and reduction of administrative costs, which likely exceeds the revenue collected solely on the basis of citizenship.170

Third, elimination of citizenship-based taxation would not impair FATCA’s goals of tracking down tax evaders and raising revenue. FATCA was designed to fight offshore tax evasion by “bad actors” whose primary reason for establishing and maintaining unreported overseas accounts was to hide income

---

163. See U.S. Model Treaty, supra note 41, Art. 1 para. 4. (“[T]his convention shall not affect the taxation by a Contracting State of its residents . . . and its citizens.”).

164. See Avi-Yonah, supra note 160, at 1–2 (noting that this rule was created at a time when the income tax applied only to the rich and when some of the rich moved overseas to avoid the draft).

165. See id. at 6–10 (addressing the benefits, ability-to-pay, and administrability justifications for citizenship-based taxation).

166. The United States also imposes income tax on residents and U.S. source income.

167. See supra notes 79–80 and accompanying text (discussing FATCA reporting requirements); see also TAXPAYER ADVOCATE SERV., supra note 23, at 132 (discussing the overwhelming complexity and administrative detail of international reporting requirements).

168. Supra notes 103–04 and accompanying text.

169. See TAXPAYER ADVOCATE SERV., supra note 23, at 133–34 (listing potentially applicable civil and criminal penalties relating to individual U.S. taxpayers with foreign assets).

and avoid paying U.S. taxes they legally owe.\textsuperscript{171} In contrast, the estimated five to seven million\textsuperscript{172} U.S. citizens living abroad generally fall into a category of “benign actors”\textsuperscript{173} whose primary reasons for establishing and maintaining overseas accounts are unrelated to tax. Nonresident citizens include a wide range of individuals, from those who choose or are assigned to live overseas due to the opportunities of globalization to “accidental citizens” who were merely born in America and left the country at a young age.\textsuperscript{174} The average nonresident citizen holds foreign assets, including bank accounts, retirement funds, insurance plans, and investments, that are necessary for living and working in his country of residence. Similarly, due to international income exclusions and credits, citizens living abroad have, at most, a \textit{de minimis} tax liability.\textsuperscript{175} Despite this inconsistency, FATCA poses serious problems for U.S. citizens living abroad.\textsuperscript{176}

Finally, termination of citizenship-based taxation would facilitate an intergovernmental approach to automatic information reporting.\textsuperscript{177} Requiring financial institutions to identify both a taxpayer’s residence and citizenship, as is currently the case, doubles the amount of work required for compliance with FATCA. Unlike other FATCA requirements under the intergovernmental approach, identifying a taxpayer’s citizenship would only benefit the United States.\textsuperscript{178} Conversely, the ability to apply a single standard for identifying taxpayers

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{171} See supra Part II.A (discussing FATCA’s goals); see also TAXPAYER ADVOCATE SERV., \textit{supra} note 23, at 194 (noting that international information reporting penalties, FBAR, and FATCA are designed to combat deliberate offshore tax evasion).
\item \textsuperscript{172} See TAXPAYER ADVOCATE SERV., \textit{supra} note 23, at 192 n.7 (citing estimates that the number of U.S. citizens residing abroad is between five to seven million, not including U.S. troops).
\item \textsuperscript{173} See id. at 194 (listing examples of “benign actors”).
\item \textsuperscript{174} Avi-Yonah, \textit{supra} note 160, at 5.
\item \textsuperscript{175} See TAXPAYER ADVOCATE SERV., \textit{supra} note 23, at 194 n.16 (noting that, in the 2009 tax year, only about 9\% of international taxpayers had a U.S. tax liability after claiming the foreign earned income exclusion and applying the foreign tax credit).
\item \textsuperscript{176} See supra Part III.A (discussing the implications of FATCA for American citizens living abroad).
\item \textsuperscript{177} \textit{Supra} note 159 and accompanying text.
\item \textsuperscript{178} See \textit{supra} note 161 and accompanying text (noting that the United States is the only country that uses citizenship as a basis for taxation).
\end{enumerate}
\end{footnotesize}
A MULTILATERAL AUTOMATIC REPORTING REGIME

Based on residence would improve efficiency and reduce the burden on financial institutions to the benefit of all FATCA partners. Further, this compromise may induce other countries to participate in the proposed intergovernmental approach. This may be especially true of countries, like Canada, that are home to a significant number of U.S. citizens.179

In sum, taxation of nonresident citizens is inconsistent with global norms, creates administrative inefficiencies, and impairs development of a multilateral FATCA regime. For these reasons, the United States should abandon the policy of citizenship-based taxation.

VI. Conclusion

Since its enactment, strong criticisms have been raised about FATCA’s approach to international tax enforcement and its potential unintended consequences. In response, the U.S. government has expressed a willingness to adopt an intergovernmental approach to FATCA implementation. This Note argues that the type of international collaboration envisioned in the Joint Statement is essential to successful FATCA implementation. It also asserts that a multilateral automatic information reporting regime supports the policy goals of FATCA and mitigates the concerns associated with a unilateral approach. Finally, this Note argues that the United States should abandon its policy of citizenship-based taxation in order to facilitate a multilateral automatic information reporting regime.

The Virtual Water Cooler and the NLRB: Concerted Activity in the Age of Facebook

Lauren K. Neal*

Table of Contents

I. Introduction ................................................................. 1716

II. The Development of the Present Concerted Activity Standard .......................................................... 1717
   A. The Statutory Language ......................................... 1717
   B. The Board’s Interpretation of the Statutory Language ................................................................. 1719

III. The Application of the Meyers I Standard in the Social Media Context ..................................................... 1721
   A. The First NLRB Report ........................................... 1725
      1. When Concerted Activity Was Present ............. 1726
      2. When Concerted Activity Was Not Present ..... 1727
      3. What the Cases in the First Report Tell Us ................................................................. 1729
   B. The Second NLRB Report ....................................... 1732
      1. When Concerted Activity Was Present ............. 1732
      2. When Concerted Activity Was Not Present ...... 1735
      3. What the Cases in the Second Report Tell Us ................................................................. 1738
   C. The ALJ Cases ................................................................. 1740
      1. Hispanics United of Buffalo, Inc. ...................... 1740
      2. Karl Knauz Motors, Inc. ................................. 1743
      3. Triple Play Sports Bar & Grille ...................... 1745
      4. What the ALJ Cases Tell Us ................................................. 1747

* Candidate for J.D., Washington and Lee University School of Law 2013; B.A., Franklin and Marshall College 2010. I would like to thank Professor Brian Murchison for being an inspiring mentor who is, above all, a wonderful human being. I am also forever grateful to my parents, Derek and Evelyn Neal, for their never-ending love and support.
I. Introduction

The National Labor Relations Act (NLRA or Act)\(^1\) gives employees the right “to engage in . . . concerted activities for the purpose of collective bargaining or other mutual aid or protection.”\(^2\) When President Roosevelt signed the Act into law in 1935, few could have imagined the new contexts in which “concerted activities” would arise. The Act’s drafters envisioned a workplace in which employees communicated with each other in person. Employee communication is no longer so limited, however. Facebook and other social networking websites have altered this traditional water cooler model, creating new spaces in which employees interact.

With these new spaces come new questions. Chief among them is how the National Labor Relations Board (NLRB or Board)\(^3\) should apply the Act’s concerted activity provision in

---


3. The Board is an independent agency created by the NLRA. \textit{Id.} § 153; \textit{see also NLRB, Who We Are}, http://www.nlrb.gov/who-we-are (last visited Sept. 27, 2012) (“The National Labor Relations Board is an independent federal agency that protects the rights of private sector employees to join together, with or without a union, to improve their wages and working conditions.”) (on file with the Washington and Lee Law Review). When this Note refers to the Board in Part I and Parts III–VI, it is referring to the agency as a whole. It is not referring to the five-member, quasi-judicial body within the agency, unless
cases involving a virtual water cooler. Do existing standards for defining concerted activity make sense when applied in the social media context, or should the NLRB alter them to better comport with the realities of contemporary interaction? Facebook firing cases—cases in which an employer fires an employee because of a Facebook posting—provide insight.

Adapting the present concerted activity standard for application in Facebook firing cases has presented the Board with numerous challenges. The source of these challenges is the very nature of social media; the forum itself makes it more difficult to distinguish reasonably between activity that is concerted and activity that is not. Furthermore, the forum alters the calculus of interest balancing in which the Board must engage to effectuate the Act’s purposes. While the Board has attempted to clarify how social media will fit into existing doctrine, uncertainty remains.

This Note examines why uncertainty remains and offers three temporally based approaches to remedy the uncertainty. Part II of this Note explores the Board’s present interpretation of concerted activity. Part III details how entities within the Board have applied the standard in Facebook firing cases. Part IV examines why the application of the standard in Facebook firing cases is problematic. Part V suggests three approaches that the Board should consider: promulgating a model social media policy; identifying factors that tend to indicate concertedness in the social media context; and applying a more stringent loss-of-protection standard to cases involving social media. These approaches are not mutually exclusive and recognize both the unique attributes of the social media context and the interests that the context implicates.

II. The Development of the Present Concerted Activity Standard

A. The Statutory Language

As is often the case, the heart of the issue lies in the ambiguity of statutory language. Section 7 of the NLRA states:

“Employees shall have the right to self-organization, to form, join, or assist labor organizations, to bargain collectively through representatives of their own choosing, and to engage in other concerted activities for the purpose of collective bargaining or other mutual aid or protection.” 4 In the list of employee rights, the right “to engage in other concerted activities” provides the most room for interpretation and, consequently, confusion. This confusion is problematic because understanding what constitutes concerted activity is a precondition for employees and employers to understand the scope of their rights and obligations. 5

Section 7’s language and the placement of the concerted activity phrase within the Section provide some insight into the meaning of concerted activity. First, the use of “other” in the phrase “and to engage in other concerted activities” suggests that the activities enumerated before the phrase—self-organizing; forming, joining, or assisting labor organizations; and bargaining collectively through representatives—are themselves examples of concerted activities. 6 Second, it appears that Congress intended for some unenumerated activities to fall within Section 7. Otherwise, the phrase would be unnecessary. Thus, Section 7’s language and structure indicate that the activities fitting within the phrase’s scope are similar to, yet different from, those activities that Congress enumerated specifically. Because neither Section 7, nor its legislative history, defines concerted activity explicitly, 7 the Board has interpreted the Section and divined its own definition, while keeping in mind the Act’s purposes. 8


5. Section 8 of the NLRA makes the rights set out in Section 7 enforceable: “It shall be an unfair labor practice for an employer . . . to interfere with, restrain, or coerce employees in the exercise of the rights guaranteed in section 157 . . . .” 29 U.S.C. § 158(a)(1). Section 8 of the NLRA corresponds to 29 U.S.C. § 158.

6. See NLRB v. City Disposal Sys., Inc., 465 U.S. 822, 831 n.8 (1984) (“Section 7 lists . . . activities initially and concludes the list with the phrase ‘other concerted activities,’ thereby indicating that the enumerated activities are deemed to be ‘concerted.’”).

7. See id. at 830 (stating that the Act does not define concerted activity); see also Meyers Indus., Inc. (Meyers I), 268 N.L.R.B. 493, 493 (1984) (stating that the Act’s legislative history does not contain a definition).

8. The Act’s purposes are set forth in its preamble and include “encouraging . . . collective bargaining and . . . protecting the exercise by
B. The Board’s Interpretation of the Statutory Language

The present definition of concerted activity originates from Meyers Industries, Inc. (Meyers I). In Meyers I, the Board adopted what it described as “the ‘objective’ standard of concerted activity.” The Board contrasted the objective standard with what it deemed the “per se standard of concerted activity,” the standard the Board employed before Meyers I. The Board enunciated the per se standard ten years earlier in Alleluia Cushion Co. It applied to cases in which “an employee [spoke] up and [sought] to enforce statutory provisions relating to occupational safety designed for the benefit of all employees.” In those circumstances, the Board stated: “[I]n the absence of any evidence that fellow employees disavow such representation, we will find an implied consent thereto and deem such activity to be concerted.” Because Alleluia did not require an outward manifestation of group involvement or support, the Board later characterized this as the per se standard of concerted activity.

Meyers I firmly rejected this standard, defining concerted activity as follows: “[T]o find an employee’s activity to be ‘concerted,’ we shall require that it be engaged in with or on the behalf of the other employees, and not solely by and on behalf of representatives of their own choosing, for the purpose of negotiating the terms and conditions of their employment or other mutual aid or protection.”

10. Id. at 496.
11. See id. at 493–97 (describing both standards of concerted activity).
13. Id. at 1000.
14. Id.
15. See Meyers I, 268 N.L.R.B. 493, 495 (1984) (“Under the Alleluia analytical framework, the Board questioned whether the purpose of the activity was one it wished to protect and . . . deemed the activity ‘concerted,’ without regard to its form.”).
16. See id. at 496 (“[W]e hold that the concept of concerted activity first enunciated in Alleluia does not comport with the principles inherent in Section 7 of the Act.”). The Board then overruled Alleluia. Id. Rather than enunciating an entirely new standard, the Board asserted that it was simply returning to the pre-Alleluia standard. Id. at 496–97.
of the employee himself.”17 Determining whether an activity meets this standard requires a highly factual inquiry.18 Applying this standard to Meyers I’s facts, the Board found that a truck driver was not engaged in concerted activity when he (1) complained to his employer and state authorities that his truck was unsafe; (2) contacted the Tennessee Public Service Commission to arrange a vehicle inspection after getting into an accident; and (3) refused to drive the truck after the accident.19 According to the Board, rather than engaging in concerted activity, the employee was acting alone.20

In Prill v. NLRB,21 the D.C. Circuit addressed Meyers I’s definition of concerted activity.22 Although the court acknowledged the deference due to the Board,23 it found that “the Board act[ed] pursuant to an erroneous view of the law . . . when it decided that its new definition of ‘concerted activities’ was mandated by the NLRA.”24 In addition, the court found that the Board misread some of its own precedent.25 Consequently, the court remanded the case, instructing the Board to reconsider its Meyers I definition because it rested “on a faulty legal premise and [was] without adequate rationale.”26

17. Id. at 497. To establish a violation of Section 8 of the NLRA, an employee must also show that “the employer knew of the concerted nature of the employee’s activity, the concerted activity was protected by the Act, and the adverse employment action . . . was motivated by the employee’s protected concerted activity.” Id.
18. See id. (“[T]he question of whether an employee engaged in concerted activity is, at its heart, a factual one . . . .”)
19. Id. at 498.
20. See id. (“Prill [the employee] acted solely on his own behalf.”).
22. Id. at 948–50.
23. See id. at 942 (“The Board has been granted broad authority to construe the NLRA in light of its expertise.”). In particular, the D.C. Circuit acknowledged the Supreme Court’s decision in NLRB v. City Disposal Systems, Inc., in which the Court stated that the Board’s interpretation of the Act, if reasonable, “is entitled to considerable deference.” NLRB v. City Disposal Sys., Inc., 465 U.S. 822, 829 (1984).
24. Prill, 755 F.2d at 942. Thus, the D.C. Circuit asserted that the Board “fail[ed] to exercise the discretion delegated to it by Congress.” Id.
25. See id. at 953–56 (challenging the Board’s assertion that the Meyers I definition represented a return to the pre-Alleluia definition).
26. Id. at 942. The court stressed that it was not suggesting that the Meyers I definition was incorrect. See id. (“We express no opinion as to the
Nevertheless, in *Meyers Industries, Inc. (Meyers II)*, the Board reaffirmed the *Meyers I* standard. While addressing the D.C. Circuit’s concerns, the Board clarified the standard, offering more insight into what constitutes concerted activity. First, the Board stressed that the standard “requires some linkage to group action.” Notably, however, the Board emphasized that an individual employee’s act can constitute concerted activity. While explaining when individual action falls within Section 7, the Board stated: “[O]ur definition of concerted activity in *Meyers I* encompasses those circumstances where individual employees seek to initiate or to induce or to prepare for group action, as well as individual employees bringing truly group complaints to the attention of management.” The Board applied this clarified standard to *Meyers I*s facts, once again finding that the employee did not engage in concerted activity but instead “acted alone and without an intent to enlist the support of other employees.” Therefore, the Board affirmed its dismissal of the employee’s complaint.

### III. The Application of the Meyers I Standard in the Social Media Context

The advancement of technology and the growth of employee use of technology have presented—and continue to present—numerous challenges. Technology has required the NLRB to correct test of ‘concerted activities’ . . . ”.


28. See id. at 889 (“The Board has reconsidered this case . . . and has decided to adhere to the *Meyers I* definition of concerted activity as a reasonable construction of Section 7 of the Act.”).

29. Id. at 884.

30. See id. at 885 (“There is nothing in the *Meyers I* definition that states that conduct engaged in by a single employee at one point in time can never constitute concerted activity within the meaning of Section 7.”).

31. Id. at 887. In addition, the Board reaffirmed its so-called “Interboro doctrine”: an individual employee who reasonably and honestly invokes a collective bargaining right is engaged in concerted activity. *Id.* at 884–85; see also NLRB v. City Disposal Sys., Inc., 465 U.S. 822, 841 (1984) (approving of the *Interboro* doctrine as a reasonable interpretation of the Act).


33. Id. at 889.
determine how to adapt the NLRA to a modern reality in which employee communication is fundamentally different than it was seventy-five years ago.\textsuperscript{34} This adaption process has involved a difficult task: attempting to “maintain[] stability in the law while simultaneously allowing for flexibility to address these new developments.”\textsuperscript{35} Most recently, social media has presented the Board with an opportunity to consider these competing values—stability and flexibility—for the purpose of balancing the competing interests at stake.\textsuperscript{36}

When confronting new issues, administrative agencies like the NLRB have two tools at their disposal: rulemaking and adjudication. Because agencies often lack ready answers to the questions that emerging issues present, adjudication provides a vehicle to explore those issues, formulate an approach, and develop a rationale. In \textit{SEC v. Chenery Corp. (Chenery II)},\textsuperscript{37} the Supreme Court characterized this process as a necessary and valuable supplement to agencies’ rulemaking authority.\textsuperscript{38} When explaining why an agency could reasonably prefer to address an issue through adjudication, the Court stated: “Not every principle essential to the effective administration of a statute can or should be cast immediately into the mold of a general rule. Some principles must await their own development, while others must be adjusted to meet particular, unforeseeable situations.”\textsuperscript{39} Thus far, the NLRB has chosen to confront the issues presented by social media through adjudication.

\textsuperscript{34} See Martin H. Malin & Henry H. Perritt, Jr., \textit{The National Labor Relations Act in Cyberspace: Union Organizing in Electronic Workplaces}, 49 U. KAN. L. REV. 1, 62 (2000) (“For six and one-half decades, the National Labor Relations Board and the courts have been developing and refining doctrine under the National Labor Relations Act in the context of traditional physically defined workplaces.”).

\textsuperscript{35} Gwynne A. Wilcox, \textit{Section 7 Rights of Employees and Union Access to Employees: Cyber Organizing}, 16 LAB. L. 253, 253 (2000).

