NEW YORK UNIVERSITY ANNUAL SURVEY OF AMERICAN LAW

VOLUME 68
ISSUE 4

NEW YORK UNIVERSITY SCHOOL OF LAW
ARTHUR T. VANDERBILT HALL
Washington Square
New York City

New York University Annual Survey of American Law
is in its seventy-first year of publication.

L.C. Cat. Card No.: 46-30523
ISSN 0066-4413

New York University Annual Survey of American Law is published quarterly at 110 West 3rd Street, New York, New York 10012. Subscription price: $30.00 per year (plus $4.00 for foreign mailing). Single issues are available at $16.00 per issue (plus $1.00 for foreign mailing). For regular subscriptions or single issues, contact the *Annual Survey* editorial office. Back issues may be ordered directly from William S. Hein & Co., Inc., by mail (1285 Main St., Buffalo, NY 14209-1987), phone (800-828-7571), fax (716-883-8100), or email (order@wshein.com). Back issues are also available in PDF format through HeinOnline (http://heinonline.org).

Manuscripts: The *Annual Survey* invites the submission of unsolicited manuscripts. Text and citations should conform to the 19th edition of *A Uniform System of Citation.* Please enclose an envelope with return postage if you would like your manuscript returned after consideration.

Editorial Office: 110 West 3rd Street, New York, N.Y. 10012
(212) 998-6540
(212) 995-4032 Fax
http: //www.law.nyu.edu/pubs/annualsurvey

For what avail the plough or sail
Or land or life, if freedom fail?
EMERSON

NEW YORK UNIVERSITY ANNUAL SURVEY OF AMERICAN LAW

2012-2013 BOARD OF EDITORS

Editor-in-Chief
THEODORE KELLY

Managing Editors
JEFFREY R. BENGEL
ANDREW OSARCHUK

Development Editors
JOSEPH M. ENO
CARLY WEINREB

Note Editors
PAUL CHAFFIN
STEPHEN ELKIND

Executive Article Editors
MAX KAPLAN
DAVID KATZ
CLIFFORD MARK LANEY
ROSE S. LEE
KEVIN EUNHO PARK
KELLY STEPHANIE ROCHESTER PARKER
EMILY WINOGRAD
DENNY WON

Senior Articles Editor
E. DANIELLE BYAM

Symposium Editor
JASON SCHEFF

Article Editors

JOSE ACOSTA
TAYLOR ALEXANDER
NICHOLAS AXELROD
AYAL BAINVOLL
MADELEINE EDDY BARENHOLTZ
ASHLEY PARKS BELTON
JEREMY H. ERSHOW
KEVIN FRICK
ANDREW J. GERSHON
GANARAJ S. HEGDE
LAUREN M. ISAACSON
ELIZABETH B. JORDAN
ERIC KAMERMAN
LORETTA LAU
WILLIAM CLAYTON LAWRENCE
JUDD LINDENFELD
MEGAN C. MANFRED
AVA MCALPIN
JAKE ORNSTEIN
MATTHEW ROTBART
JESSE B. SACKS
PAUL L. SANDLER
SOPHIE SAVRYN
P. DAVID SEIBERT
ELANA BROWNSTEIN SIEGEL
SARAH E. SIEGEL
YOAV SIMCHONI
JENNA E. SMALL
PAIGE CATHLEEN SPENCER
RIKKI STERN
CHARLES TANENBAUM
JORDAN WELLS
KARI WOHLSCHLEGEL