\textsuperscript{36} See Malin & Perritt, supra note 34, at 62 (“Adapting the NLRA to electronic workplaces will continue a process of balancing employee rights to engage in concerted activities against employer property and entrepreneurial rights.”).

\textsuperscript{37} SEC v. Chenery Corp. (Chenery II), 332 U.S. 194 (1947).

\textsuperscript{38} See id. at 203 (“There is . . . a very definite place for the case-by-case evolution of statutory standards.”).

\textsuperscript{39} Id. at 202.
By late 2011, the Board had reviewed approximately 130 cases involving social media, all existing at varying stages of development.\textsuperscript{40} The Board's involvement in Facebook firing cases, in particular, began on October 27, 2010, when Region 34 of the NLRB issued a complaint against a Connecticut ambulance company in \textit{American Medical Response of Connecticut, Inc.}\textsuperscript{41} The complaint alleged, in part, that a company employee who criticized her supervisor on Facebook engaged in concerted activity, and that the company terminated her employment because of her Facebook postings.\textsuperscript{42} Because this case, and many subsequent cases, settled prior to adjudication,\textsuperscript{43} guidance on how the Board will treat cases involving the intersection of labor law and social media is lacking. In addition, guidance is lacking because many of these cases have not yet progressed past the initial stages of development.\textsuperscript{44}

\textsuperscript{40} U.S. Chamber of Commerce, \textit{A Survey of Social Media Issues Before the NLRB} (2011), www.uschamber.com/reports/survey-social-media-issues-nlrb (last visited Sept. 27, 2012) (on file with the Washington and Lee Law Review). This number includes decisions by the five-member board, ALJ decisions, settlement agreements, complaints, memoranda, and charges that contain social media components. \textit{Id.}\textsuperscript{2} It includes a broader range of factual circumstances than is the focus of this Note: concerted activity in the context of Facebook.

\textsuperscript{41} Complaint & Notice of Hearing, Am. Med. Response of Conn., Inc., N.L.R.B. No. 34-CA-12576 (Region 34 Oct. 27, 2010), http://documents.jdsupra.com/daa37777-f93f-4fe0-be1f-82c65d02f3ac3.pdf. Typically, a case proceeds through the following stages: First, the employee or the employee's representative files a charge—"a one-page form alleging that an employer or union has committed an unfair labor practice"—with one of the Board's regional offices. \textit{U.S. Chamber of Commerce, supra} note 40, at 3. Second, the regional office investigates and "makes a determination as to whether the charge has merit." \textit{Id.}\textsuperscript{4} If the regional office finds that the charge has merit, it issues a complaint, and the parties either settle the case, or the case proceeds to adjudication. \textit{Id.} The NLRB has thirty-two regional offices. NLRB, \textit{Regional Offices}, http://www.nlrb.gov/who-we-are/regional-offices (last visited Sept. 27, 2012) (on file with the Washington and Lee Law Review).


\textsuperscript{43} \textit{U.S. Chamber of Commerce, supra} note 40, at 2; \textit{see also} Press Release, NLRB, \textit{Settlement Reached in Case Involving Discharge for Facebook Comments} (Feb. 8, 2011), http://www.nlrb.gov/news/settlement-reached-case-involving-discharge-facebook-comments (last visited Sept. 27, 2012) ("A settlement has been reached in a case involving the discharge of a Connecticut ambulance service employee for posting negative comments about a supervisor on her Facebook page."). (on file with the Washington and Lee Law Review).

\textsuperscript{44} \textit{U.S. Chamber of Commerce, supra} note 40, at 1.
Nevertheless, some guidance exists. Of particular importance are two reports: one issued by the Board’s Acting General Counsel on August 18, 2011 (First NLRB Report or First Report) and one issued by the Board’s Acting General Counsel on January 24, 2012 (Second NLRB Report or Second Report). The First Report “presents recent case developments arising in the context of today’s social media,” including “issues concerning the protected and/or concerted nature of employees’ Facebook and Twitter postings.” After the Acting General Counsel issued the First Report, Administrative Law Judges (ALJs) heard the first three Facebook firing cases: Hispanics United of Buffalo, Inc., Karl Knauz Motors, Inc., and Triple Play Sports Bar & Grille.


47. First NLRB Report, supra note 45, at 2.

48. ALJs are Article I judges who are similar to “trial court judges hearing a case without a jury.” NLRB, Administrative Law Judge Decisions, http://www.nlrb.gov/cases-decisions/case-decisions/administrative-law-judge-decisions (last visited Sept. 27, 2012) (on file with the Washington and Lee Law Review). After a regional office issues a complaint, an ALJ “hears the case and issues a decision and recommendation order.” Id.


51. Triple Play Sports Bar & Grille, N.L.R.B. No. 34-CA-12915, 2012 WL 76862 (Div. of Judges Jan. 3, 2012). See infra Part III.C.3. Although the respondent in this case is Three D, LLC d/b/a Triple Play Sports Bar & Grille, this Note will refer to the case as “Triple Play Sports Bar & Grille” or “Triple Play” because it is commonly referred to by those names.
Following these cases, the Acting General Counsel issued the Second Report, which discusses cases containing “emerging issues in the context of social media.”52 These two reports, along with the ALJ cases, provide insight into how the Board will treat future Facebook firing cases.53

A. The First NLRB Report

The First NLRB Report contains summaries of fourteen cases involving social media, nine of which pertain to the concerted nature of online postings.54 In addition, the First Report discusses whether the General Counsel’s Division of Advice (Division of Advice) found each case meritorious.55 Its purpose, as stated by

52. SECOND NLRB REPORT, supra note 46, at 2.
53. On May 30, 2012, the Acting General Counsel issued a third report (Third NLRB Report or Third Report) discussing social media cases. See NLRB OFFICE OF THE GEN. COUNCIL, MEMORANDUM OM 12-59, REPORT OF THE ACTING GENERAL COUNSEL CONCERNING SOCIAL MEDIA CASES (2012) [hereinafter THIRD NLRB REPORT], www.nlrb.gov/publications/operations-management-memos (last visited Sept. 27, 2012) (on file with the Washington and Lee Law Review). Unlike the first two reports, the Third NLRB Report focuses solely on social media policies and rules. Id. at 2. Although this Note touches on the Board’s role in providing guidance in this area, the Board’s analysis of social media policies and rules is not the focus of this Note. See Part V.A (discussing the connection between providing guidance in the area of social media policies and rules and solving the concerted activity problem). Thus, this Note does not discuss the Third Report in-depth.
54. FIRST NLRB REPORT, supra note 45 (citing Triple Play Sports Bar & Grille, N.L.R.B. No. 34-CA-12915 (Div. of Advice); Karl Knauz Motors, Inc., N.L.R.B. No. 13-CA-46452 (Div. of Advice); Hispanics United of Buffalo, Inc., N.L.R.B. No. 3-CA-27872 (Div. of Advice); Martin House, N.L.R.B. No. 34-CA-12950, 2011 WL 3223853 (Div. of Advice July 19, 2011); Wal-Mart, N.L.R.B. No. 17-CA-25030, 2011 WL 3223852 (Div. of Advice July 19, 2011); JT’s Porch Saloon & Eatery, Ltd., N.L.R.B. No. 13-CA-46689, 2011 WL 2960994 (Div. of Advice July 1, 2011); Rural Metro, N.L.R.B. No. 25-CA-31802, 2011 WL 2960970 (Div. of Advice June 29, 2011); Lee Enters., Inc. N.L.R.B. No. 28-CA-23267 (Div. of Advice Apr. 21, 2011); Am. Med. Response of Conn., Inc., N.L.R.B. No. 34-CA-12576 (Div. of Advice Oct. 5, 2010)). One of the nine cases—Lee Enterprises, Inc.—involves the concerted nature of Twitter postings and is included in the discussion because it implicates similar issues. The five other cases involve matters beyond the scope of this Note. Furthermore, the First Report does not identify explicitly the cases that it discusses. This was determined independently.
55. Id. The First Report states that the cases it discusses “were decided upon a request for advice from a Regional Director.” Id. at 2. When a regional
the Board’s Acting General Counsel, Lafe Solomon, is to “encourage compliance with the Act and cooperation with Agency personnel” by “keep[ing] the labor-management community fully aware of the activities of [his] office.”

1. When Concerted Activity Was Present

The Division of Advice found that concerted activity existed in four of the nine cases. Three of these four cases—Hispanics United, Karl Knauz Motors, and Triple Play—went on to become the first Facebook firing cases heard by an ALJ. The other case—American Medical—settled. Thus, the First Report and the Advice Memorandum it summarizes contain the only existing analysis of American Medical.

In American Medical, the Division of Advice found that an employee engaged in concerted activity when she posted negative comments about her supervisor on her Facebook page. The employee posted the comments after her employer denied her request for a union representative to assist her in the preparation of a written incident report. Coworkers responded to the postings, and eventually the employee was terminated. Because the employee “discuss[ed] supervisory actions with coworkers in her Facebook post,” the Division of Advice found that her activity was both protected and concerted.
2. When Concerted Activity Was Not Present

The Division of Advice found that there was no concerted activity in five of the nine cases. For example, in *Lee Enterprises, Inc.*, the Division of Advice found that an employee’s Twitter postings did not constitute concerted activity. The case involved a newspaper reporter who created a Twitter account after his employer, a newspaper company, encouraged him to do so. After creating the account, he posted tweets criticizing his copy editors and a local television station. In addition, he posted tweets concerning local homicides and others containing sexual content. In finding that his eventual termination did not violate the Act, the Division of Advice stated that the employee’s “conduct was not protected and concerted: it did not relate to the terms and conditions of his employment or seek to involve other employees in issues related to employment.”

Likewise, in *JT’s Porch Saloon & Eatery, Ltd.*, the Division of Advice found that an employee’s Facebook posting did not constitute concerted activity. In response to a question posed by his stepsister, the employee, a bartender, expressed frustration with his employer’s tipping policy. Although the employee had previously discussed his concerns about the tipping policy with a coworker, the Division of Advice found that the posting was not concerted because the employee “did not discuss the posting with his coworkers, and none of them responded to the posting.”

---

64. *Id.* at 12–14 (citing *Lee Enters., Inc.*, N.L.R.B. No. 28-CA-23267 (Div. of Advice Apr. 21, 2011)).
65. *Id.* at 13.
66. *See id.* at 12 (“[T]he [e]mployer encouraged employees . . . to use social media to get news stories out . . . .”).
67. *Id.* at 12–13.
68. *Id.* at 13.
69. *Id.*
70. *Id.* at 14–15 (citing *JT’s Porch Saloon & Eatery, Ltd.*, N.L.R.B. No. 13-CA-46689, 2011 WL 2960964 (Div. of Advice July 7, 2011)).
71. *Id.*
72. *Id.* at 14. The employee’s sister responded in agreement, saying that the policy “sucked.” *Id.*
73. *Id.* at 15 (emphasis added).
In Rural Metro, the Division of Advice found that a person’s Facebook posting on a United States Senator’s Facebook page did not constitute concerted activity. After the Senator made a Facebook posting concerning federal grants to fire departments, the person, who worked for a company that contracted with fire departments, responded by complaining about her wages and the way that the state handled emergency medical services. Although the employee “had discussed wages with other employees after [her] [e]mployer had announced a wage cap,” the Division of Advice found that the posting was not concerted, emphasizing the fact that “she did not discuss her posting with any other employee.”

In Martin House, the Division of Advice found that an employee of a nonprofit facility for homeless people was not engaged in concerted activity when she posted comments about the facility’s mentally disabled clients on her Facebook page. Because the employee “did not discuss her Facebook posts with any of her fellow employees” and “none of her coworkers responded to the posts,” the activity was not concerted.

Finally, in Wal-Mart, the Division of Advice found that a Wal-Mart employee who posted comments critical of the store’s management was not engaged in concerted activity, even though several coworkers responded to the postings. The Division of Advice emphasized that the postings were “expression[s] of an individual gripe.” The coworkers’ responses did not make the

74. Id. at 15–16 (citing Rural Metro, N.L.R.B. No. 25-CA-31802, 2011 WL 2960970 (Div. of Advice June 29, 2011)).
75. Id. at 16.
76. Id. at 15.
77. Id. at 15–16 (emphasis added).
78. Id. at 16–17 (citing Martin House, N.L.R.B. No. 34-CA-12950, 2011 WL 3223853 (Div. of Advice July 19, 2011)).
79. Id. For example, the employee “stated that it was spooky being alone overnight in a mental institution, that one client was cracking her up, and that [she] did not know whether the client was laughing at her, with her, or at the client’s own voices.” Id.
80. Id. (emphasis added).
81. Id. 17–18 (citing Wal-Mart, N.L.R.B. No. 17-CA-25030, 2011 WL 3223852 (Div. of Advice July 19, 2011)).
82. Id.
83. Id.
Facebook activity concerted because the responses “merely indicated that [the coworkers] had found the employee’s first posting humorous, asked why the employee was so ‘wound up,’ or offered emotional support.”

3. What the Cases in the First Report Tell Us

Although the General Counsel did not identify explicitly the factors that the Division of Advice considered when evaluating concertedness in the social media context, the Division of Advice considered certain factors regularly in its analysis. Such factors include: (1) whether the posting grew out of prior non-Facebook group activity,85 (2) whether the posting contemplated future non-Facebook group activity,86 (3) whether coworkers responded to the posting,87 (4) whether the employee who made the posting discussed the posting itself with coworkers,88 (5) the intent of the

84. Id.


86. See Martin House, 2011 WL 3223853, at *2 (“[T]he Charging Party was not seeking to induce or prepare for group action . . . .”); Wal-Mart, 2011 WL 3223852, at *2 (“They [the postings] contain no language suggesting the Charging Party sought to initiate or induce coworkers to engage in group action . . . .”).


88. See Martin House, 2011 WL 3223853, at *2 (“The Charging Party did not discuss her Facebook posts with any of her fellow employees . . . .”); JT’s Porch, 2011 WL 2960964, at *2 (“[H]e did not discuss his Facebook posting with any of his fellow employees either before or after he wrote it . . . .”); Rural Metro, N.L.R.B. No. 25-CA-31802, 2011 WL 2960970, at *2 (Div. of Advice June 29, 2011) (“The Charging Party did not discuss her Facebook posting with any other employee . . . either before or immediately thereafter.”).
employee who made the posting, as reflected by the language and context of the posting, and (6) how coworkers interpreted the posting, as reflected by their responses.

Factors one and two relate closely to the standard developed in Meyers I and its progeny: if individual action grows out of prior group action or prepares for future group action, it is likely concerted. One question unique to the social media context is whether the Facebook activity in question must be an outgrowth of, or preparation for, in-person group activity, or whether it can be an outgrowth of, or preparation for, additional online group activity. The Division of Advice’s analysis suggests that a lack of connection to in-person group activity will weigh against finding that Facebook activity is concerted.

As factors three and four indicate, the Division of Advice focused not only on the original posting and the person who made it, but also on whether coworkers became involved in the posting. The Division of Advice examined two forms of involvement: whether coworkers responded to the posting online and whether the posting itself became a topic of in-person conversation. Typically, the Division of Advice focused on the absence of these

89. See Martin House, 2011 WL 3223853, at *2 (finding the postings revealed the employee’s intent to “communicat[e] with her personal friends about what was happening on her shift”); Wal-Mart, 2011 WL 3223852, at *2 (finding the postings revealed the employee’s intent to “express ... his frustration regarding his individual dispute with the Assistant Manager”); JT’s Porch, 2011 WL 2960964, at *2 (finding the postings revealed the employee’s intent to “respond[] to a question from his step-sister about how his evening at work went”); Rural Metro, 2011 WL 2960970, at *2 (finding the postings revealed the employee’s intent “to make a public official aware of the state of emergency medical services in Indiana”).

90. See, e.g., Wal-Mart, 2011 WL 3223852, at *2 (finding that coworkers’ responses were reactions to an individual gripe rather than the beginning of group action). For example, the Division of Advice characterized a coworker’s “‘hang in there’-type comment” as suggesting that the coworker viewed the employee’s “postings to be a plea for emotional support.” Id.

91. See NLRB v. Mike Yurosek & Son, Inc., 53 F.3d 261, 265 (9th Cir. 1995) (stating that individual action may be concerted when it is a “logical outgrowth’ of prior concerted activity”); see also Meyers II, 281 N.L.R.B. 882, 887 (1986) (stating that “individual employees seek[ing] to initiate or to induce or to prepare for group action” may be involved in concerted activity).

92. See supra notes 85–86.

93. See supra notes 87–88. Discussion of a posting can occur within a conversation about common concerns, or it can provide a means of entering into a conversation about common concerns.
factors, which are objective in nature, when finding that Facebook activity was not concerted. In contrast, when the Division of Advice found that Facebook activity was concerted, it put less of an emphasis on whether coworkers discussed the posting itself with each other in person. Perhaps this is because in the cases in which the Division of Advice found concertedness, coworkers usually responded to the original posting by making postings of their own. Thus, their responses reflected knowledge of the original posting, and an examination of whether the original posting was a topic of in-person conversation became unnecessary. Even so, when finding concerted activity, the Division of Advice tended to address the significance of coworker responses only implicitly—by characterizing the Facebook activity as a “discussion” or “conversation” when coworkers responded—rather than stating clearly that coworker responses indicated concertedness.

Factors five and six represent an implicit focus in the Division of Advice’s analysis on the subjective intent of those involved in the Facebook activity. In each case, the Division of Advice analyzed what the person who made the posting was “seeking to” or “trying to” do. Likewise, it looked at coworkers’ responses to discern how they interpreted the posting: whether their responses reflected an intent to participate in group action or simply conveyed, for example, that they found the posting amusing.

94. See, e.g., Martin House, N.L.R.B. No. 34-CA-12950, 2011 WL 3223853, at *2 (Div. of Advice July 19, 2011) (finding that the Facebook activity was not concerted in part because “[t]he Charging Party did not discuss her Facebook posts with any of her fellow employees, and none of her coworkers responded to the posts”).


96. See, e.g., id. at 9 (characterizing the online activity as a “discussion” of supervisory actions).

97. See, e.g., Wal-Mart, N.L.R.B. No. 17-CA-25030, 2011 WL 3223852, at *2 (Div. of Advice July 19, 2011) (“They [the postings] contain no language suggesting the Charging Party sought to initiate or induce coworkers to engage in group action; rather they express only his frustration regarding his individual dispute . . . .”) (emphasis added).

98. See, e.g., id. (“Employee 1 merely indicated that he found Charging Party’s first Facebook posting humorous . . . .”).
Although it is possible to identify factors that the Division of Advice considered regularly, it is difficult to discern the weight that the Division of Advice attributed to each factor. The Division of Advice neither attributed a uniform weight to each factor nor indicated which factor(s) it considered more heavily than the others.

**B. The Second NLRB Report**

Like the First Report, the Second NLRB Report contains summaries of cases involving social media, ten of which include a discussion of the concerted nature of Facebook postings.99 Because, as the Board’s Acting General Counsel acknowledged, the issues raised by these cases “continue to be a ‘hot topic’ among practitioners, human resource professionals, the media, and the public,” the Second Report’s purpose is to continue “to provide guidance as this area of law develops.”100

1. *When Concerted Activity Was Present*

The Division of Advice found that concerted activity existed in five of the ten cases. For example, in Case A,101 the Division of Advice found that an employee engaged in concerted activity when she posted a status update on her Facebook page about her employer.102 The employee made the posting after her employer

---

99. *SECOND NLRB REPORT*, supra note 46 (citing Children’s Nat’l Med. Ctr., N.L.R.B. No. 05-CA-36658 (Div. of Advice Nov. 14, 2011); Frito-Lay, Inc., N.L.R.B. No. 36-CA-10882 (Div. of Advice Sept. 19, 2011); Buel, Inc., N.L.R.B. No. 11-CA-22936 (Div. of Advice July 28, 2011); *Case A*; *Case B*; *Case C*; *Case D*; *Case E*; *Case F*; *Case G*). The NLRB does not release all Advice Memoranda to the public. See NLRB, *Advice Memos* http://www.nlrb.gov/cases-decisions/advice-memos (last visited Sept. 27, 2012) (“Two categories of advice memoranda are released to the public: memoranda directing dismissal of the charge . . . and memoranda in closed cases that . . . are released in the General Counsel’s discretion.”) (on file with the Washington and Lee Law Review). Moreover, the Second Report does not contain case names. Unless the name of a case discussed in the Second Report was found independently, the case was assigned a letter, and this Note refers to it by the letter throughout.

100. *SECOND NLRB REPORT*, supra note 46, at 2.

101. *Id.* at 3–6 (citing *Case A*).

102. *Id.* at 5. “Using expletives, she stated the Employer had messed up and
transferred her to a less lucrative position. Coworkers and former coworkers responded, expressing support, anger, and frustration, and one former coworker suggested that the employees initiate a class action lawsuit against the employer. A few days after making the posting, the employee was terminated. In supporting its finding of concerted activity, the Division of Advice stated: “[T]he Charging Party’s initial Facebook statement, and the discussion it generated, . . . clearly fell within the Board’s definition of concerted activity, which encompasses employee initiation of group action through the discussion of complaints with fellow employees.”

In Case B, the Division of Advice found that an employee was engaged in concerted activity when she posted several comments on her Facebook page. The employee posted the first comment after a company manager made a sexist remark. A coworker who was with the employee when the manager made the remark responded to the posting. The employee later made a second series of postings after a coworker was fired, and the company’s president reprimanded her for getting involved in coworkers’ work-related problems. Subsequently, the employee was terminated—a termination that the Division of Advice concluded was due to her “engag[ement] in discussions with her coworkers about working conditions.”

Similarly, the Division of Advice found that an employee’s Facebook activity was concerted in Case C. After a coworker

103. Id.
104. Id.
105. Id.
106. Id. at 5. A Charging Party is an employee who files a charge against his or her employer, alleging that the employer committed an unfair labor practice.
107. Id. at 18–20 (citing Case B).
108. Id. at 20.
109. Id. at 18. The employee sent an email about the remark to her supervisor and a human resources assistant but did not receive responses. Id.
110. Id.
111. Id. at 19.
112. Id. at 20. The Division of Advice characterized the termination as a “pre-emptive strike because of the [e]mployer’s fear of what those discussions might lead to.” Id.
113. Id. at 20–22 (citing Case C).
was promoted, the employee had separate discussions with two other coworkers about her disagreement with the promotion.\footnote{Id. at 20.} She made a Facebook posting reflecting her frustration, and three coworkers responded in agreement.\footnote{Id. at 20–21.} The employer terminated the employee who made the posting, along with one of the coworkers who responded, and the employer disciplined the two other coworkers who responded.\footnote{Id. at 21.} The activity was concerted, the Division of Advice reasoned, because the employee’s posting “sparked a collective dialogue that elicited responses from three of her coworkers.”\footnote{Id. at 22.} The Division of Advice noted: “While the concerted actions expressed in the posts were of a preliminary nature . . . the movement toward concerted action was halted by the [e]mployer’s pre-emptive discharge and discipline of all the employees involved.”\footnote{Id.}

In \textit{Case D}, the Division of Advice found that an employee’s Facebook posting constituted concerted activity.\footnote{Id. at 22–25 (citing Case D).} Before making the posting, the employee participated in discussions with coworkers about a superior’s negative attitude and brought the concern to management’s attention.\footnote{Id. at 22.} A coworker posted on her Facebook page “that there had been so much drama” at work, and an online conversation with coworkers ensued.\footnote{Id. at 23.} The employee—who was eventually discharged—responded by stating that “she hated [her workplace] and couldn’t wait to get out of there” because of the negative work environment.\footnote{Id. at 23.} The Division of Advice concluded that the Facebook activity was concerted “because it was a continuation of the earlier group action that included complaints to management . . . and because it was part of a discussion of employees’ shared concerns.”\footnote{Id.}
Finally, in Case E, the Division of Advice found that an employee's Facebook postings made during a seven-month period constituted concerted activity. Tension between the employee, a nurse, and his employer, a hospital, arose after “a recently discharged hospital employee killed one supervisor and critically wounded another.” After the incident, the employee criticized his employer publicly—by writing letters to a local newspaper, for example. In addition, the employee made numerous Facebook postings that referenced an on-going labor dispute, commented negatively on the employer’s management style, and discussed management’s mistreatment of employees. The postings received comments from coworkers expressing support. The employer disciplined the employee and later discharged him. In concluding that the postings constituted concerted activity, the Division of Advice stated that they “were the logical outgrowth of other employees’ collective concerns or were made with or on the authority of other employees.”

2. When Concerted Activity Was Not Present

The Division of Advice found that concerted activity was not present in the five other cases. In Case F, for example, the Division of Advice found that an employee’s two Facebook postings did not constitute concerted activity. After the employee’s “supervisor reprimanded her in front of the Regional Manager for failing to perform a task that she had never been instructed to perform,” she made a Facebook posting “that consisted of an expletive and the name of the [e]mployer’s

125. Id. at 26–30 (citing Case E).
126. Id. at 28.
127. Id. at 26.
128. Id.
129. Id. at 26–27.
130. Id. at 28–29.
131. Id. at 26–27.
132. Id. at 28.
133. Id. at 6–8 (citing Case F).
134. Id. at 7.
store.” One coworker responded by “Liking” the posting. Then, the employee made a second posting stating that her “employer did not appreciate its employees,” to which coworkers did not respond. In the following days, the employee discussed the incident with her coworkers, and she was fired. In concluding that the Facebook activity was not concerted, the Division of Advice characterized the postings as “merely an expression of an individual gripe.”

In Case G, the Division of Advice determined that an employee did not engage in concerted activity when she “posted angry profane comments on her Facebook wall” about her coworkers and employer. One coworker responded with empathy, saying “that she had gone through the same thing.” Subsequently, the employer fired the employee. The Division of Advice reasoned that the postings were not concerted because they “expressed [the employee’s] personal anger with coworkers and the [e]mployer, were made solely on her own behalf, and did not involve the sharing of common concerns.”

Similarly, in Children’s National Medical Center, the Division of Advice found that two Facebook postings did not constitute concerted activity. The first posting expressed the employee’s frustration with the way that a doctor had treated the employee, a respiratory therapist, and the second posting

135. Id. at 6.
136. Id.; see also Facebook, Like, http://www.facebook.com/help/like (last visited Sept. 27, 2012) (“Clicking Like under something you or a friend posts on Facebook is an easy way to let someone know that you enjoy it, without leaving a comment.”) (on file with the Washington and Lee Law Review).
137. SECOND NLRB REPORT, supra note 46, at 6 (citing Case F).
138. Id. at 6–7.
139. Id. at 7.
140. Id. at 11–13 (citing Case G).
141. Id. at 11. The employee “had a history of conflict with several coworkers.” Id.
142. Id. at 12.
143. Id.
144. Id. The Division of Advice characterized the postings as “rants” and “general profanities.” Id.
145. Id. at 30–32 (citing Children’s Nat’l Med. Ctr., N.L.R.B. No. 05-CA-36658 (Div. of Advice Nov. 14, 2011)).
146. Id. at 32.
expressed the employee’s irritation with a coworker who was sucking his teeth at work. The Division of Advice’s analysis of concertedness was brief because it found that the content of the postings was not protected. Nevertheless, the Division of Advice stated that even if the first posting were protected, it was not concerted because it “was merely a personal complaint.”

In *Buel, Inc.*, the Division of Advice found that an employee’s Facebook activity was not concerted. The employee, a truck driver, discovered that roads were closed because of the weather. He attempted, unsuccessfully, to reach his employer’s on-call dispatcher and discussed his inability to reach the dispatcher with other drivers. He also made several Facebook postings, commenting on the situation. One of the employee’s “Facebook friends,” the employer’s operations manager, responded critically, and a conversation ensued. Following, the employer revoked the employee’s “status as a leader operator,” and the employee resigned, claiming he was forced to do so.

Despite the discussion that the employee had with other drivers before posting on Facebook, the Division of Advice found that “there [was] insufficient evidence that his Facebook activity was a continuation of any collective concerns.” Rather, the employee “was simply expressing his own frustration and boredom about his inability to reach the on-call dispatcher.”

Finally, in *Frito-Lay, Inc.*, the Division of Advice found that an employee was not engaging in concerted activity when he

---

147. *Id.* at 30–31.
148. *Id.* at 31–32.
149. *Id.* at 32. The Division of Advice characterized the second posting in similar terms. *Id.* at 31–32.
150. *Id.* at 32–34 (citing *Buel, Inc.*, N.L.R.B. No. 11-CA-22936 (Div. of Advice July 28, 2011)).
151. *Id.* at 32.
152. *Id.*
153. *Id.*
154. *Id.*
155. *Id.* at 32–33.
156. *Id.* at 33.
157. *Id.*
158. *Id.*
159. *Id.* at 34–35 (citing *Frito-Lay, Inc.*, N.L.R.B. No. 36-CA-10882 (Div. of Advice Sept. 19, 2011)).
posted on Facebook but was “just venting.” The employee, who was feeling sick, talked to his supervisor about leaving work early. The supervisor told the employee that he could leave early but would forfeit an attendance point. Because the employee did not want to lose an attendance point, he completed his shift. After his shift, the employee made a Facebook posting, for which he was later discharged, “indicating that it was too bad when your boss doesn’t care about your health.” The Division of Advice concluded that the Facebook activity was not concerted; it was not an outgrowth of prior group action, did not attempt to initiate group action, and did not receive responses from coworkers.

3. What the Cases in the Second Report Tell Us

The factors that the Division of Advice considered regularly in its analysis of the First Report’s cases also appeared in its analysis of the Second Report’s cases. Frequently, the Division of Advice considered: (1) whether the posting grew out of prior non-Facebook group activity, (2) whether the posting contemplated future non-Facebook group activity, (3) whether coworkers responded to the posting.

160. Id. at 35.
161. Id. at 34.
162. Id.
163. Id.
164. Id.
165. Id. at 35.
166. See id. at 35 (citing Frito-Lay, Inc., N.L.R.B. No. 36-CA-10882, at 4 (Div. of Advice Sept. 19, 2011)) (finding that the postings were not “an outgrowth of prior employee meetings”); Id. at 33 (citing Buel, Inc., N.L.R.B. No. 11-CA-22936, at 4 (Div. of Advice July 28, 2011)) (“Although he had discussed with other drivers the fact that the on-call dispatcher was not reachable, there is insufficient evidence that his Facebook activity was a continuation of any collective concerns.”); Id. at 21 (citing Case C) (“[P]rior to her Facebook postings, the Charging Party spoke to two coworkers . . . .”); Id. at 23 (citing Case D) (describing the postings as “a continuation of . . . earlier group action that included employee complaints to management about the Employer’s Operations Manager”); Id. at 28 (citing Case E) (describing the postings as a “logical outgrowth” of prior concerted activity); Id. at 7 (citing Case F) (“[T]he post did not grow out of a prior discussion . . . with her coworkers.”).
167. See id. at 32 (citing Children’s Nat’l Med. Ctr., N.L.R.B. No. 05-CA-36658, at 3 (Div. of Advice Nov. 14, 2011)) (“The Charging Party was not seeking to induce or prepare for group action . . . .”); Id. at 35 (citing Frito-Lay, Inc.,
posting,168 (4) whether the employee who made the posting discussed the posting itself with coworkers,169 (5) the intent of the employee who made the posting, as reflected by the language and context of the posting,170 and (6) how coworkers interpreted the posting, as reflected by their responses.171

Furthermore, the Division of Advice considered an additional factor that it did not consider in the cases discussed

N.L.R.B. No. 36-CA-10882, at 3–4 (Div. of Advice Sept. 19, 2011) (same); Id. at 33 (citing Buel, Inc., N.L.R.B. No. 11-CA-22936, at 4 (Div. of Advice July 28, 2011)) (same); Id. at 7 (citing Case F) (“[T]he post contained no language suggesting that she sought to initiate or induce coworkers to engage in group action . . . .”); Id. at 12 (citing Case G) (same).

168. See id. at 32 (citing Children’s Nat’l Med. Ctr., N.L.R.B. No. 05-CA-36658, at 3 (Div. of Advice Nov. 14, 2011)) (“[N]one of her coworkers responded.”); Id. at 35 (citing Frito-Lay, Inc. N.L.R.B. No. 36-CA-10882, at 3 (Div. of Advice Sept. 19 2011)) (“[N]one of his coworkers responded to the postings with similar concerns.”); Id. at 33 (citing Buel, Inc., N.L.R.B. No. 11-CA-22936, at 4–5 (Div. of Advice July 28, 2011)) (“[N]one of his coworkers responded to his complaints about work-related matters.”); Id. at 5 (citing Case A) (“[C]oworkers and former coworkers responded.”); Id. at 22 (citing Case C) (“The Charging Party’s Facebook post sparked a collective dialogue that elicited responses from three of her coworkers . . . .”).

169. See id. at 32 (citing Children’s Nat’l Med. Ctr., N.L.R.B. No. 05-CA-36658, at 3 (Div. of Advice Nov. 14, 2011)) (“The Charging Party did not discuss her Facebook post with any of her fellow employees . . . .”); Id. at 33 (citing Buel, Inc., N.L.R.B. No. 11-CA-22936, at 3 (Div. of Advice July 28, 2011)) (“The Charging Party did not discuss his Facebook posts with any of his fellow employees . . . .”).

170. See id. at 32 (citing Children’s Nat’l Med. Ctr., N.L.R.B. No. 05-CA-36658, at 3 (Div. of Advice Nov. 14, 2011)) (finding the postings revealed the employee’s intent to express “a personal complaint about something that had happened on her shift.”); Id. at 35 (citing Frito-Lay, Inc., N.L.R.B. No. 36-CA-10882, at 4 (Div. of Advice Sept. 19, 2011)) (finding the postings revealed the employee’s intent to “vent.”); Id. at 33 (citing Buel, Inc., N.L.R.B. No. 11-CA-22936, at 4 (Div. of Advice July 28, 2011)) (finding the postings revealed the employee’s intent to “express[] his own frustration and boredom while stranded by the weather.”); Id. at 7 (citing Case F) (finding the postings revealed the employee’s intent to express “an individual gripe.”); Id. at 12 (citing Case G) (finding the postings revealed the employee’s intent to “express[] her personal anger with coworkers and the Employer”).

171. See id. at 29 (citing Case E) (“[F]ellow employees posted many messages of support for the Charging Party’s statements and general encouragement for his activity on his Facebook page . . . .”); Id. at 7 (citing Case F) (“Although one of her coworkers offered her sympathy and indicated some general dissatisfaction with her job, [the coworker] did not engage in any extended discussion with the Charging Party over working conditions or indicate any interest in taking action with the Charging Party.”).
by the First Report: whether the firing that resulted from the employee’s Facebook activity constituted a “pre-emptive” strike. In other words, sometimes Facebook activity itself does not quite rise to the level of concerted activity or is concerted only in a preliminary sense. Nevertheless, the Division of Advice indicated that it would consider the activity concerted under the Act if it represented a movement towards concerted activity that could never occur because of the firing. For example, when explaining the Division of Advice’s findings in Case C, the Acting General Counsel stated:

While the concerted actions expressed in the posts were of a preliminary nature, we concluded that the movement toward concerted action was halted by the Employer’s pre-emptive discharge and discipline of all the employees involved in the Facebook posts. Thus, we concluded that the Employer unlawfully prevented the fruition of the employees’ protected concerted activity.172

This “pre-emptive strike” concept relates closely to the second factor identified in Part III.A.3: whether the posting contemplated future non-Facebook group activity. The language, however, is unique to the Second Report and the cases contained therein, indicating that perhaps the concept is slightly different.

C. The ALJ Cases

1. Hispanics United of Buffalo, Inc.

Until September 2011, an ALJ did not have occasion to apply the Meyers I standard in a case involving a Facebook firing.173 Other than the First NLRB Report and the Advice Memoranda summarized therein, little guidance existed on how the standard

172. Id. at 22 (citing Case C); see also id. at 20 (citing Case B) (“We therefore concluded that Charging Party was discharged . . . as a ‘pre-emptive strike’ because of the Employer’s fear of what those discussions might lead to.”).

173. One reason was choice—the NLRB did not begin issuing Facebook firing complaints until October 2010. See Complaint & Notice of Hearing, Am. Med. Response of Conn., Inc., N.L.R.B. No. 34-CA-12576 (Region 34 Oct. 27, 2010) (first Facebook firing complaint). Another reason was opportunity—many Facebook firing cases settled prior to adjudication. U.S. CHAMBER OF COMMERCE, supra note 40, at 5.
would apply in the social media context. This changed when an ALJ issued the first decision of its kind in *Hispanics United*. *Hispanics United* held that employees’ Facebook postings constituted protected concerted activity, making their termination unlawful under the NLRA. Hispanics United of Buffalo, Inc. (HUB) employed Lydia Cruz-Moore as a domestic violence social worker. Cruz-Moore was often critical of other HUB employees’ job performance and communicated this criticism through text messages and in-person conversations. In particular, she told one HUB employee, Mariana Cole-Rivera, of her plans to raise her concerns with HUB’s Executive Director, Lourdes Iglesias. Following this conversation, Cole-Rivera posted a message on her Facebook page: “Lydia Cruz, a coworker feels that we don’t help our clients enough at HUB I about had it! My fellow coworkers how do u feel?” Several HUB employees responded by posting comments. After Cruz-Moore complained to Iglesias about the Facebook activity, he met individually with five of the employees involved in the Facebook activity and fired each of them.