Staff Editors

SIOBHAN C. ATKINS
PAUL S. BALIK
GABRIEL BRUNSWICK
MARY BRUST
MICHAEL A. CANENCIA
JOSEPH M. CASTELLI
WALTER A. CIACCI
ILYSSA L. COGHLAN
JOSHUA S. COHN
GRAHAM B. COLE
MATTHEW WARREN DEAVERS
DANIELLE ELIZABETH DEBOLD
JOSHUA M. DRAPEKIN
NICOLE A. ESCOBAR
MEAGAN L. FROEMMING
JULIAN J. GINOS
DAVID A. GIROUX
MICHAEL F. GOON
ALEX J. GORMAN
JORDANA L. HAVIV
PETER HUR
GAIL L. HYMAN
NICOLE N. IDOKO
CHRISTINE A. KUVEKE
HAROLD LAIDLAW
CHARLES MICHAEL LUPICA
LEAH S. MARTIN
CHIP MCCORKLE
ERIC A. MINTZ
BENJAMIN E. NOTTERMAN
GIL OFIR
JANE PARK
JULIA C. PILCER
JEFFREY A. RITHOLTZ
ARYEH L. ROSKIES
THEODORE D. SAMETS
ALEXANDRA SAMOWITZ
NICHOLAS C. SCHOLTEN
MELISSA B. SIEGEL
AMANDEEP SINGH
CHRISTOPHER BRODERICK SMITH
NAOMI R. SOSNER
GREGORY SPRINGSTED
MAX TIERMAN
PATRICK R. TOTARO
YAEL TZIPORI
PETER VAN VALKENBURGH
DANIEL WOLFF
KYUNGEUN K. WON

SUMMARY OF CONTENTS

LAW AND MECHANISM DESIGN: PROCEDURES TO INDUCE HONEST BARGAINING
Steven J. Brams & Joshua R. Mitts 729

AN APPEALING CHOICE: AN ANALYSIS OF AND A PROPOSAL FOR CERTIFICATES OF APPEALABILITY IN "PROCEDURAL" HABEAS APPEALS
David Goodwin 791

THE LAW OF INTENDED CONSEQUENCES: DID THE AMERICANS WITH DISABILITIES ACT AMENDMENTS ACT MAKE IT EASIER FOR CANCER SURVIVORS TO PROVE DISABILITY STATUS?
Barbara Hoffman 843

FACEBOOK USED TAKEDOWN AND IT WAS SUPER EFFECTIVE! FINDING A FRAMEWORK FOR PROTECTING USER RIGHTS OF EXPRESSION ON SOCIAL NETWORKING SITES
Kevin Park 891

LAW AND MECHANISM DESIGN: PROCEDURES TO INDUCE HONEST BARGAINING

STEVEN J. BRAMS & JOSHUA R. MITTS†*

ABSTRACT

A classic challenge in contract and property law is unstructured negotiation between two parties with asymmetric information (i.e., each party has different private information) under bilateral monopoly (each party must negotiate with the other to try to reach an agreement), which often leads to prohibitively high transaction costs and, if the parties fail to agree, social costs as well. In these situations, the law should incorporate principles of mechanism design, a methodology that employs structured procedures to give the parties incentives to reach agreement. In terms of contract theory, mechanisms constitute algorithmic altering rules that reduce if not eliminate inefficient transaction costs. We review two bargaining mechanisms that inherently elicit honesty by making it a dominant strategy and discuss two extensions for legal applications. In particular, we show that algorithmic procedures would reduce transaction costs and lead to more efficient bargaining in pretrial settlement negotiations and blockholder disclosure under section 13(d) of the Securities Exchange Act of 1934. The former is a straightforward application of mechanism design to a negotiation situation wherein the social externalities of nonagreement justify inducing the honest disclosure of reservation prices, or "bottom lines." The latter is an example of using mechanism design to facilitate negotiated settlements in situations presently subject to a suboptimal mandatory rule.

Introduction ..... 730 R
I. Negotiation and the Mechanism Design Literature ... 733 R
A. Transaction Costs in Bargaining Under Bilateral Monopoly with Assymetric Information ..... 733 R
B. Negotiation Mechanism as Procedural Altering Rules ..... 739 R

* Professor of Politics, New York University.

† Associate, Sullivan & Cromwell, LLP. The views in this article should not be attributed to Mr. Mitts's employer.