The ALJ—faced with the issue of whether the HUB employees’ Facebook postings constituted protected concerted activity—applied the standard of *Meyers I* and its progeny. The ALJ asserted that the employees “were taking a first step towards taking group action to defend themselves against the accusations they could reasonably believe Cruz-Moore was going to make to management.” Thus, the Facebook postings

---

174. The Acting General Counsel did not issue the Second Report until after the three ALJ cases discussed in this Note had been decided.
175. *Hispanics United of Buffalo, Inc., N.L.R.B. No. 3-CA-27872, at 7–9, 2011 WL 3894520 (Div. of Judges Sept. 2, 2011).* The case is on appeal before the five-member board.
176. *Hispanics United, N.L.R.B. No. 3-CA-27872, at 4.*
177. *Id.*
178. *Id.*
179. *Id.*
180. *Id.* at 4–6. Cole-Rivera responded to some of the comments. *Id.* at 5–6. In addition, Cruz-Moore posted: “Marianna stop with ur lies about me I’ll b at HUB Tuesday.” *Id.* at 6.
181. *Id.*
182. *Id.* at 7–8.
183. *Id.* at 8–9.
constituted concerted activity. Because the ALJ found that the other elements of a Section 8 violation were present, he asserted that HUB had terminated its employees in violation of the Act.

Although the Board asserted in *Meyers I* that whether an activity is concerted is a separate inquiry from whether an activity is protected, the ALJ in *Hispanics United* fused both inquiries together and reversed the analysis. The ALJ first determined that the activity was protected, and then he determined that the activity was concerted. As a consequence, the ALJ did not provide a thorough rationale for his determination that the Facebook postings constituted concerted activity. Interestingly, the ALJ did not address the unique context in which the activity arose, likely because HUB “concede[d] that regardless of whether the comments and actions of the five terminated employees took place on Facebook or ‘around the water cooler’ the result would be the same.”

---

184. *Id.* Before progressing to the ALJ stage, the Division of Advice issued a memorandum, advising the regional office to issue a complaint. The Division of Advice, as noted by the Acting General Counsel in the First Report, found the that the activity involved was concerted: “[T]he Facebook discussion here was a textbook example of concerted activity, even though it transpired on a social network platform.” *First NLRB Report*, supra note 45, at 4.

185. *Hispanics United of Buffalo, Inc.*, N.L.R.B. No. 3-CA-27872, at 10, 2011 WL 3894520 (Div. of Judges Sept. 2, 2011). First, the ALJ determined that HUB knew of the concerted nature of employees’ activity. *See id.* at 9 (“[T]he fact that [HUB] lumped the discriminates together in terminating them, establishes that [HUB] viewed the five as a group and that their activity was concerted.”). Second, the ALJ determined that the activity was protected. *See id.* at 8 (“I conclude that their Facebook communications with each other, in reaction to a co-worker’s criticisms of the manner in which HUB employees performed their jobs, are protected.”). Finally, the ALJ noted HUB’s concession that the Facebook postings motivated the firings. *Id.*

186. *See Meyers I*, 268 N.L.R.B. 493, 497 (1984) (“Once an activity is found to be concerted, an 8(a)(1) violation will be found if, *in addition*, . . . the concerted activity was protected by the Act . . . .”) (emphasis added).

187. This is problematic because whether an activity is concerted concerns the character of the activity, while whether an activity is protected concerns the content of the activity.

188. *Hispanics United*, N.L.R.B. No. 3-CA-27872, at 8. It is unclear why HUB made this concession.
2. Karl Knauz Motors, Inc.

A few weeks after the *Hispanics United* decision, another ALJ issued the second Facebook firing decision. *Karl Knauz Motors* held that an employee’s Facebook posting on one topic constituted protected concerted activity, while the employee’s posting on a different topic did not constitute protected concerted activity. Because the ALJ determined that the latter posting caused the employee’s termination, the ALJ found that the discharge did not violate the NLRA.

The dispute involved Karl Knauz Motors, Inc. (Karl Knauz) and its employee, Robert Becker, a car salesman at its BMW dealership. The dealership was organizing an “Ultimate Driving Event” at which it planned to introduce a redesigned BMW 5 Series automobile. Prior to the event, Becker’s supervisor, Phillip Ceraulo, held a meeting at which he informed the salespeople of the dealership’s plan to have a hot dog cart at the event. Both during and after the meeting, Becker and other salespeople allegedly commented on the disconnect between the hot dog cart and BMW’s status as a luxury brand and the effect that the disconnect could have on their commissions. While at the event, Becker took pictures of the food, including pictures of salespeople holding hot dogs. He later posted those pictures with descriptions on his Facebook page, and some of Becker’s Facebook friends posted comments, to which Becker responded.

Several days after the event, an accident occurred at a Land Rover dealership, also owned by Karl Knauz, that was located adjacent to the BMW dealership. Becker took pictures of the accident and posted pictures, along with comments, on his

190. *Id.* at 9.
191. *Id.* The case is on appeal before the five-member board.
192. *Id.* at 1–2.
193. *Id.* at 2.
194. *Id.*
195. *Id.*
196. *Id.* at 3.
197. *Id.* at 3–4.
198. See *id.* at 3 (describing the accident).
Facebook page. Some of Becker’s Facebook friends, including other Karl Knauz employees, posted comments, to which Becker responded. After Karl Knauz discovered the postings of the event and the accident, it terminated Becker’s employment.

Thus, the ALJ had to determine whether Becker’s Facebook postings of (1) the Ultimate Driving Event (event posting), or (2) the accident at the Land Rover dealership (accident posting), or (3) both constituted protected concerted activity, making Becker’s termination unlawful under the NLRA. In applying the standard of *Meyers I* and its progeny, the ALJ found that Becker’s event posting constituted concerted activity:

As both Larsen [another salesperson] and Becker spoke up at the meeting commenting on what they considered to be the inadequacies of the food being offered at the event, and the subject was further discussed by the salespersons after the meeting, even though only Becker complained further about it on his Facebook pages without any further input from any other salesperson, other than the Facebook pictures of [two other salespeople], I find that it was concerted activities . . . .

Because Becker’s individual action was a “logical outgrowth of prior concerted activity,” his individual action was concerted.

In contrast, the ALJ found that Becker’s accident posting did not constitute concerted activity. Supporting his conclusion, the ALJ emphasized that “[i]t was posted solely by Becker,

---

199. *Id.* at 4.
200. *Id.*
201. *Id.* at 5–6
202. *Id.* at 7.
203. *Id.* at 8. In addition, the ALJ determined that the activity was protected. See *id.* (“[I]t was protected . . . as it could have had an effect upon [Becker’s] compensation.”). Before progressing to the ALJ stage, the Division of Advice issued a memorandum, advising the regional office to issue a complaint. The Division of Advice, as noted by the Acting General Counsel in the First Report, found that the activity involved was concerted: “The Facebook activity was a direct outgrowth of the earlier discussion among the salespeople that followed the meeting with management.” *First NLRB Report*, *supra* note 45, at 8.
205. *Id.* at 9.
apparently as a lark, without any discussion with any other employee.”206 Because the ALJ found that Becker’s accident posting, rather than his event posting, motivated his discharge, the ALJ found that Karl Knauz did not terminate Becker in violation of Section 8 of the Act.207

3. Triple Play Sports Bar & Grille

The first ALJ Facebook firing decision of 2012—the third decision overall—was issued on January 3, 2012. Triple Play held that two employees engaged in concerted activity when they responded to a Facebook posting by a former coworker about their employer’s tax withholding practices.208 Triple Play Sports Bar and Grille (Triple Play) employed Jillian Sanzone as a waitress and bartender and Vincent Spinella as a cook.209 When Sanzone filed her tax returns, she realized that she owed taxes to the state.210 She talked about the tax issue with coworkers, and her supervisors arranged a staff meeting to discuss the issue.211

Before the staff meeting took place, a former employee of Triple Play made the following Facebook posting: “Maybe someone should do the owners of Triple Play a favor and buy it from them. They can’t even do the tax paperwork correctly!!! Now I OWE money . . . Wtf!!!!”212 The posting received a number of responses from customers and former coworkers, including Sanzone.213 In addition, the former employee who made the original posting replied to some of the responses.214 After eleven responses had been made to the original posting, all expressing frustration with owing taxes, Sanzone stated: “I owe too. Such an

206. Id. In addition, the ALJ determined that the activity was not protected. See id. (“It is so obviously unprotected . . . .”).
207. Id.
210. Id. at 3.
211. Id.
212. Id.
213. Id. at 3–4.
214. Id.
Although Spinella did not respond with a textual comment, he clicked the “Like” button under the initial posting. After two supervisors learned of the Facebook posting and the responses, they fired both Sanzone and Spinella.

Consequently, the ALJ had to determine whether the two employees’ Facebook postings constituted protected concerted activity, making their termination unlawful under the NLRA. In answering this question, the ALJ focused on the fact that the Facebook postings were “part of an ongoing sequence of events” involving the tax issue. Employees of Triple Play had discussed the matter before the Facebook activity, and a meeting was scheduled to discuss the matter in the near future. In addition, the ALJ notably asserted: “The specific medium in which the discussion takes place is irrelevant to its protected nature.” Therefore, the ALJ concluded that the Facebook activity constituted concerted activity.

215. Id.
216. Id. at 4. After clicking “Like,” the text “Vincent VinnyCenz Spinella and Chelsea Molloy like this” appeared automatically below the original posting. Id. Spinella testified that he clicked “Like” after the fifth response had been made, a comment by the person who made the original posting: “It’s all Ralph’s [one of the supervisor’s] fault. He didn’t do the paperwork right. I’m calling the labor board to look into it because he still owes me about 2000 in paychecks.” Id.
217. Id. at 4–5.
218. Id. at 8–9.
219. Id. at 8.
220. Id. Thus, the Facebook activity was a “logical outgrowth” of prior concerted activity. See id. (“[T]he Facebook discussion was part of a sequence of events, including other, face-to-face employee conversations . . . .”).
221. Id. Thus, the Facebook activity related to the preparation of group action. See id. (“The employees who posted comments . . . specifically discussed the issues they intended to raise at [the] upcoming meeting and avenues for possible complaints to government entities.”).
222. Id. The ALJ’s use of the phrase “protected nature” appears to encompass both whether the activity is protected and whether the activity is concerted.
223. Id. Before progressing to the ALJ stage, the Division of Advice issued a memorandum, advising the regional office to issue a complaint. The Division of Advice, as noted by the Acting General Counsel in the First Report, found the that the activity involved was concerted: “[T]he conversation that transpired on Facebook not only embodied ‘truly group complaints’ but also contemplated future group activity.” First NLRB Report, supra note 45, at 10.
Perhaps the most interesting part of the ALJ’s discussion came when she discussed the implications of Spinella clicking “Like,” rather than responding in text form. The ALJ stated that clicking the “Like” button “constituted participation in the discussion that was sufficiently meaningful as to rise to the level of concerted activity.” Because the ALJ determined that the other elements of a Section 8 violation were present and that the Facebook activity did not lose the protection of the Act, she found that Triple Play had terminated both employees in violation of the Act.

4. What the ALJ Cases Tell Us

In deciding the previous three cases, the ALJs considered some of the same factors that the Acting General Counsel and Division of Advice considered in the two reports and the Advice Memoranda cited therein. For example, the ALJs in both Triple Play and Karl Knauz Motors afforded weight to the fact that the postings grew out of prior non-Facebook activity—factor one. In addition, the ALJs in both Triple Play and Hispanics United found that the postings contemplated future non-Facebook activity—factor two.

224. Triple Play Sports Bar & Grille, N.L.R.B. No. 34-CA-12915, at 8–9, 2012 WL 76862 (Div. of Judges Jan. 3, 2012). The ALJ emphasized that “the Board has never parsed the participation of individual employees in otherwise concerted conversations, or deemed the protections of Section 7 to be contingent upon their level of engagement or enthusiasm.” Id. at 9.

225. Id. at 22. In addition to finding that the activity was concerted, the ALJ determined that the activity was protected. See id. at 8 (“It is beyond question that issues related to wages, including the tax treatment of earnings, are directly related to the employment relationship and may form the basis for protected concerted activity . . . .”). The ALJ also found that Triple Play knew that its employees were engaged in protected concerted activity and that Facebook activity motivated the firings. Id. at 14–15.

226. See id. at 8 (discussing the in-person conversations between employees that had taken place before the Facebook posting); Karl Knauz Motors, Inc., N.L.R.B. No. 13-CA-46452, at 8, 2011 WL 4499437 (Div. of Judges Sept. 28, 2011) (same).

Furthermore, at least implicitly, the ALJs focused on the intent of those involved—factors four and five. For example, in *Hispanics United*, the ALJ indicated that the original poster and the coworkers who responded intended “to defend themselves against . . . accusations.” The *Hispanics United* decision suggests that if the intent of an employee who makes a Facebook posting aligns with the intent of the coworkers who respond, the Facebook activity is likely concerted. In both *Triple Play* and *Karl Knauz Motors*—cases that presented different factual circumstances than *Hispanics United*—the ALJs focused on the intent of the parties to continue a discussion that had begun previously or to prepare for a future discussion that was planned before the commencement of the Facebook activity.

The ALJs, however, did not afford nearly as much significance to two of the factors identified in Part III.A.3—factors three and four. Although the ALJs in all three cases mentioned whether coworkers responded to the initial posting, they did not emphasize coworker responses as essential to their analysis. In fact, in *Karl Knauz Motors*, the ALJ found that the posting to which coworkers did not respond—the accident posting—was not concerted, while the posting to which coworkers did not respond—the event posting—was concerted. Moreover, none of the coworkers responded to the original posting because they knew that their coworker would do so.


229. In *Triple Play*, the person who made the original posting was not a current employee of the company charged with an unfair labor practice, but a former one. *Triple Play*, N.L.R.B. No. 34-CA-12915, at 3. In *Karl Knauz Motors*, the posting that the ALJ found was concerted received no coworker responses. *Karl Knauz Motors*, N.L.R.B. No. 13-CA-46452, at 8.

230. See *Triple Play*, N.L.R.B. No. 34-CA-12915, at 8 (finding that the employees’ responses to the original posting indicated an intent to “discuss[] the issues they intended to raise at [the] upcoming meeting and avenues for possible complaints to government entitles”); *Karl Knauz*, N.L.R.B. No. 13-CA-46452, at 8 (finding that the event posting reflected an intent to continue expressing complaints about management’s choice of food expressed previously at a meeting).

231. See *Karl Knauz Motors*, N.L.R.B. No. 13-CA-46452, at 3 (stating that multiple coworkers responded to the accident posting).

232. See *id.* at 7 (“[E]ven though only Becker complained further about it on his Facebook pages without any further input from any other salesperson . . . I find that it was concerted . . . .”).
the ALJs considered whether the person who made the original posting discussed the posting itself with coworkers.

Finally, although the ALJs did not use the “pre-emptive strike” terminology that the Division of Advice employed, the ALJ in Hispanics United invoked the concept by stating: “By discharging the discriminatees . . . [the employer] prevented them [from] taking any further group action.”

IV. Why the Concerted Activity Standard Is Problematic in the Social Media Context

Many have accused the Board of playing politics, arguing that the Board’s decision to enter into the social media realm is part of a larger pro-union effort to expand its role. Still, others have argued that the Board’s role in enforcing the NLRA has never been more important, given the decline in union membership that has occurred over the past few decades. Politics aside, however, there are problems with the way that the Board has applied the traditional concerted activity standard in the emerging context of social media.

A. Where the Board Draws the Line

The root of the difficulty in applying the standard is the same in the social media context as it is in other contexts: determining

233. See supra note 172 and accompanying text.
235. See Dave Jamieson, Wilma Liebman, Outgoing NLRB Chair, Finds ‘Silver Lining’ in Political Rancor, HUFFINGTON POST (Sept. 5, 2011, 9:09 AM), http://www.huffingtonpost.com/2011/09/05/wilma-liebman-nlrb-chairwoman-interview_n_947258.html (last visited Sept. 27, 2012) (“Corporations and their allies have decried the board as ‘out of control’ during Liebman’s tenure as chairwoman.”) (on file with the Washington and Lee Law Review). Those who argue that the Board is “out of control” point to other recent Board actions, in conjunction with those related to social media. Id.
236. See William R. Corbett, Waiting for the Labor Law of the Twenty-First Century: Everything Old Is New Again, 23 BERKLEY J. EMP. & LAB. L. 259, 267 (2002) (“As union representation continues to decline, particularly in the private sector, a broad interpretation and application of section 7 in the nonunion workplace is even more important today than it was ten or twenty years ago.”).
the point at which individual activity transforms into concerted activity. As the Supreme Court noted in *NLRB v. City Disposal Systems, Inc.*, challenging issues often arise concerning “the precise manner in which particular actions of an individual employee must be linked to the actions of fellow employees” in order for the individual action to be concerted. On the one hand, if the Board interprets and applies the *Meyers I* standard too broadly, the standard will lose all meaning. On the other hand, if the Board interprets and applies the standard too narrowly, the Board will fail to effectuate the Act’s purposes. To avoid these undesirable alternatives, the Board has attempted to draw a line between individually initiated activity that is concerted and that which is not.

In drawing this line, the Board distinguishes between situations in which an individual acts with the “intent[] to enlist the support of other employees in a common endeavor” and situations in which that intent is lacking. In *Mushroom Transportation Co.*, an oft-cited opinion, the Third Circuit explained that “a conversation may constitute concerted activity although it involves only a speaker and a listener.” However, this speaker-listener concept has limits:

Activity which consists of mere talk must, in order to be protected, be talk looking toward group action. If its only purpose is to advise an individual as to what he could or should do without involving fellow workers or union representation to protect or improve his own status or working position, it is an individual, not a concerted activity, and, if it looks forward to no action at all, it is more than likely to be mere “griping.”

Thus, “mere griping” is the antithesis of concerted activity and is not protected by the Act. The Board has used this concerted

---

238. *Id.* at 831.
240. *See id.* at 885 (stating that “an employee’s activities engaged in ‘solely by and on behalf of the employee himself,’” do not constitute concerted activity).
242. *Id.* at 685.
243. *Id.*
244. *See Robert A. Gorman & Matthew W. Finkin, The Individual and the*
activity—mere griping dichotomy to guide its analysis of Facebook firing cases. In a press release accompanying the Second Report, the Board emphasized: “An employee’s comments on social media are generally not protected if they are mere gripes not made in relation to group activity among employees.”

B. Why the Board’s Line Drawing Is Problematic

Stating that “mere griping” does not constitute concerted activity is nothing new. The social media forum, however, presents new challenges to those attempting to distinguish between concerted activity and mere griping. First, social media is, as its name suggests, social. Because it necessarily involves interaction to varying degrees, almost any individual action in a social media forum could be considered concerted. Therefore, in cases involving social media, it is easier to establish the “linkage to group action” that the Board requires. After all, employees are often Facebook friends with coworkers. If an employee makes a Facebook posting, coworkers can read the posting without responding, “Like” the posting, or respond in text form. If

---

Requirement of Concert Under the National Labor Relations Act, 130 U. PA. L. REV. 286, 290 (1981) (“[W]hen an individual employee protests alone, without any consultation with and authorization by fellow employees, his legal rights under section 7 may be drastically curtailed, even when he purports to voice the concerns of others but especially when he is speaking only for himself . . . .”).


246. See Gorman & Finkin, supra note 244, at 290 (“The prevailing principle of law—endorsed both by the courts of appeals and the NLRB—is that section 7 does not protect ‘personal gripes’ by individual employees.”).


concerted activity need only involve “a speaker and a listener” and an intent on the speaker’s part to initiate group action, it becomes difficult to imagine when Facebook activity concerning working conditions would not fit within the current definition of concerted activity. This is especially true given that the Board has indicated neither the degree of linkage to group action that it requires, nor how it determines employee intent, in the social media context.

At the same time, social media lends itself to griping. As anyone with a Facebook page knows, many Facebook users employ their page as a platform to air personal complaints—both work-related and otherwise. Facebook friends can respond to such complaints—offering sympathy or humor—with informality and ease. Consequently, Facebook does not fit into the present concerted activity standard easily because the forum is inherently concerted, yet it fosters activity that is the opposite of concerted: griping.

V. Possible Approaches

When thinking about how the Board can increase certainty in this area of law, it is important to be realistic. First, the NLRA is likely here to stay. Second, it is unlikely that the Board will completely re-construe Section 7 of the Act. Its current


250. See Corbett, supra note 236, at 264 (“While there may be a need to substantially overhaul the body of law regulating the employment relationship in the United States, it is doubtful that such a project will be undertaken by lawmakers absent an economic catastrophe.”). One could argue that the United States has experienced an economic catastrophe in the past few years. However, lawmakers’ response to the current economic crisis will likely be different from lawmakers’ response to the Great Depression—out of which the NLRA was born—because of the recent political controversies surrounding the Board’s use of power. Instead of overhauling the NLRA completely, lawmakers will likely continue the trend of enacting more individual employment rights laws, if they legislate in this area at all. Id.

251. Christine Neylon O’Brien, The First Facebook Firing Case Under Section 7 of the National Labor Relations Act: Exploring the Limits of Labor Law Protection for Concerted Communication on Social Media, 45 SUFFOLK U. L. REV. 29, 33 (2011) (“At the moment, it seems as though the current NLRB is poised to adapt existing legal doctrines to craft new rules and remedies regarding employer rules and restrictions concerning employee use of... social
construction of the Section, as reflected by the Meyers I standard, endured a challenge by the D.C. Circuit and has remained for over twenty-five years. Consequently, the three approaches that I suggest all work within the Board’s current framework.

A. Addressing the Problem at the Front-End

In addition to resolving the concerted activity issue in Facebook firing cases, the Board often determines the legality of employers’ social media policies and other rules in employers’ handbooks. As the Board stated in Lafayette Park Hotel, if an employer maintains rules that “are likely to have a chilling effect on Section 7 rights, the Board may conclude that their maintenance is an unfair labor practice, even absent evidence of enforcement.” To determine whether a rule has an impermissible chilling effect, the Board examines whether “(1) employees would reasonably construe the language [of the rule] to prohibit Section 7 activity; (2) the rule was promulgated in response to union activity; or (3) the rule has been applied to restrict the exercise of Section 7 rights.”

The ALJs in both Triple Play and Karl Knauz Motors examined employer rules in light of the above factors. In Triple Play, the ALJ analyzed the employer’s social media policy and concluded that the policy did not violate the Act. In contrast, the ALJ in Karl Knauz Motors found that rules in the employer’s handbook did violate the Act. Interestingly, the ALJ addressed

---

255. Id. at 825.
256. Lutheran Heritage Village-Livonia, 343 N.L.R.B. 646, 647 (2004). This inquiry is appropriate when a “rule does not explicitly restrict activity protected by Section 7.” Id. If a rule restricts Section 7 activity explicitly, the Board will find a violation of the Act without further inquiry. Id. at 646.
258. Karl Knauz Motors, N.L.R.B. No. 13-CA-46452, at 11. The ALJ found
the lawfulness of the rules even after finding that the employer did not discharge the employee because of the event posting—the posting that the ALJ found constituted protected concerted activity.259

One approach that the Board could take is to promulgate a model social media policy through its rulemaking process.260 Ideally, the promulgation of a model policy would lead to less Facebook firing complaints and subsequent adjudication because it would increase certainty for both employers and employees. When an employer became aware of employee Facebook activity that concerned the employer, he or she could analyze the activity in terms of how it fit within the contours of the model policy. The employer could then take action, or refrain from taking action, based on his or her determination of whether the employee activity violated the model policy. Because the employer’s determination would be based on a lawful model policy that would guide the employer’s decisionmaking, less adjudication of the lawfulness of employee discipline or discharge would likely result. The promulgation of a model policy would also benefit employees by informing their judgment when they use social media.

This approach would address the concerted activity problem at the front-end. The model policy would be preventative, in that it would hopefully lead to fewer unfair labor practice disputes in the first place. As a consequence, the Board would be called on less often to construe the concerted activity provision of Section 7 in the social media context.

the following rules unlawful: (1) a rule prohibiting employees from “be[ing] disrespectful or use[ing] profanity or any other language which injures the reputation of the dealership” and (2) rules “prohibit[ing] employees from participating in interviews with, or answering inquiries concerning employees from, practically anybody.” Id. at 9.

259. Id. at 9 (finding that the accident posting, which did not constitute protected concerted activity, caused the discharge). Moreover, the ALJ addressed the rules even though the employer rescinded them voluntarily prior to the hearing. Id. at 11.

260. In its Third Report, the Board analyzed seven social media policies. THIRD NLRB REPORT, supra note 53. Perhaps the NLRB could use the one policy that it deemed lawful in its entirety as a starting point to formulate a model policy. See id. at 19–24 (explaining why the policy was lawful and providing a copy of the policy).
This approach is not without problems, however. First, the model policy would not eliminate disputes altogether because employers and employees could misinterpret or misapply the policy, and the Board would have to resolve those disputes. Second, the Board tends to favor adjudication over rulemaking, a preference that has a persuasive rationale in certain circumstances. Finally, given the controversy surrounding the Board's recent actions, the Board may not be willing to risk the possibility of more backlash that may result from a visible use of power.

B. Addressing the Problem Head-On

Most likely, the Board will continue to address the problem head-on, by attempting to clarify through adjudication how the Act’s concerted activity provision—as interpreted by the Board in Meyers I and Meyers II—applies in cases involving social media. The Board's present attempts at clarification, however, have not provided the guidance that employers and employees need. While the Board provided details of social media cases in its reports, it neither identified nor explained the commonalities among the cases that it found significant. Likewise, the Board did not distinguish the cases from each other to show how the presence


262. See Chenery II, 332 U.S. 194, 202 (1947) (stating that an “agency may not have had sufficient experience with a particular problem to warrant rigidifying its tentative judgment into a hard and fast rule.”).

263. See Kathleen Furey McDonough, Labor: Political Controversy Surrounds the NLRB, INSIDE COUNSEL (Jan. 9, 2012), http://www.insidecounsel.com/2012/01/09/labor-political-controversy-surrounds-the-nlrb (last visited Sept. 27, 2012) (“In a political environment mired in controversy, the National Labor Relations Board . . . is one of the federal agencies receiving more than the usual dose of criticism.”) (on file with the Washington and Lee Law Review).

264. See FIRST NLRB REPORT, supra note 45; SECOND NLRB REPORT, supra note 46.
or absence of a certain fact affected its analysis of concerted activity.

Because the evaluation of Facebook firing cases is fact-specific, identifying factors unique to the social media context that tend to indicate concertedness, or lack thereof, could help remedy the uncertainty. This Note identifies factors that the Board seemed to consider significant to its analysis of concerted activity in Facebook firing cases. A similar identification of factors by the Board could help organize and guide the Board’s analysis in future cases and provide clarity for both employers and employees.

C. Addressing the Problem at the Back-End

Even if the Board finds that an employee’s Facebook activity constitutes protected concerted activity, the employee may nevertheless lose the protection of the Act under one of the Board’s several loss-of-protection standards. For example, in Atlantic Steel Co., the Board asserted: “[E]ven an employee who is engaged in concerted protected activity can, by opprobrious conduct, lose the protection of the Act.” Moreover, statements that constitute protected concerted activity but are disloyal can lose the protection of the Act if “they are made ‘at a critical time in the initiation of the company’s’ business and . . . constitute ‘a sharp, public, disparaging attack upon the quality of the company’s product and its business policies, in a manner reasonably calculated to harm the company’s reputation and reduce its income.’”


266. See supra Parts III.A.3, III.B.3, and III.C.4.


268. Id. at 816. “The decision as to whether the employee has crossed that line depends on several factors: (1) the place of the discussion; (2) the subject matter of the discussion; (3) the nature of the employee’s outburst; and (4) whether the outburst was, in any way, provoked by an employer’s unfair labor practice.”

Perhaps the Board could address the concerted activity problem by applying existing loss-of-protection standards more strictly to employee activity in the social media context or by developing a social media specific loss-of-protection standard. If the Board construes the Meyers I standard broadly when applying it in cases involving social media, it may be appropriate for the Board to apply a more stringent loss-of-protection standard subsequently to ensure that the activity in question is activity that Congress intended for the NLRA to protect. Social media cases warrant such a stringent application because social media alters the mix of employer and employee interests involved. In addition, the Board’s notice posting rule will lessen the long-standing concern that many, if not most, non-unionized employees are unaware of their rights under the NLRA. Because the rule alleviates this concern, it is reasonable to require employees to conform their actions more closely to what the Act requires in order to obtain the Act’s protection.


270. This broad construction could be intentional or unintentional. As explained previously, it is difficult to apply the Meyers I standard in the social media context because social media inherently involves some degree of interaction yet fosters griping. See supra Part IV.B.

271. See O’Brien, supra note 251, at 47 n.83 (“[E]mployee use of social media to vent discontent is more likely to extend beyond the workplace and damage a company’s reputation than an on-site verbal confrontation.”); Trottman, supra note 265 (discussing the reputational concerns that online postings implicate).


273. See Peter D. DeChiara, The Right to Know: An Argument for Informing Employees of Their Rights Under the National Labor Relations Act, 32 HARV. J. ON LEGIS. 431, 433 (1995) (“American workers are largely ignorant of their rights under the NLRA, and this ignorance stands as an obstacle to the effective exercise of such rights.”).
VI. Conclusion

When introducing the bill that eventually became the NLRA, the bill’s sponsor, Senator Robert F. Wagner, stated: “When employees are denied the freedom to act in concert even when they desire to do so, they cannot exercise a restraining influence upon the wayward members of their own groups, and they cannot participate in our national endeavor to coordinate production and purchasing power.”274 While emphasizing the importance of providing employees with an enforceable right to engage in concerted activity, Senator Wagner nevertheless acknowledged: “[E]mployers are tremendously handicapped when it is impossible to determine exactly what their rights are. Everybody needs a law that is precise and certain.”275

The need for precision and certainty is as great today as it was then. Precision and certainty in the Act’s application benefits both employees and employers and creates confidence in the Board. Because the Board’s foray into Facebook firing cases is recent, it would be unrealistic to expect the Board to have a ready-made solution to the issues that the social media context presents. Still, the Board must begin to acknowledge the differences that exist between the virtual water cooler and the traditional water cooler, and it must devise an approach. This Note presents three approaches—approaches that the Board can implement individually or in concert—formulated with an understanding of the interests involved and the context’s distinct attributes.

275. Id. The act that preceded the NLRA, the National Industrial Recovery Act, conferred on employees similar rights to engage in concerted activities, but it did not contain an enforcement mechanism. Id.
Do End-Users Get the Best of Both Worlds?—Title VII of Dodd–Frank and the End-User Exception

Carney Simpson*

Table of Contents

I. Introduction ................................................................... 1760

II. Over-the-Counter Derivative Market ........................... 1766
   A. Market Overview ..................................................... 1766
   B. Types of OTC Contracts .......................................... 1770
      1. Options ................................................................ 1770
      2. Forwards ............................................................ 1771
      3. Swaps ................................................................... 1771

III. End-User Participation in the OTC Market ................. 1773

IV. Regulation of the OTC Market ...................................... 1777
   A. Past Regulation ....................................................... 1777
   B. Dodd–Frank Regulation .......................................... 1781

V. End-User Exception to the Mandatory Clearing
   of Swaps ......................................................................... 1784
   A. Rule Design: Narrow Versus Broad ......................... 1785
   B. Alternative Rule Design .......................................... 1787
      1. Alternative Approaches to OTC
         Derivatives Regulation ........................................ 1788

VI. Amendment to the End-User Exception ....................... 1790
   A. Proposal for End-User Disclosure ........................... 1791

VII. Conclusion ...................................................................... 1794

* Candidate for J.D., Washington and Lee University School of Law, May 2013; B.B.A., University of Georgia, May 2009. I would like to give profound thanks to my faculty advisor, Professor Christopher Bruner, my Note editor, Gibson Wright, and the Editorial Board of the Washington and Lee Law Review for their input and support.
I. Introduction

“Derivatives are financial weapons of mass destruction, carrying dangers that, while now latent, are potentially lethal.”

Warren Buffett made this statement to Berkshire shareholders in 2002. The potential threat of the derivatives market was known then and became real in 2008. Now the government is taking action to control it.

The financial crisis of 2008 was one of the largest in American history and almost led to the collapse of the U.S. financial system. Factors likely contributing to the crisis include the repeal of the Glass-Steagall Act, the influx of foreign money, the popularity of hedge funds and private equity, and the rise of mortgage-backed securities. Although the financial crisis of 2008 had numerous causes, the over-the-counter (OTC) derivative market was one of the most noted.


DO END-USERS GET THE BEST OF BOTH WORLDS?

Some say that the OTC market’s influence on the financial crisis was a result of the market’s large financial volume and a lack of corresponding government regulation. As of July 2010, the OTC derivatives market had a notional value of approximately $300 trillion in the United States. This amount is roughly twenty times the size of the American economy. The notional amount is a way in which derivatives are priced, but it does not consider the risk involved. This risk can be much smaller or larger than the notional contract amount. Because this market is large and rapidly growing, regulation is necessary but will never completely eliminate the risks the market poses because of the complexity of the contracts and the rate of innovation in the market.

The widespread use of OTC derivatives for speculative purposes exposes market participants to systemic risk. Systemic risk arises when investors hold highly leveraged positions that could trigger a crisis like that of 2008. Due to the number of

6. See Lynn A. Stout, How Deregulating Derivatives Led to Disaster, and Why Re-Regulating Them Can Prevent Another 1 (UCLA Sch. of Law, Law-Econ Research, Paper No. 09-13) [hereinafter Stout, Deregulating Derivatives] (theorizing that the banking system failure in 2008 was caused primarily by the deregulation of the derivatives market in 2000); see also infra Part IV.A (providing a background on past derivatives regulation and why the government chose not to act until the 2008 crisis).


8. Id.


10. See id.

11. See Henry T.C. Hu, Hedging Expectations: “Derivative Reality” and the Law and Finance of the Corporate Objective, 73 TEX. L. REV. 985, 1013–14 (explaining how the unprecedented rate of financial innovation in the OTC market is largely due to persons with quantitative or physical science backgrounds playing vital roles in pricing the contracts).


13. See Frank Partnoy & David A. Skeel, Jr., The Promise and Perils of
OTC contracts and their interconnection with other trading instruments, a small market shift in the value of an OTC derivative could lead to a major international liquidity problem.\(^\text{14}\) OTC derivatives enhance systemic risk dramatically because they lack transparency.\(^\text{15}\) Improving transparency in the OTC market is precisely why regulation and disclosure are necessary for all participants.

To protect the economy from systemic risk, Congress passed the Dodd–Frank Wall Street Reform and Consumer Protection Act\(^\text{16}\) (Dodd–Frank) in an effort to increase transparency, regulate pricing in the derivatives market, and, most importantly, minimize the risk to the American people.\(^\text{17}\) Title VII of Dodd–Frank proposes guidelines that the Securities and Exchange Commission (SEC) and the U.S. Commodity Futures Trading Commission (CFTC) are required to follow when promulgating regulations for the derivatives market.\(^\text{18}\) One of the most important changes Dodd–Frank requires is the formulation of a clearing organization for OTC derivatives, particularly swaps.\(^\text{19}\)

The regulation for clearing organizations requires market participants to set initial margin requirements, post or recover collateral at the end of each day, and provide certain disclosures that were previously not required.\(^\text{20}\) This only applies to “swap

---

\(^{\text{Credit Derivatives, 75 U. CINN. L. REV. 1019, 1040 (2010) (discussing how the size of the credit default swap market could cause a ripple effect throughout financial markets even when only one small thing goes awry).}}\)

\(^{\text{14. See id.}}\)

\(^{\text{15. See Ariail, supra note 12, at 179–80 (indicating that the opaqueness of the market is a major risk associated with the OTC market).}}\)


\(^{\text{18. Dodd–Frank § 717.}}\)

\(^{\text{19. Id. § 725. This section pertains specifically to CFTC’s jurisdiction over non-security based swaps. Id.}}\)

dealers” and “major swap participants.” These definitions apply to participants who hold large positions that would create substantial counterparty exposure and increase the threat of systemic risk.

When defining these terms, the CFTC examined the appropriate risk factors associated with these users and the role they allegedly played in the 2008 financial crisis. Nonfinancial corporations who participate in the market, commonly classified as “end-users,” are not subject to the mandatory clearing of swaps. An end-user is a corporation that utilizes the OTC market to enter into customized contracts that hedge an already exposed risk, such as fluctuation in interest rates or foreign currency. A survey of the world’s largest 500 companies revealed that 94% of them use the derivatives market as end-users to manage and hedge business risk. Thus, a large number of market participants are not subject to regulation.

The end-user exception defines end-users narrowly and provides minimal regulation requirements. It applies to a nonfinancial entity that is a participant and uses swaps to hedge or mitigate commercial risk. The only real requirement for end-users is to elect the exception and then notify the Commission as to how it will generally meet its financial obligation.

23. See id.
25. See infra notes 83–87 and accompanying text (introducing the corporate end-user and its role in the OTC market).
28. Id.
29. Id.; see infra Part V.A for an in-depth description of the end-user exception.
Commercial end-users disfavor the narrow interpretation and the minimal level of regulation that it imposes. Commercial end-users wanted the CFTC to broadly cast the definition of end-user to include affiliate companies who utilize the market and, further, allow commercial end-users to hedge risk in the OTC market with minimal transaction costs. This Note examines the CFTC’s rule pertaining to the end-user exception and analyzes whether the CFTC’s narrow interpretation is the most beneficial approach. This Note proposes that the CFTC broaden the end-user definition to exempt more users, such as small banks and corporate affiliates. But to broaden the exception, this Note argues that regulators must impose stricter disclosure and reporting requirements to monitor the market for abuse of the exception. The end-user exemption’s aim is to maintain low transaction costs, but regulators must monitor closely the potential abuse of this exception. The risk is that participants could disguise speculative uses of the market in the form of bona-fide hedging.