C. What is Mechanism Design? A Review of the Literature ..... 743 R
II. Applying Mechanism Design to Contractual Negotiations ..... 748 R
A. Truth-Telling and Collusion in Bargaining: The Chatterjee-Samuelson and Bonus Procedures .... 748 R
B. The Two-Stage Procedure: Overview, Visualization, and Implementation ..... 753 R
1. Overview of the Mechanism ..... 753 R
2. Visualization ..... 756 R
3. Implementation ..... 757 R
C. Extensions of the Procedures for Legal Applications ..... 761 R
1. A Budget-Balanced Bonus: Taxing Under the Bonus Procedure ..... 761 R
2. Utilizing Truthful Information to Improve Regulatory Policy ..... 763 R
III. Applying Mechanism Design to Settlement Negotiations and Securities Regulation ..... 765 R
A. Settlement Negotiations ..... 766 R
1. An Overview of Mandatory Pretrial Settlement Negotiations ..... 766 R
2. Applying the Bonus and Two-Stage Procedures to Settlement Negotiations ..... 770 R
3. Social Benefits of Honesty-Inducing Procedures for Settlement Negotiations ..... 771 R
B. Securities Regulation: Reconsidering Mandatory Blockholder Disclosure ..... 773 R
1. Overview, Private Ordering, and Efficient Trade ..... 773 R
2. Applying the Bonus and Two-Stage Procedures to Negotiating Delayed Blockholder Disclosure ..... 783 R
3. Regulating Social and Macroeconomic Effects ..... 787 R
Conclusion ..... 789 R

INTRODUCTION

Freedom of contract is a fundamental principle of the American legal system. At its core, it reflects a free-market orientation that entrusts actors with the discretion to transact as they wish. We suggest, however, that a few *structured* restraints on freedom of contract could bring substantial improvements to legal fields as diverse as

alternative dispute resolution and securities regulation. This is not such a radical idea: the law already curtails unstructured negotiation for the sake of other socially beneficial goals.

A classic example from property law is the case of *Boomer v. Atlantic Cement Co.*, which involved a nuisance claim by landowners against a neighboring factory for pollution.[1] Law and economics scholars have long debated whether the right to remain free from pollution should be protected through a property rule—thereby forcing the parties to negotiate—or a liability rule—thereby permitting one party essentially to force a judicial sale of this right by bringing a lawsuit.[2]

The predominant view is that in cases like *Boomer*, a liability rule would be more efficient because this is a *bilateral monopoly*. There is no "free market" here: neither party can simply decide to trade with a different neighbor. They are "stuck with each other," in one scholar's words.[3] But the Coase theorem suggests that, in the absence of transaction costs, the parties would still negotiate to the efficient outcome. Why abrogate contractual freedom by forcing a sale? Why not mandate a bargained-for solution?

There is a major source of transaction costs in cases like *Boomer*—namely, asymmetric information, in which each party has different private information. The Coasean ideal of efficient bargaining only applies in a frictionless world where each side knows how much the other values the activity. When one side does not know the other's *reservation price*, or "bottom line" (as is typically the case), the price at which it is indifferent between transacting or not, it is in his or her best interest to engage in *strategic* negotiation. Often, this means opening with an extreme offer and yielding ground very slowly. This renders negotiations expensive and cumbersome, which is why scholars often advocate liability rules in bilateral monopoly with asymmetric information. A judicial forced sale of the entitlement—even at a price far from the parties' actual valuation—is often more efficient than costly bargaining.

We suggest that this is a false dichotomy. The law need not choose between unstructured negotiation and forced sales by the judiciary. The 2007 Nobel Prize in Economics was awarded for the development of the field of *mechanism design*, which provides a third path. Mechanism design permits structuring the rules of interaction, or protocols to be followed, to reduce or eliminate the natural

1. 257 N.E.2d 870 (N.Y. 1970).

2. *See* discussion *infra* Section I.A.

3. Carol M. Rose, *The Shadow of the Cathedral*, 106 YALE L.J. 2175, 2183 (1997).

incentive of parties to posture and exaggerate. It is an interdisciplinary approach that has been adopted in economics, political science, and even computer science. The time has come for the law to embrace this methodology as well.

Applying mechanism design to the law means instituting algorithmic procedures that impose *structural* limitations on negotiations between the parties. Unlike liability rules, these procedures do not eliminate freedom of contract because they still permit the parties to negotiate and transact at a price that reflects their subjective valuations. Yet, by replacing unstructured negotiations with bargaining procedures, mechanism design can reduce or eliminate the incentives that prevent reaching agreement in these bilateral monopoly situations. In the language of contract theory, mechanism design provides structured altering rules that render contract formation more efficient. Indeed, the *ABA Journal* recently pointed out the increasing use of procedures like these by parties wishing to facilitate transactions and settle legal disputes.[4]

In this Article, we discuss two legal contexts that are particularly suited for mechanism design: settlement negotiations and mandatory disclosure under the securities laws. In both, we argue that imposing algorithmic negotiation procedures is justified because the failure to reach agreement imposes a cost on society. Mechanism design should be mandatory when transaction costs lead to impasse and significant negative externalities in negotiations.