This Note emphasizes the CFTC’s definition of end-user because the CFTC has jurisdiction over all commodity-based swaps (including interest rate swaps), options, and futures. Based on statistics from the Bank of International Settlements, interest rate contracts on the OTC market comprised 78% of all


31. See id. at 6–7 (addressing the particular concerns Kraft Foods, Inc. has pertaining to the definition of the end-user).

32. See id. at 2 (discussing how Kraft’s affiliate corporations utilize the OTC derivative market as agents for Kraft and Kraft’s subsidiary companies to hedge exposed market risk). Kraft has a complex structure that calls for two affiliate companies to enter into swap transactions with Kraft, its subsidiary corporations, and at times, swap dealers. Id. Due to the affiliate corporations’ active participation in the OTC market, Kraft submitted a comment to the CFTC end-user exception to guarantee that the affiliate’s participation and use would fall within the definition of end-user. See id. at 4–5.

33. See 156 Cong. Rec. 105, S5293 (daily ed. July 15, 2010) (statement of Sen. Blanche Lincoln) (stating that non-narrow-based security index swaps and credit default swaps may be the only swaps that will fall within the jurisdiction of the SEC).
outstanding contracts on the market in June 2011. Because interest rate contracts are highly traded among OTC market participants, especially end-users, this Note examines how the CFTC approaches the regulation.

Part II of this Note provides an overview of the OTC derivatives market. It introduces the different kinds of contracts traded in the OTC market and elaborates on the type of risks associated with the OTC market. Part III examines who qualifies as an end-user and how they utilize the market. Part IV gives a brief history of past regulation in the OTC derivative market. It also introduces Title VII of Dodd–Frank and the requirements it imposes on the OTC market, especially end-users.

Although it is vital that commercial end-users have the ability to hedge risk with individualized contracts at minimized costs, a narrow exception with minimal regulation is not in the American public’s best interest. Part V introduces alternative approaches to regulating OTC derivatives. This Note argues that the regulation should exempt end-users, but that the current requirements of end-users are not sufficient to meet Title VII’s aims to promote transparency in the OTC market and to protect the American public from systemic risk. This Note tackles the question of how broad to make the end-user definition and how to then regulate those falling within the definition. This Note argues for a broader definition to accommodate the number of users who cannot afford to participate in the market if subjected to mandatory clearing requirements.

Part VI proposes that end-users should comply with stringent disclosure and reporting requirements in order not to exacerbate another financial crisis. The proposed disclosure model mirrors the ISDA’s Master Agreement, a standard form agreement already used by a number of market participants.


36. See Kraft Foods Letter, supra note 30.

37. See infra notes 131–34 and accompanying text (introducing the ISDA
The goal is to strike a balance between a broad end-user exemption and protecting the market from users abusing this exception through speculation and highly leveraged bets. This Note argues that such a balance is achievable so long as regulators demand more disclosure of end-users, particularly in regard to their ability to meet financial obligations. This will promote transparency and allow regulators to easily monitor abuse of the exception. While this Note acknowledges that the rapidly changing face of financial innovation in the OTC derivatives market makes it difficult for regulators to implement an efficient regulatory scheme, this Note’s proposal could nevertheless work by adapting disclosure requirements to changes in financial innovation.

II. Over-the-Counter Derivative Market

A. Market Overview

A derivative contract is “a bilateral contract or payments exchange agreement whose value derives . . . from the value of an underlying asset or underlying reference rate or index.” Derivatives are traded in two kinds of markets: exchanges and OTC markets. This Note focuses on the OTC market where contracts are bilaterally negotiated collateral agreements with flexible terms that mature over time.

The OTC market consists of privately negotiated and traded agreements. These market characteristics enable market participants to tailor derivative contracts to their specific needs.
This market appeals to commercial users hedging risk and noncommercial investors speculating in the market to obtain large profits. Although derivatives contracts allow participants to obtain large arbitrage profits and hedge exposed risks, they have a potentially large downside.

The resulting losses from Long Term Capital Management’s (LTCM) failure in 1998 and the AIG bailout in 2008 revealed this downside. These failures presented issues of counterparty credit risk. AIG sold credit derivatives in the OTC market that essentially insured corporate credit. AIG, a dealer of credit default swaps, was able to undertake a great amount of exposure in the OTC market by selling insurance on the risk of default where the underlying asset was a mortgage-backed security. Unfortunately, such contracts required large payments to counterparties when the subject of the credit default swap worsened. With the decrease in credit ratings in 2008 due to the downfall of the economy, AIG became responsible for posting collateral to numerous outstanding credit default swap contracts. Due to the multiple positions AIG held in credit default swaps, the company could not meet all of the collateral

43. See, e.g., EMILIOS AVLGOULEAS, THE MECHANICS AND REGULATION OF MARKET ABUSE 43 (2005) (listing why the derivatives market is attractive to different users).

44. See RECHTSCHAFFEN, supra note 5, at 160–62 (discussing the systemic risk effects of the failure of LTCM, Bear Stearns, and AIG due to their large speculative positions in the OTC markets).

45. See id. (describing the effect of one party defaulting on its contract due to failure and shifting the risk to the counterparty). This dislocation of risk has a great impact on the market, and will likely create a ripple effect for the economy as a whole. Id. See Gubler, supra note 5, at 87–88 (stating that AIG underwrote approximately $80 billion in notional amount of credit default swaps derived from mortgage-backed securities).

46. See RECHTSCHAFFEN, supra note 5, at 173–74 (providing a brief overview of how AIG was able to insure multiple times the value of the outstanding credit of the companies subject to the transaction). Market participants use credit default swaps to transfer credit risk to another party at a set cost. Id. at 179.

47. See Gubler, supra note 5, at 87–88 (discussing the role of OTC derivatives and AIG in the 2008 financial crisis).

48. See RECHTSCHAFFEN, supra note 5, at 173–74.

demands and turned to the government to help meet its obligations, avoiding a worse crisis than the nation was already facing.50

Participants in the OTC market, especially major participants and dealers, hold large positions.51 If one party was to default, the loss endured by the counterparties to all of their derivative contracts could disrupt market functioning.52 This fear of disruption is exactly why the government intervened on multiple occasions to bail out financial institutions and related entities.53 While speculation is not normally pertinent to end-users, this Note proposes the idea that minimally regulated end-users could manipulate the exception and speculate while claiming the hedging exemption.54 The general problem of distinguishing between hedging and speculation is a difficult task.55 In May 2012, JPMorgan Chase lost $2 billion over a six week period in a trading portfolio used to hedge risks to which the company exposed itself.56 This shows how transactions designed to hedge exposed risk could seem like speculation and result in large losses.57

50. See id. at 166 (stating that the only other alternative for AIG would have been bankruptcy, which would have left the counterparties to the swap transactions with no way to receive collateral payments).

51. See id. at 162 (stating that if the Federal Reserve did not facilitate the private sector recapitalization, then counterparties’ losses could have been somewhere from $3 billion to $5 billion).

52. See id.

53. See id. at 163–66.

54. See infra note 209 and accompanying text (discussing how the lack of disclosure proposed by the end-user exception does not effectively enable regulators to monitor the market for abuse of the exception).


57. See id. (“Concern [is] ‘that a large, supposedly sophisticated institution, even something called a ‘hedge’ can contain all kinds of hidden risks that the senior people don’t understand.’”).
Speculation in the derivatives market is a dangerous endeavor because the market exposes investors to a large amount of risk from highly leveraged bets. By imposing minimal regulation on end-users, transparency in the market is not achieved and bad actors can more easily hide trading activities from regulators. This is why regulators should subject these participants to more stringent disclosure and reporting requirements.

Another common characteristic of derivatives is the amount of leverage on which these instruments are traded. Leverage allows one to make an investment with little or no upfront monetary payment. In the OTC derivatives market, no exchange of funds may be required until maturity or performance. This enables an investor to hold exponentially larger positions than the amount committed. Although, in theory, this seems ideal because it reduces risk, in reality, leverage amplifies risk by spreading and multiplying it among multiple complex financial transactions, creating systemic hazards.

Because of these complex characteristics, derivatives can be extremely dangerous, not only to the OTC market, but to the economy as a whole. This potential danger stems largely from “users’ lack of knowledge and their under appreciation of the

58. See Lynn A. Stout, Why the Law Hates Speculators: Regulation and Private Ordering in the Market for OTC Derivatives, 48 DUKE L. J. 701, 772 (1999) (discussing how speculation in the derivatives market is sometimes cheap and encourages investors to accept uncompensated risk). Stout introduces the idea that derivatives speculation increases systemic risk because of the number of financial firms exposing themselves to a great level of risk. Id.

59. See Schapiro, supra note 17, at 165 (stating the importance that some sort of regulatory framework exists to monitor those who could abuse the OTC market and increase the threat of systemic risk).

60. See RECHTSCHAFFEN, supra note 5, at 164 (introducing the concept of leverage and how it is beneficial to transactions in the derivatives market).

61. See id. (pointing out how an OTC derivative may not require any advancement until maturity, whereas an exchange-traded derivative requires a margin). See infra notes 165–68 and accompanying text for a description of margin.

62. See Topham, supra note 4, at 139–40 (discussing how financial leverage in the derivatives market can lead to an increased ability to take risks, therefore raising the chances of failure and systemic risk).

63. Id. at 140.
risks involved.”64 This factor caused a number of losses stemming from derivatives use, including the economic crisis of 2008.65

B. Types of OTC Contracts

1. Options

An option contract gives “the purchaser the right to buy (call option) or sell (put option) a specified quantity of a commodity or financial asset at a particular price (exercise price) on or before a certain future date.”66 Option contracts function by having the purchaser pay the seller (writer) an option premium for the right to buy or sell.67 The purchaser’s loss is limited to the price of the premium, enabling the purchaser to limit the downside of investment.68 In contrast, the seller of an option receives the premium in return for risk exposure.69

Like other derivatives, options may serve as hedging and speculating instruments, but, most importantly, options enable a buyer to eliminate all downside risk by paying a premium upfront.70 Options do not expose the owner of an option to market risk because they create only a right to buy or sell, not an obligation.71 Options do expose the seller to some risk associated

---


65. See id. at 629–30 (stating how users’ lack of knowledge caused a number of financial losses in the 1990s, but these losses resulted in few lawsuits).

66. See GAO REPORT, supra note 9, at 27.

67. Id.

68. See RECHTSCHAFFEN, supra note 5, at 170–71 (discussing the operation and use of options).

69. See id.

70. See id. (noting that downside risk of an option is shifted completely to the seller).

71. See id. at 171 (explaining that the owner of an option is exposed to no market risk and, if they choose not to exercise an option, then they only lose the price of the premium).
with the underlying asset. The seller can hedge this risk, however, by investing in the underlying asset.

2. Forwards

A forward contract “obligate[s] the holder to buy or sell a specific underlying [asset] at a specified price, quantity, and date in the future.” Forwards are customized contracts that allow market participants to hedge assets and liabilities by locking in a future purchase or sale price. For example, if a U.S. importer plans to buy a product at a future date for a price quoted in a foreign currency, the U.S. importer can enter into a forward contract to fix the U.S. dollar cost of the product. This allows the U.S. importer to hedge against currency fluctuations between the purchase and delivery dates. Alternatively, market participants use forwards to speculate on market movements and profit from decreases or increases in future prices or rates.

3. Swaps

A swap is a complex instrument traded on the OTC market. In a swap agreement, “two parties agree to exchange ‘cash flows’ on a ‘notional amount’ over a period of time in the future.” This notional amount is a reference upon which the payment stream is

72. See id. (looking to how sellers of an option contract can limit their exposure to market risk).
73. See id.
74. See GAO REPORT, supra note 9, at 26 (defining a forward and describing the instrument’s use in the market).
75. See id. at 26 (providing an example of how a hedging participant could benefit from using a customized forward contract).
76. See id. at 5 (providing an example of how a party would utilize a forward contract).
77. See id.
78. See id. at 27 (discussing how a speculator attempts to profit in the forwards market).
79. See RECHTSCHAFFEN, supra note 5, at 172.
80. See id. (looking at swap agreements and discussing the characteristics of such contracts).
A commonly transacted swap is the interest rate swap, which allows counterparties to exchange a fixed rate for a floating rate. For example, Party A invests in a bond with a par value of $1000 and a fixed interest rate of 5% compounded annually. Party B, however, owns a bond that pays a floating interest rate tied to the London Interbank Offered Rate (LIBOR). Party B may prefer a more steady interest payment to hedge against the risk of a decreasing interest rate. Party A, however, may want to trade his steady interest payment for a potentially larger interest payment. If the two enter into an interest rate swap, then Party A is guaranteed his annual interest payment of $50 plus any additional profit over $50 on the two bonds. Party B, on the other hand, is always guaranteed $50. Party B pays any amount over $50 to Party A, while Party A, in turn, compensates Party B for any loss under $50. Because swaps are traded on the OTC market, parties can negotiate a contract that specifically reflects their needs, whether

---

81. See id. at 175 (describing the elements of a swap transaction prior to an overview of the types of swap transactions most frequently used in the market). Common transactions include interest rate swaps, currency swaps, and credit-default swaps. Id.

82. See id. at 175–77 (setting out how a “plain vanilla” interest rate swap works and provides benefits to the counterparties); see also Gogel, supra note 5, at 8–9 (providing a hypothetical example of how an interest rate swap works). These described transactions are constructed similarly to other swap contracts. Id.

83. See Gogel, supra note 5, at 8–9 (depicting this example).

84. See id. (Party B has a floating interest rate payment and wants to guarantee a specific payment so he transacts with Party A, who wishes to increase his guaranteed payment). For a description of LIBOR, see British Banking Ass’n, Understanding BBA LIBOR, (Jan. 8, 2011), http://www.bba.org.uk/media/article/understanding-bba-libor (last visited Sept. 27, 2012) (on file with the Washington and Lee Law Review). LIBOR is the London Interbank Offered Rate, which reflects the rate at which banks, such as the U.S. Federal Reserve and the European Central Bank, borrow money from one another daily. Id.

85. See Gogel, supra note 5, at 8–9.

86. Id.

87. Id.

88. Id.

89. Id.
to hedge exposed credit risk from an owned asset or to profit by speculating on pricing inefficiencies in the market.90

Swaps enable extreme flexibility to meet the contracting parties’ needs, but, concurrently, expose these parties to high credit risk.91 Credit risk is the risk that a counterparty to the transaction will default on the contract.92 Ideally, credit risk is reflected in the contractual terms, but, due to information asymmetry, this is not always achievable.93 AIG’s issuance of multiple credit default swaps and inability to meet financial obligations when they became due provides a picture of how asymmetrical information can have damaging effects.94 After the 2008 financial crisis, regulators set out to address these risks and pricing inefficiencies within the OTC swap market.95 The next portion of this Note examines commercial end-users’ role in the swap market and how proposed legislation may affect their participation.

III. End-User Participation in the OTC Market

End-users are a category of market participants that utilize the OTC market to hedge exposed market risk and minimize volatility of their overall earnings.96 This category mainly

90. See Mark A. Guinn & William L. Harvey, Taking OTC Derivative Contracts as Collateral, 57 BUS. LAW. 1127, 1132 (2002) (explaining that OTC derivatives are used for both speculation and hedging, but they are more commonly used for hedging purposes).
91. See Rechtschaffen, supra note 5, at 172.
92. See Adams & Runkle, supra note 64, at 663 (providing a description of credit risk and how to manage it).
93. See Gogel, supra note 5, at 5 (discussing how lack of transparency in the market and nondisclosure between counterparties leaves regulators and participants unaware of potential risks building up in the financial market). This is a problem that Title VII of the Dodd–Frank Act aims to address. See Schapiro, supra note 17, at 165.
94. See supra notes 44–50 and accompanying text for a brief overview of the AIG bailout.
95. See Schapiro, supra note 17, at 164 (discussing the goals of Dodd–Frank with respect to regulating the OTC derivative market).
96. See Practical Derivatives: A Transaction Approach 10 (Jonathan Denton ed., 2006) (“Derivatives are typically used by corporate end users to reduce or extinguish their exposure to discrete risks and thus reduce the volatility of their earnings.”); Adams & Runkle, supra note 64, at 674 (“When
includes nonfinancial corporations, but asset managers and other financial institutions also qualify. These users, and their investors, benefit greatly from using the OTC market to hedge risk.

For example, Shell Co. takes positions in the U.S. energy derivatives market to respond to internal forecasts of supply and demand, enabling Shell to be ahead of foreseeable price movements. It also participates in the swap market to offset credit risk and assist actual transactions. Kraft Foods employs risk management strategies to handle risks associated with volatility in interest rates, commodity prices, and foreign currency rates that entail entering into forward, option, and swap contracts.

Although a positive duty to hedge exposed risks may not exist, at least one court found that the board of directors of a company owes its shareholders a duty of care to instruct managers adequately about the use of hedging with derivatives. Such a ruling demonstrates that hedging in the derivatives market is a common practice among corporations, and it is in a board’s best interest to examine whether it is beneficial for a corporation to take this risk management route.

properly used to hedge against risk, derivatives are an essential corporate tool."

97. See ISDA Survey: Interest Rate Swaps, supra note 35 (providing a list of categories of survey respondents).


99. See id.

100. See Kraft Foods Letter, supra note 30 (describing how Kraft Foods and its affiliate companies utilize the derivative market).

101. See Brane v. Roth, 590 N.E. 2d 587, 591–93 (Ind. Ct. App. 1992) (concluding that the board of directors failed “to attain knowledge of the basic fundamentals of hedging to be able to direct the hedging activities and supervise the manager properly”).