It is easy to see how society pays for settlement negotiations that fail because of strategic bargaining. As a case in point, excessive litigation leads to greater costs of maintaining the court system. In certain contexts (e.g., labor negotiations), third parties, including the public, may be directly harmed by the failure to reach agreement.[5]

4. *See* Deborah L. Cohen, *Not Playing Games: Firm Takes Decision-Making Theory into Transactions*, ABA JOURNAL (Jan. 1, 2013, 3:40 AM), http://www.abajournal.com/magazine/article/not_playing_games_firm_takes_decision-making_theory_into_transactions/. One of the authors (Brams) is chairman of the advisory board of Fair Outcomes, Inc., the firm discussed in the article.

5. The students harmed in Chicago's 2012 teacher strike are just one example. *See, e.g.*, Charles Lane, Editorial, *Students Are Victims in Chicago Fight Over Clout*, WASHINGTON POST, Sept. 10, 2012, http://www.washingtonpost.com/opinions/charles-lane-students-are-victims-in-chicago-fight-over-clout/2012/09/10/ec4b47c2-fb66-11e1-b2af-1f7d12fe907a_story.html; Monica Davey, *As Chicago Strike Goes On, the Mayor Digs In*, N.Y. TIMES, Sept. 17, 2012, at A1, *available at* http://www.nytimes.com/2012/09/18/education/chicago-teachers-strike-enters-second-week.html.

Securities regulation, however, is a less obvious context for mandating mechanism design. We suggest that this is an example of the power of mechanism design to facilitate bargaining when the law has traditionally imposed a suboptimal mandatory rule. The recent controversy over blockholder disclosure under Section 13(d) of the Securities Exchange Act of 1934 has generally assumed that the disclosure duration must be fixed. We show that this is a situation ripe for bargaining, but an algorithmic procedure is essential to fostering settlements.

This Article proceeds as follows. In Part I, we summarize the problem of transaction costs in bargaining under bilateral monopoly with asymmetric information, show how mechanisms may be understood as structured altering rules, and summarize the literature on mechanism design. In Part II, we consider the application of mechanism design to contract law. We discuss the shortcomings of traditional bargaining theory, review two procedures that induce honest revelation of reservation prices in bargaining, and propose two extensions of these procedures for legal applications. In Part III, we apply mechanism design to specific legal settings. We analyze its potential for improving the efficiency of bargaining in settlement negotiations and discuss specific procedures to permit negotiations in situations presently subject to a mandatory rule under the securities laws.

I.
NEGOTIATION AND THE MECHANISM DESIGN LITERATURE

A. Transaction Costs in Bargaining Under Bilateral Monopoly with Asymmetric Information

A fundamental issue in the economic analysis of law is the problem of high transaction costs resulting from bargaining under bilateral monopoly conditions with asymmetric information.[6] The initial insight was Ronald Coase's claim that in the absence of transaction costs, parties will bargain to the efficient outcome regardless of the initial allocation of rights.[7] Ever since Calabresi and Me-

6. For a general discussion of transaction costs under a bilateral monopoly with asymmetric information, see RICHARD POSNER, ECONOMIC ANALYSIS OF LAW 251 (2003); Oliver E. Williamson, *Transaction-Cost Economics: The Governance of Contractual Relations*, 22 J.L. & ECON. 233, 241–42 (1979); see also William Samuelson, *A Comment on the Coase Theorem, in* GAME-THEORETIC MODELS OF BARGAINING 321, 324–31 (Alvin E. Roth ed., 1985).

7. *See* Ronald H. Coase, *The Problem of Social Cost*, 3 J. LAW & ECON. 1 (1960). The Coase theorem has spawned a vast literature on the role of transaction costs in

lamed's well-known argument that the presence of prohibitively high transaction costs supports the imposition of liability rules rather than property rules,[8] legal scholars have sought to prescribe effective rules in the bilateral-trade context where two parties are "stuck with each other."[9] In these thin, illiquid markets, the presence of private information, which is not shared and therefore asymmetric, gives each party an incentive to misrepresent his or her bargaining offer and thereby render negotiations protracted and costly, if they succeed at all.