102. See PRACTICAL DERIVATIVES, supra note 96, at 13 (discussing the court’s finding in Brane v. Roth and a corporation’s duty to hedge or, at a minimum, look to the mechanics of such a trade to determine if it is within the best interests of the corporation).
Companies, as mentioned previously, use the derivative market to manage risk and optimize value.\textsuperscript{103} The OTC market enables end-users to tailor their derivative contracts to their needs.\textsuperscript{104} High negotiation costs of OTC derivatives contracts makes trading on the market costly.\textsuperscript{105} But this is the price end-users pay for a uniquely tailored contract. Prior to Dodd–Frank, participants in the market enjoyed the added benefit of no margin requirements.\textsuperscript{106}

According to a 2010 International Swaps and Derivatives Association (ISDA) survey, end-users stated that they use the OTC derivatives market primarily for interest rate swaps, currency swaps, credit default swaps, equity swaps, and energy swaps.\textsuperscript{107} The results show that 80% use the OTC market for interest rate swaps, 59% for currency swaps, 27% for credit default swaps, 25% for equity swaps, and 32% for energy/commodity swaps.\textsuperscript{108} The ISDA classified end-users as companies utilizing the market to manage exposed risk.\textsuperscript{109} The group included nonfinancial corporations, asset managers, and other financial institutions.\textsuperscript{110} These instruments are so widely

\begin{itemize}
\item \textsuperscript{103} See Kraft Foods Letter, \textit{supra} note 30; see also Shell Letter, \textit{supra} note 98 (discussing for what purposes Shell uses the derivative market and how the proposed end-user definition will affect Shell’s ability to utilize the market via its subsidiary companies).
\item \textsuperscript{104} See Thomas C. Singher, Note, \textit{Regulating Derivatives: Does Transnational Regulatory Cooperation Offer a Viable Alternative to Congressional Action?}, 18 \textit{Fordham Int’l L.J.} 1397, 1406–08 (1995) (addressing the uses of the derivatives market to commercial users). One of the uses discussed is speculation, which, although not common for end-users, supports an argument to regulate these participants in some fashion. \textit{Id.} at 1410–11.
\item \textsuperscript{105} See Gogel, \textit{supra} note 5, at 9 (stating that swap contracts tend to have high transaction costs due to the extensive negotiations over delivering a party’s specific needs).
\item \textsuperscript{106} See \textit{infra} notes 122–28 and accompanying text for a description on Dodd–Frank’s margin requirements.
\item \textsuperscript{107} See \textit{ISDA Survey: Interest Rate Swaps, supra} note 35, at 2 (listing results from a survey of 295 respondents from North America and Europe who use the OTC derivatives market).
\item \textsuperscript{108} See \textit{id.}
\item \textsuperscript{109} See \textit{id.} The ISDA study includes more than just nonfinancial entities utilizing the OTC market for hedging exposed risks. \textit{Id.} This Note proposes that the regulatory definition of end-user should include those surveyed as end-users by the ISDA.
\item \textsuperscript{110} See \textit{id.}
\end{itemize}
used because companies across all industries utilize interest rate
swaps and currency swaps. Specifically, industries utilize
commodity, equity, and credit derivatives. For example,
financial industries primarily use equity and credit derivatives,
while companies focusing on utilities and basic materials utilize
commodity derivatives.

Because end-users typically use the OTC market specifically
for hedging exposed market risk, Congress excluded these users
from regulation under Title VII of Dodd-Frank. If mandatory
clearing pertained to end-users, they would be required to post
margin and use clearinghouses for all of their trades. This
would increase the cost of using the OTC market and drive many
end-users away from the market. As a result, corporate end-users
would hedge their risks in nonbeneficial ways, such as through
insurance contracts that increase transaction costs, which the
company passes on to the consumers and investors.

Although increased costs will drive end-users from the
market, imposing minimal regulation for end-users is not the
proper solution. It is possible that end-users are attracted to the
market not only because of low transaction costs, but also because
of the lack of regulation. Although most end-users are utilizing
the market for hedging purposes, misusing these instruments is
easy due to their complexity and large potential gains.

111. See ISDA News Release, supra note 26, at 2 (breaking down results of
a 2009 survey of end-user participation in the OTC derivative market by
industry and country).

112. See id. at 4 (displaying a chart of the survey results broken down by
industry sector and usage of each type of swap contract).

113. See.

114. See Ariail, supra note 12, at 189 (explaining that Congress’s rationale
for excluding end-users from regulation stemmed from concerns for consumers
who would suffer through increased costs due to corporations’ inability to hedge
risk and reduce overall losses).

115. See Dodd-Frank § 723, 7 U.S.C. § 2(h); infra Part IV.B (discussing Title
VII of Dodd-Frank).

Barney Frank & Rep. Colin Peterson (June 30, 2010) (stating Congress’s
intention in exempting end-users from the market) (on file with the Washington
and Lee Law Review).

117. See Guinn & Harvey, supra note 90, at 1128, 1130 (commenting on the
possibility of abuse in the market due to some participant’s lack of knowledge of
the complex OTC derivatives).
example, one claiming the end-user exception could determine a way to abuse the exception by disguising speculation as hedging and increasing the threat of systemic risk.\(^ {118}\) Does exemption from Dodd–Frank mandatory clearing provide end-users the best of both worlds—virtually no regulation and low transaction costs? This Note will analyze whether it is possible to broaden the class of end-users and keep the derivatives market attractive to end-users by maintaining low transaction costs.

**IV. Regulation of the OTC Market**

**A. Past Regulation**

Before introducing the current regulation of the OTC market, it is important to discuss briefly the evolution of regulation in the derivatives market, specifically of OTC derivatives. The regulatory framework of the derivatives market is a combination of the CFTC, the SEC, and a number of self-regulatory institutions.\(^ {119}\) Although traders utilized OTC derivatives for a long period of time, the market rose to popularity in the 1980s.\(^ {120}\) Due to regulators’ unfamiliarity with these new, complex instruments, OTC derivatives remained largely unregulated for some time.\(^ {121}\) Originally, regulators focused on preventing manipulation and fraud in the derivatives market.\(^ {122}\) Because OTC derivatives are less susceptible to manipulation, regulators did not see these instruments as a threat deserving of regulation.\(^ {123}\) Unfortunately,

---

118. *See infra* Part VI.


120. *See* Gary Gensler Testimony, *supra* note 7 (addressing why the OTC market was unregulated prior to 2008).

121. *See* Krippel, *supra* note 119, at 278.

122. *See* Gogel, *supra* note 5, at 18–23 (discussing the history of derivative regulation from the passage of the CEA in 1936 to the creation of the CFTC in 1974). Congress attempted to address issues of fraud and manipulation of the unregulated market. *Id.* Ultimately, this led to exempting the OTC derivatives from regulations in the 1992 amendments to the CEA, which attracted many critics. *Id.* at 23.

123. *See* U.S. GOVT ACCOUNTABILITY OFFICE, GAO-00-00, *Commodity
the 1992 amendments to the 1936 Commodity Exchange Act (CEA), which governs regulation of the derivatives market, did not completely resolve this issue.\textsuperscript{124} The 1992 amendments gave the CFTC power to exempt OTC derivatives from regulation.\textsuperscript{125} Regulators justified nonregulation with the fact that participants were self-interested and sufficiently sophisticated to self-regulate.\textsuperscript{126}

Because government regulators failed to address issues beyond manipulation and fraud, private regulators began appearing in the 1990s to increase market transparency and disclosure.\textsuperscript{127} In 1994, the Derivatives Policy Group introduced a voluntary oversight framework that would help address the public policy issues of the OTC derivatives market.\textsuperscript{128} Its goal was to have more OTC market participants report their use and risk exposure in the market.\textsuperscript{129}

The Financial Accounting Standards Board (FASB) and the ISDA have made progress in making the derivatives market more transparent and efficient.\textsuperscript{130} The ISDA uses a form document, known as the ISDA Master Agreement, which parties sign prior to entering into any derivatives contract.\textsuperscript{131} The ISDA Master Agreement...

\textsuperscript{124} See id. Because the OTC derivative market was not as susceptible to fraud and manipulation, regulators were not concerned with subjecting them to regulation. See id. at 28.

\textsuperscript{125} See id.

\textsuperscript{126} See Gary Gensler Testimony, supra note 7.

\textsuperscript{127} See Singher, supra note 104, at 1431–34 (looking at all the different private actors who have taken steps to help promote efficiency and transparency in the OTC market).


\textsuperscript{129} See id.

\textsuperscript{130} See Singher, supra note 104, at 1431–33 (discussing additional regulatory influences on the OTC derivatives market beyond statutory guidelines).

Agreement specifies a number of things, including the obligations of the parties and the relevant events of default. If the standard form contract is unfavorable to one party, it can either negotiate this term in each of its contracts or accept the standard terms. This practice promotes efficiency for members, and the documents further provide guidance in judicial decisions.

While private groups were encouraging self-regulation, some advocated for statutory regulation of the OTC market. Though OTC derivatives did not pose a risk of manipulation in the late 1990s when the argument to regulate was introduced, some saw that the rapidly changing structure of this market and the recent large losses in the market required a regulatory response. The Government Financial Officers Association presented the idea that derivatives posed a bigger threat than the benefit they offered. Most notably, Brooksley Born, Commissioner of the CFTC from 1996–1999, testified before Congress on the perils that the derivatives market posed and how regulators should address these dangers.

132. See id. at 6 (outlining what the ISDA Master Agreement and Schedule specifies to reduce counterparty risk).
134. See id. at 6–10 (presenting the idea that the ISDA’s private law, addressing issues unanticipated by judges and regulators, will continue to be used in judicial opinions).
135. See CFTC Concept Release, Over-the-Counter Derivatives (May 7, 1998), http://www.cftc.gov/opa/press98/opamtn.htm (last visited Sept. 27, 2012) (proposing a regulatory scheme for the OTC market) (on file with the Washington and Lee Law Review). This proposal was opposed by financial regulators, Greenspan and Rubin, who essentially stated that such regulation would cause a financial crisis. See Krippel, supra note 119, at 279 (discussing the negative reaction to Brooksley Born’s regulatory proposal).
Failure to keep pace with the changing market would stifle the capacity of U.S. firms to meet global competitive challenges, would create a cloud of legal uncertainty over the applicability of outdated rules to new products and innovative transactions, and would erode the regulatory system’s ability to protect customers and to preserve the financial integrity of that market.139

Her proposal considered the clearing of derivatives, OTC derivatives market users’ registration and reporting to the CFTC, capital requirements, requirement of risk management controls for derivatives dealers, and restrictions on dealers’ sale practices.140 This idea met strong opposition from other regulators who believed that regulation would hinder the efficiency of the OTC derivative market.141

Born’s fear became real in 2008. But, in 1999, proponents for deregulation won the day in Congress. Alan Greenspan, then Federal Reserve Chairman, claimed that the proposed regulatory scheme would “distort the efficiency of [the U.S.] market system and as a consequence [impede] growth and improvements in standards of living.”142

With Greenspan’s encouragement, Congress rejected regulation of OTC derivatives and passed the Commodity Futures Modernization Act of 2000 (CFMA).143 With this Act, Congress ensured that the CFTC would not regulate OTC derivatives by expressly exempting them from the CFMA legislation.144 Essentially, the CFMA’s goal was to promote innovation, enhance legal certainty, and provide greater stability in the derivatives

139. Id.
140. See CFTC Concept Release, supra note 135.
142. See id.
144. See 7 U.S.C. § 2(q)(1) (2010) (exempting agreements entered into by eligible contract participants that are “subject to individual negotiation by parties”).
market. As a result, private parties could negotiate OTC derivatives contracts without being subject to any regulation.

Some feel that perhaps Born and Greenspan were both wrong in their regulatory endeavors. Lynn Stout argues that the deregulation of derivatives led to disaster by making swap contracts legally enforceable through the CFMA, which was unprecedented. Stout proposes a form of self-regulation that would hold dealers personally responsible for risky, speculative decisions, by making speculative OTC contracts not legally enforceable. This is how the law viewed derivatives contracts prior to 2000 and Stout believes this is the best way to regulate now. In essence, Lynn Stout’s answer is to treat an OTC swap contract as a gambling contract, just as it was once done centuries ago. This Note will address this idea further in Part V.B.1.

B. Dodd–Frank Regulation

For years, regulators recognized that the OTC market lacked oversight. In 1994, the Comptroller General of the United States called for uniform regulation of the OTC market.

145. See Gogel, supra note 5, at 24 (discussing the intention and effect of the CFMA).

146. See id. (stating the result of deregulation in the OTC market). Gogel introduces the idea that deregulation of OTC transactions directly contrasts with the intention of the CFMA to promote transparency, decrease systemic risk, and provide stability in the market. Id.

147. See Stout, Deregulating Derivatives, supra note 6, at 30 (“[The] rule of unenforceability encouraged speculators to rely on private ordering and to develop and police their own private markets.”).

148. See id.

149. See id; Lynn A. Stout, Regulate OTC Derivatives by Deregulating Them, BANKING & FINANCE 34 (Fall 2009) [hereinafter Regulate OTC Derivatives] (proposing an idea to treat speculative contracts as gambling contracts to give dealers more accountability for their actions).

150. See Stout, Regulate OTC Derivatives, supra note 149, at 34–35; see infra notes 196–99.

151. See Gogel, supra note 5, at 31–32 (“On October 8, 2008, Christopher Cox, then Chairman of the SEC, characterized the lack of oversight of the OTC derivatives market as a ‘regulatory black hole.’”).

152. See GAO REPORT, supra note 9, at 126–27 (“Given the weaknesses and gaps that impede regulatory preparedness for dealing with a crisis associated
1998, Brooksley Born reaffirmed the importance of regulating this vast market. Unfortunately, the government did not act to regulate the OTC market until it was too late.

Title VII of Dodd–Frank addresses the regulation of the OTC derivatives market, particularly swaps. The legislation’s overall goal is to lower risk to the American public and to promote transparency in the OTC market. The major rules that the CFTC and SEC are currently promulgating are set out below.

Dodd–Frank gives the CFTC and the SEC jurisdiction over swap regulation. The SEC has jurisdiction over security-based swaps and the CFTC has exclusive jurisdiction over all nonsecurities based swaps, including interest rate swaps and currency swaps. Dodd–Frank requires the CFTC to coordinate with the SEC and other agencies prior to issuing rules or orders in connection with swap regulation. Dodd–Frank also mandates that the CFTC and SEC engage in joint rulemaking to define terms, including “swap,” “security-based swap,” “swap dealer,” and “major swap participant.”

One of the major changes that Dodd–Frank imposes on the OTC market is the mandatory clearing of swaps. This regulatory approach requires all qualifying swaps to be traded through a heavily regulated third-party, called a derivatives clearing organization (DCO). This aims to reduce systemic risk with derivatives, we recommend that Congress require federal regulation of the safety and soundness of all major U.S. OTC derivatives dealers.

---


154. See id. at 40 (introducing different sections of the Financial Reform Act and what each section aims to regulate).


156. Dodd–Frank Act § 722(a).

157. Id.

158. Id. § 712.

159. Id. § 712(d)(1). The definitions of these terms are finalized. See 77 Fed. Reg. 30596 (May 23, 2012) (to be codified at 17 C.F.R. pt. 1).

160. See Gogel, supra note 5, at 48–50 (looking at the costs and benefits of derivatives clearing).

161. See Gubler, supra note 5, at 86–87 (discussing approaches to managing counterparty risk and the effects of these approaches, specifically the central clearing requirement).
DO END-USERS GET THE BEST OF BOTH WORLDS?

by putting OTC transactions on the books of regulated clearinghouses and not large financial institutions. Central clearing also aims to reduce counterparty risk, leading to more accurate pricing.

Central clearing will require participants, with some exceptions, to set initial margin requirements. Margin is a cash balance in a trader’s account with the clearinghouse. If the balance in this account falls below a certain amount, the trader must post additional collateral to maintain a balance established by the clearinghouse. A portion of the contract settlement amount determines the margin requirement. This guarantees that parties back up investments with an adequate amount of capital. Due to uncertainty in derivatives pricing and fluctuation of an underlying asset’s price, a trader may be required to post a considerable amount of collateral. While providing upfront collateral is important, it is an incredibly thorny topic because of the difficulty in pricing derivatives and determining an accurate settlement price on which to base a margin calculation.

162. See OTC Derivatives Markets Act of 2009 Hearing, supra note 155, at 4 (addressing how the clearing organizations will help to lower risk to the American public).
163. See Gubler, supra note 5, at 92 (explaining how the DCO in the OTC market will overcome counterparty risk by providing an example). Although the DCO is designed to eliminate counterparty risk and increase transparency, the use of a central party clearing may increase costs in obtaining information that in the current bilateral framework is not public. Id. at 94.
164. See Ariail, supra note 12, at 185 (providing an overview of central party clearing in the OTC market).
165. See id. at 185 n.74 (explaining briefly the basics of margin).
166. See id.
167. See Gogel, supra note 5, at 50 (explaining the purpose of margin requirements).
168. See id.
170. See Gubler, supra note 5, at 105–06 (examining the problems collateral requirements may pose for OTC market users).
These regulations apply only to swap dealers and major participants.\(^\text{171}\) Congress's goal in enacting this section of Dodd–Frank was to guarantee that margin and collateral requirements would not hinder end-users who utilize the market for pure hedging purposes.\(^\text{172}\) Congress wanted these participants to use funds for business investment and job creation, rather than margin requirements.\(^\text{173}\) Determining which participants utilize the market for these reasons and, consequently, whom to exempt from regulation is a problem.\(^\text{174}\) The next portion of this Note examines the end-user exception, which carries out this goal of Congress.

V. End-User Exception to the Mandatory Clearing of Swaps

Dodd–Frank gives the CFTC authority to adopt rules exempting certain swaps from the clearing mandate.\(^\text{175}\) Aiming to protect the American public by regulating systemic risk in the OTC market, the CFTC promulgated a rule defining the end-user.\(^\text{176}\)

\(^{171}\) See Dodd–Frank § 723(a)(3); Gogel, supra note 5, at 45–46 (discussing the Dodd–Frank regulations of the OTC swap market).

\(^{172}\) See 156 Cong. Rec. S5905 (daily ed. July 15, 2010) (statement of Sen. Blanche Lincoln) (stating that the legislation may not require some OTC market users to post margin, it will require them to satisfy public reporting requirements as set out in Dodd–Frank).

\(^{173}\) See John Carney, Will the CFTC Kill the End User Exemption?, CNBC NETNET (Oct. 4, 2010), http://www.cnbc.com/id/39506763/Will_The_CFTC_Kill_The_End_User_Exemption (last visited Sept. 27, 2012) (discussing whether the CFTC has the authority to subject end-users to margin requirements) (on file with the Washington and Lee Law Review).

\(^{174}\) See Ariail, supra note 12, at 194 (“[T]he most problematic part of policing the end-user exemption . . . is distinguishing when an end-user is ‘hedging or mitigating commercial risk’ rather than taking a speculative position in derivatives markets.”).

\(^{175}\) See End-User Exception to Mandatory Clearing of Swaps, 75 Fed. Reg. 80747 (proposed Dec. 23, 2010) (to be codified at 17 C.F.R. pt. 39) (“[T]he Dodd–Frank Act provides the Commission with the authority to adopt rules governing the end-user clearing exception and to prescribe rules, issue interpretations, or request information from persons claiming the end-user clearing exception necessary to prevent abuse of the exception.”).