Scholars have proposed various ways to induce the truthful disclosure of reservation prices to reduce transaction costs in negotiations under bilateral monopoly conditions with asymmetric information. In a highly influential piece, Ian Ayres and Eric Talley propose dividing entitlements to induce uncertainty as to whether a party "will ultimately emerge as a seller or a buyer."[10] In their view, "[t]his form of rational ambivalence . . . can lead the bargainers to represent their valuations more truthfully."[11]

As a case in point, a procedure called "Fair Buy-Sell,"[12] useful in cases of joint ownership, leaves uncertain which party will be the

bargaining and trade. *E.g.*, James M. Buchanan, *Rights, Efficiency and Exchange: The Irrelevance of Transaction Costs*, *reprinted in* THE LEGACY OF RONALD COASE IN ECONOMIC ANALYSIS (S.G. Medema, ed. 1995); Robert D. Cooter, *The Cost of Coase*, 11 J. LEGAL. STUD. 1, 1–2 (1982); Robert C. Ellickson, *Of Coase and Cattle: Dispute Resolution Among Neighbors in Shasta County*, 38 STAN. L. REV. 623 (1986); Joseph Farrell, *Information and the Coase Theorem*, 1 J. ECON. PERSPECTIVES 113 (1987); Herbert Hovenkamp, *Coasean Markets*, 31 EUR. J. L. & ECON. 63 (2011); Steven G. Medema, *A Case of Mistaken Identity: George Stigler, "The Problem Of Social Cost," and the Coase Theorem*, 31 EUR. J.L. & ECON. 11 (2011) (Neth.); Donald H. Regan, *The Problem of Social Cost Revisited*, 15 J.L. & ECON. 427 (1972); *see also* ROBERT C. ELLICKSON, ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES (1994).

8. Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089 (1972).

9. IAN AYRES, OPTIONAL LAW: THE STRUCTURE OF LEGAL ENTITLEMENTS 20 (2005) ("[W]hile contracts may serve as a fine paradigmatic example, the option approach is a powerful way to analyze any bilateral monopoly situation—that is any situation where there are two (or a small number of) people who 'are stuck with each other.'") (quoting Carol M. Rose, *The Shadow of the Cathedral*, 106 YALE L.J. 2175, 2183 (1997) ("Ayres and Talley are interested in situations in which two parties are stuck with each other, thin markets instead of 'thick' ones. Neighboring landowners seem to fit that bill.")).

10. Ian Ayres & Eric Talley, *Solomonic Bargaining: Dividing a Legal Entitlement to Facilitate Coasean Trade*, 104 YALE L.J. 1027, 1030 (1995).

11. *Id.*

12. For an informal description of fair buy-sell, see *Fair Buy-Sell*, FAIR OUTCOMES, INC., http://www.fairoutcomes.com/fb.html (last visited Jan. 22, 2014). Fair Outcomes, Inc. is a firm utilizing the patented fair buy-sell procedure to provide online dispute resolution and negotiation consulting services. For a formal

buyer and which the seller. Each party is thereby motivated to offer the other party a price that makes it indifferent between being the buyer and being the seller. Because the price is set at the mean of the two offers—with the party that offers more becoming the buyer and the party that offers less becoming the seller—each party does better than its offer and so profits from the transaction. Stergios Athanassogloua, Steven J. Brams, and Jay Sethuraman have shown that if the parties are equally endowed, each has an incentive to offer its truthful reservation price.[13] Another approach to truthful revelation, as Ayres and Talley suggest, is to divide an entitlement by protecting the parties by liability rules, which has an "information-forcing quality . . . [that] can induce entitlement holders to signal credible information about their valuation."[14]

The bulk of the law and economics literature responding to Ayres and Talley has focused not on the question of inducing honest disclosure in bargaining, but rather on whether property rules or liability rules are more efficient in the asymmetric information setting. For example, Louis Kaplow and Steven Shavell argue that while total welfare is higher under a liability rule, welfare gains from bargaining are lower under a liability rule than a property rule.[15] Ayres and Talley reply that litigation costs may still render liability rules more efficient.[16] In a separate piece, Kaplow and Shavell argue that the choice between property and liability rules is indeterminate: "examples can be constructed in which either the liability rule is superior to property rules or the reverse is true."[17] Yet they conclude that a liability rule "tends to be superior" because "before any bargaining occurs . . . the liability rule is ahead of the property rules. . . . [A]fter imperfect bargaining occurs, the liability rule will remain ahead of the property rules, although not as far ahead."[18] These arguments led to a series of articles by other schol-