\(^{176}\) See id.
DO END-USERS GET THE BEST OF BOTH WORLDS? 1785

The end-user exception is limited to nonfinancial entities and does not completely eliminate their transactions from regulation.177 A nonfinancial entity includes users who do not hold investment positions for profit in the OTC market.178 The rule disqualifies any swap used for any speculative or trading purposes, or if used to hedge another swap.179 The rule requires that a nonfinancial entity claiming exempt status must disclose to the Commission its intention of claiming the exemption and how it generally plans to meet financial obligations.180 Some commenting on the proposed rule feel this regulation is burdensome and will not effectively increase price transparency in the swap market.181 Companies, like Shell, argue that overly aggressive regulation could undermine efficiencies appreciated in the OTC market and could distort pricing without promoting transparency.182

A. Rule Design: Narrow Versus Broad

By passing Title VII of Dodd–Frank, Congress wanted to increase transparency in the OTC market and decrease the threat of systemic risk.183 Regulators aim to promote public transparency in addition to transparency to regulators.184 The CFTC hopes that public transparency will improve the function of the OTC market, much as it has existing securities and futures

179. See 75 Fed. Reg. 80747 (explaining that the Commission finds all swaps held for appreciation in value to be speculative).
180. See id. (proposing that end-users notify the Commission each time the clearing exception is elected by providing specified information as set out in the swap data recordkeeping and reporting rules).
181. See Shell Letter, supra note 98, at 5 (explaining how compliance with data reporting for end-users will not provide an accurate portrayal of one’s exposure in the market).
182. See id. at 2 (stating that real time reporting requirements for swaps between an affiliate and a corporation do not promote the aims of Dodd–Frank).
183. See Schapiro, supra note 17, at 166.
184. See Gary Gensler Testimony, supra note 7 (discussing the Dodd–Frank legislation and what it aims to promote in the OTC market).
markets.\(^{185}\) Although Title VII of Dodd–Frank regulates the vast and dangerous OTC market, Congress did not want the SEC and CFTC to regulate certain users.

The end-user exception rule allows for commercial end-users to utilize the OTC derivatives market without the added cost of margin requirements and central clearing.\(^{186}\) Section 2(h)(7) of the Commodity Exchange Act (CEA), as amended by Dodd–Frank, provides:

> [T]hat a swap otherwise subject to mandatory clearing is subject to an elective exception from clearing if one party to the swap is not a financial entity, is using swaps to hedge or mitigate commercial risk, and notifies the Commission, in a manner set forth by the Commission, how it generally meets its financial obligations associated with entering into non-cleared swaps.\(^{187}\)

Commentators to the end-user exception raised questions pertaining to a participant who does not qualify as a nonfinancial entity, but (a) cannot afford to meet hedging requirements, such as small banks, which are not systemically significant, and (b) uses the OTC market to hedge or mitigate risk, such as corporations’ affiliate companies.\(^{188}\) These commentators were concerned that they would not qualify for the end-user exemption and therefore must comply with mandatory clearing requirements or leave the market.\(^{189}\)

In response to these concerns, Congress introduced a bill that would require the CFTC to exempt inter-affiliate swaps from mandatory clearing.\(^{190}\) Inter-affiliate swaps are swaps that have

---

185. See id.


188. See Letter from Nat’l Rural Util. Cooperative Fin. Corp., to David A. Stawick, Secretary of the Comm’n, CFTC (Jan. 12, 2011) [hereinafter NRUCFC Letter] (addressing the issue of small financial entities utilizing the OTC market for reasons and in ways far different from large financial lenders, which would justify the small lenders inclusion in the end-user exception) (on file with the Washington and Lee Law Review); See Kraft Foods Letter, supra note 30, at 9 (arguing that affiliate corporations, formed to act as a centralized hedging center to hedge exposed risk for an entire corporate group, should be included in the CFTC’s end-user definition).

189. See id.

190. See To Exempt Inter-affiliate Swaps from Certain Regulatory
a corporation on one side of the transaction and, on the other side, a party that is controlled or under common control of the corporation that is a counterparty to the transaction. This bill has yet to be enacted and may force the CFTC to broaden its narrowly proposed definition.

The problem this poses is that the current design of the rule does not support a broad end-user definition. The CFTC defines end-user to include more participants, such as small banks, but does not subject the users to more regulatory requirements. Congress’s intention is to exempt end-users from the mandatory clearing of swaps to maintain low transaction costs, but imposing minimal regulatory requirements only hampers regulators’ efforts in creating a more transparent market. This Note argues that more disclosure is necessary for the end-users exempted from the mandatory clearing of swaps.

B. Alternative Rule Design

Based on the difficulty of pricing derivatives, the complexity of calculating margin requirements, and the uncertainty of financial innovation, can regulators effectively increase market transparency and decrease systemic risk in the OTC derivative market? This Note argues that more disclosure will help increase transparency, but that regulators will have difficulty in maintaining transparency with the ever changing landscape of financial innovation.


192. See infra Part V.B.1.
1. Alternative Approaches to OTC Derivatives Regulation

The CFMA’s deregulation of the OTC derivatives market proved ineffective, as it is alleged to be a cause of the 2008 financial crisis.\textsuperscript{193} By flying under the radar of regulators, the OTC market grew to an extraordinary size within the years leading up to the financial crisis.\textsuperscript{194} Title VII of Dodd–Frank aims to protect the American public from the systemic risk that the OTC market poses by increasing transparency and disclosure in the market.\textsuperscript{195} Part IV.B. and Part V examine Title VII’s efforts to achieve these goals. This Note now introduces alternative, perhaps better, approaches to derivatives regulation.

First, as mentioned in Part IV.A, Lynn Stout proposes that regulators should treat OTC derivatives contracts as gambling contracts.\textsuperscript{196} Traditionally, speculative derivatives contracts were legally unenforceable wagers.\textsuperscript{197} Stout argues that because derivatives are technically bets on future market conduct, it makes sense for regulators to treat these contracts as gambling contracts.\textsuperscript{198} It encourages participants to be more careful in choosing counterparties and to take responsibility for their bets.\textsuperscript{199}

Second, voluntary reporting institutions, such as the ISDA and the Derivatives Policy Group mentioned in Part IV.B., allow

\begin{itemize}
  \item \textsuperscript{193} See Topham, supra note 4, at 147–49 (discussing how the CFMA’s deregulation of derivatives resulted in participants not being aware of the underlying risks associated with these instruments).
  \item \textsuperscript{194} See id. at 148 (stating the notional amount of the derivatives market was estimated at $604 trillion in 2009, with major commercial banks holding positions that totaled $204 million).
  \item \textsuperscript{195} See Schapiro, supra note 17, at 164 (listing the aims of the Dodd–Frank legislation).
  \item \textsuperscript{196} See supra notes 147–50 and accompanying text (introducing Lynn Stout’s theory on regulating derivatives as gambling contracts).
  \item \textsuperscript{197} See Stout, Regulate OTC Derivatives, supra note 149, at 30 (discussing how derivatives were regulated prior to the passage of the CFMA in 2000).
  \item \textsuperscript{198} See id. at 30–31 (stating that, for centuries, speculative derivative contracts were treated as legally non-enforceable contracts just as gambling contracts due to the level of risk in the bet).
  \item \textsuperscript{199} See id. at 32 (presenting the idea that speculative participants in the derivatives market are encouraged to choose counterparties wisely and to create a private market on which speculative participants trade, therefore minimizing systemic risk).
\end{itemize}
for participants to trade under standardized contracts. This enables a more concrete report of activity in the OTC market. These institutions have made progress in reducing counterparty risk, but unfortunately could not prevent the 2008 financial crisis.

A third approach would regulate the conduct of users and not the specific instruments. This view looks to the major issues of the OTC market that regulators must address, namely counterparty risk and financial innovation. This approach seems similar to Dodd–Frank in that the legislation carves out an exception for end-users who utilize the market for hedging and requires speculators in the market to meet collateral requirements and mandatory clearing.

The final regulatory method builds on the last approach. It proposes that regardless of the conduct of the participant, all participants should adhere to disclosure requirements greater than just checking a box, as the current proposed rule sets out for end-users. The CFTC must set more stringent disclosure requirements so regulators can monitor the market for abuse of the end-user exception. Regulators should require end-users to

---

200. See supra notes 128–34 and accompanying text for an introduction of these institutions and their policy goals.

201. See id.


203. See Willa E. Gibson, Are Swap Agreements Securities or Futures?: The Inadequacies of Applying the Traditional Regulatory Approach to OTC Derivatives Transactions, 24 J. CORP. L. 379, 414 (1999) (introducing the idea to regulate OTC derivatives by conduct, i.e. speculation and hedging, and not by the type of instrument).

204. See id. at 412–15 (regulating conduct helps address major concerns of the OTC market while “achiev[ing] a balance between market efficiency and market integrity”).

205. See supra Part V.A.

206. See 75 Fed. Reg. 80747; Letter from Michael Greenberger, Professor, Univ. of Md. Sch. of Law, to David A. Stawick, Sec., CFTC (Feb. 22, 2011) [hereinafter Greenberger Letter] (proposing to the CFTC that the end-user exception rule should enhance disclosure requirements to be more effective and align with the aims of Dodd–Frank).

207. See Greenberger Letter, supra note 206, at 3 (stating that a “check-the-box” approach is inadequate because it does not give regulators the proper
give more specific detail of how they will meet financial obligations, rather than just a general overview as the rule suggests. By adding this disclosure and other reporting requirements that provide a detailed outline of how end-users are hedging risk, regulators would be aware if end-users are adequately mitigating risk and not abusing the exception with speculation.

VI. Amendment to the End-User Exception

The mandatory clearing of swaps aims to protect the American public from a financial crisis like that of 2008, but exempting end-users from other forms of regulation does not entirely meet the goals of Dodd–Frank. Exempting end-users from mandatory clearing is necessary for corporations to run their businesses effectively and efficiently by minimizing any potential devastating downside risk that could harm consumers. Additionally, if regulators require end-users to post margin and collateral, then the additional cost incurred would most likely be transferred to the consumer. This Note proposes that the CFTC amend the end-user exception to apply to a broader number of participants and, in doing so, require more stringent disclosure similar to that of the ISDA Master Agreement.

208. See 75 Fed. Reg. 80747. See also Greenberger Letter, supra note 206.

209. See Letter from Americans for Fin. Reform to Elizabeth M. Murphy, Sec., SEC (Feb. 4, 2011) [hereinafter Am. for Fin. Reform Letter] (presenting a proposal for enhanced disclosure and reporting for end-users claiming the exception to the mandatory clearing of swaps) (on file with the Washington and Lee Law Review); Greenberger Letter, supra note 206.


211. See Katharine Rose, Annuity Issuers Eye Dodd–Frank Act, NAT'L UNDERWRITER/ LIFE & HEALTH FIN. SERV., Vol. 114 Issue 16, 12 (Aug. 23, 2010) (addressing concerns of insurers, such as MetLife and Harvard Financial Services Group, that an increased price of using derivatives would result in high costs to customers using financial services).
A. Proposal for End-User Disclosure

Before this Note proposes how the CFTC should model end-users’ disclosure requirements, it will briefly discuss why the other alternative approaches in Part V.B.1. do not sufficiently address the concerns posed by the OTC market. First, Lynn Stout’s theory to view OTC speculative contracts as gambling contracts does not meet the aims of Dodd–Frank. This approach would encourage investors to take responsibility for their risky behavior by making them individually responsible for any losses incurred on the contract.212 Unfortunately, this may not decrease systemic risk and promote transparency because it keeps regulators uninformed about market activity. Also, regulators have a difficult task in dissecting the difference between speculative and legitimate hedging contracts.213 Due to the market disruption that OTC derivatives allegedly caused in 2008, regulators cannot ignore the threat the market poses and must do more than just make derivatives legally unenforceable.

The second approach to have nongovernmental institutions, such as the ISDA and the Derivatives Policy Group, continue in their efforts of promoting efficiency and transparency in the OTC market could be successful. Dodd–Frank focuses on the United States’ regulation of the OTC market, and not necessarily the international market, as do these other groups.214 While these institutions have made extraordinary improvement in the OTC market, it was not sufficient to prevent the 2008 crisis.215

The third and fourth approaches introduced in Part V.B.1 combine to make the best approach to regulating OTC derivatives, specifically end-users. This Note advocates maintaining an exemption of end-users from the mandatory clearing of swaps because they are not engaging in purely

212. See Stout, Regulate OTC Derivatives, supra note 149, at 31–32.
213. See Kreitner, supra note 55, at 1135–36 (introducing the idea that the increasing popularity of OTC derivatives caused much discussion on the difficulty in distinguishing investment from gambling).
215. See supra note 5 and accompanying text (listing a number of suggested theories of how OTC derivatives contributed to the 2008 financial crisis).
speculative behavior. This mirrors the approach of regulating behavior in contrast to regulating a particular instrument. This Note proposes that the CFTC should define end-user broadly to enable more bona-fide hedging participants to claim the exemption. A broader exception would allow more users to effectively hedge against exposed risks for low transaction costs.

The more broadly the CFTC defines an end-user, however, the greater the danger that the proposed regulation will not adequately protect against systemic risk and increase transparency. This stems from the idea that the more participants who qualify as end-users, and are thus subject to minimal regulation, the less effective the legislation will be in promoting transparency. More stringent disclosure requirements will help address this issue. Some end-users wish that their trading positions remain private so others will not mimic their contracts and make their hedging strategies ineffective. This concern does not change the necessity of regulators’ awareness of OTC market activity to assure that no participants are abusing the end-user exception.

As discussed in Part V.A, the end-user exception provides that those claiming the exemption must only disclose that they intend to use the exemption and how they generally plan to meet financial obligations. Such a boilerplate disclosure statement will not adequately address Dodd–Frank’s aims to increase transparency and decrease systemic risk in the OTC market. This Note argues that disclosure and reporting requirements should resemble the terms of the ISDA Master Agreement, which is

---

216. See supra notes 203–04 and accompanying text.

217. See Greenberger Letter, supra note 206 (explaining how not enhancing disclosure and reporting requirements will prevent regulators from adequately monitoring the market for abuse of the end-user exception).

218. See id.

219. See Am. for Fin. Reform Letter, supra note 209; supra notes 207–09 and accompanying text.


221. See supra note 187 and accompanying text.
already utilized by many participants in the OTC market.\textsuperscript{222} For over 20 years, the ISDA has tackled the issue of promoting transparency and pricing efficiency in the OTC market while adapting to the constant changes in financial innovation.\textsuperscript{223} The CFTC could look to how the private sector promoted financial stability in the market and build off that model.

The ISDA Master Agreement provides for increased documentation of parties’ transactions and for close-out netting in the event of default.\textsuperscript{224} To promote economic certainty, the Master Agreement allows for the nondefaulting party to elect an early termination date and potentially receive money if the party incurs a loss while entering into a new derivatives contract.\textsuperscript{225} These provisions lower credit and legal risk for participants.\textsuperscript{226}

The CFTC could follow the standard form contract, which promotes low transaction costs, and require parties to provide additional disclosure of obligations or contract terms where they differ from the standard form.\textsuperscript{227} While the Master Agreement provides for disclosure of types of collateral thresholds, exposure calculations, payment schedules, netting, and standard contract terms, the CFTC must ensure that parties provide concrete specifics of their ability to meet financial obligations.\textsuperscript{228} Parties need to have adequate information to reduce counterparty risk, while regulators need the same data to moderate systemic risk.

The regulation focuses on requirements and needs of the OTC derivatives market as it stands today. With the rapidly changing landscape of financial markets, especially the OTC derivatives market, regulatory schemes may quickly become

\textsuperscript{222} See Partnoy, supra note 131, at 6 (providing an overview of what the ISDA Master Agreement requires the parties to disclose).

\textsuperscript{223} See ISDA Safe, Efficient Markets, supra note 214 (describing the ISDA’s aim to promote efficiency, accuracy, and stability in the OTC market).

\textsuperscript{224} See PRACTICAL DERIVATIVES, supra note 96, at 29 (discussing the concept of close-out netting, an important aspect of the ISDA Master Agreement).

\textsuperscript{225} See id.

\textsuperscript{226} See id. at 29–30.

\textsuperscript{227} See Partnoy, supra note 131, at 9 (stating that standard form derivatives documentation is cost reducing).

\textsuperscript{228} See PRACTICAL DERIVATIVES, supra note 96, at 27 (displaying the structure of the ISDA documentation at the relationship and transactional levels).
weak or ineffective. Because of the rapid growth of the OTC market and the complexity of financial innovation, regulators have difficulty in adequately addressing the concerns and risks this market poses. This Note argues that end-users should meet disclosure requirements as set out in the ISDA Master Agreement. It will allow regulators to guarantee that the market is more transparent to address the issues that they see fit.

VII. Conclusion

The regulation of OTC derivatives has long been a controversial topic, as seen in the debate between Brooksley Born and Alan Greenspan in the late 1990s. In 2000, with the passage of the CFMA, regulators believed the best route was to deregulate the OTC market. Unfortunately, this led to a vast, complex, unregulated market that many claim played a role in the financial crisis of 2008. The financial crisis demonstrates to regulators that OTC derivatives carry a threat of systemic risk that needs to be controlled.

This led Congress to pass Title VII of Dodd–Frank, which aims to address the threat of systemic risk by promoting transparency in the market. The CFTC proposed legislation that requires the mandatory clearing of swaps, which entails margin requirements that are quite costly to the parties of the transaction. Pursuant to Congressional intent, the CFTC proposed an exemption for nonfinancial entities who utilize the

---

229. See Iman Anabtawi & Steven Schwartz, Regulating Systemic Risk: Towards an Analytical Framework, 86 Notre Dame L. Rev. 1349, 1361 (2011) (presenting the idea that it is difficult to regulate financial markets because they are rapidly changing and adapting to new innovations).

230. See, e.g., id.

231. See supra notes 207–09 and accompanying text (introducing the proposal of more stringent disclosure requirements to allow a broad end-user exception).

232. See supra notes 138–44 and accompanying text.

233. See supra notes 143–44 and accompanying text.

234. See supra notes 4–5 and accompanying text (introducing theories as to how the OTC market played a role in the 2008 financial crisis).


236. See supra Part IV.B.
market to hedge underlying risk.\textsuperscript{237} This exemption is meant to maintain low transaction costs for these participants. While this is a legitimate concern for end-users, the CFTC should impose more stringent disclosure requirements than just checking the exemption box and generally disclosing how the party will meet financial obligations.\textsuperscript{238}

This proposal is to aid regulators in monitoring the end-users to determine who is validly claiming the exemption and who is abusing the exception by speculating.\textsuperscript{239} While this may increase costs to end-users to some extent, it will decrease the threat of systemic risk to the American public and promote transparency in the OTC market, the aims of Dodd–Frank.

Foreseeing and preventing a financial crisis of the magnitude of the one in 2008 is an extremely difficult task. Financial innovation and the complexity of OTC derivatives make this task far more difficult. While the government is taking measures to address the systemic threat OTC derivatives pose, these efforts may soon prove ineffective. The knowledge of derivatives traders and the rapid change in financial innovation make regulating this market a daunting task.\textsuperscript{240} But this Note’s proposal of more stringent disclosure and reporting requirements can help increase market transparency which, in turn, makes regulating the OTC market easier.

\textsuperscript{237} See supra note 116 and accompanying text (stating Congress’s intent to exempt end-users from the mandatory clearing of swaps).
\textsuperscript{238} See supra Part VI.A.
\textsuperscript{239} See Greenberger Letter, supra note 206.
\textsuperscript{240} See e.g., Anabtawi & Schwartz, supra note 229.