academic analysis of fair buy-sell, see Stergios Athanassoglou, Steven J. Brams & Jay Sethuraman, *A Note on the Inefficiency of Bidding over the Price of a Share*, 60 MATHEMATICAL SOC. SCI., 191 (2010). For a related mechanism applicable to dividing assets among players, see Peter Cramton, Robert Gibbons & Paul Klemperer, *Dissolving a Partnership Efficiently*, 55 ECONOMETRICA 615 (1987).

13. Athanassoglou, Brams & Sethuraman, *id.* at 191.

14. Ayres & Talley, *supra* note 10, at 1100.

15. Louis Kaplow & Steven Shavell, *Do Liability Rules Facilitate Bargaining? A Reply to Ayres and Talley*, 105 YALE L.J. 221, 227–29 (1995).

16. Ian Ayres & Eric Talley, *Distinguishing Between Consensual and Nonconsensual Advantages of Liability Rules*, 105 YALE L.J. 235, 249 (1995).

17. Louis Kaplow & Steven Shavell, *Property Rules Versus Liability Rules: An Economic Analysis*, 109 HARV. L. REV. 713, 735 (1996).

18. *Id.*

ars considering the relative superiority of property and liability rules.[19]

In our view, however, a crucial aspect of the discussion has gone unnoticed. The scholarship thus far has considered the choice between property and liability rules against the backdrop of unstructured negotiation. There seems to be a shared assumption that the legal system faces a choice between two fundamental approaches: either compel the parties to engage in unstructured, free-for-all negotiation (property rule), or permit one party to force an involuntary "cashing out" of an entitlement on the other via the judicial process (liability rule). But there is a third option: compel the parties to negotiate in order to transfer the entitlement but restrict the *procedural* rules governing the negotiation. The right process just might eliminate the incentive to engage in costly strategic bargaining, reduce transaction costs, and thereby facilitate efficient trade.

Legal scholarship has yet to consider the question of whether procedural improvements can eliminate much of the transaction costs resulting from unstructured negotiation under asymmetric information. Interestingly, a substantial portion of the law deals with rules of procedure for judicial proceedings.[20] Many articles have suggested ways to improve speed and efficiency in the courtroom.[21] Some authors have even considered the implications of different procedural rules in alternative dispute resolution.[22] But the law and economics literature has yet to consider whether a type of contracting procedure could reduce the transaction costs lying at the heart of the Calabresi and Melamed framework and much of mod-

19. *E.g.*, Richard A. Epstein, *A Clear View of The Cathedral: The Dominance of Property Rules*, 106 YALE L.J. 2091, 2104–05 (1997); Carol M. Rose, *The Shadow of the Cathedral*, 106 YALE L.J. 2175 (1997);

20. These include the federal rules of civil and criminal procedure and their corresponding state analogues.

21. *E.g.*, Patrick E. Longan, *The Shot Clock Comes to Trial: Time Limits for Federal Civil Trials*, 35 ARIZ. L. REV. 663 (1993); John Burritt McArthur, *The Strange Case of American Civil Procedure and the Missing Uniform Discovery Time Limits*, 24 HOFSTRA L. REV. 865 (1996); Carrie E. Johnson, Comment, *Rocket Dockets: Reducing Delay in Federal Civil Litigation*, 85 CAL. L. REV. 225 (1997).

22. *E.g.*, Benjamin Aaron, *Some Procedural Problems in Arbitration*, 10 VAND. L. REV. 733 (1957); Steven J. Brams & Samuel Merrill, III, *Binding Versus Final-Offer Arbitration: A Combination Is Best*, 32 MGMT. SCI. 1346 (1986); Dao-Zhi Zeng, Shinya Nakamura & Toshihide Ibaraki, *Double-Offer Arbitration*, 31 MATHEMATICAL SOC. SCI. 147 (1996); Ellen E. Deason, *Procedural Rules for Complementary Systems of Litigation and Mediation—Worldwide*, 80 NOTRE DAME L. REV. 553, 562–565 (2005); Katherine Stone, *Procedural Justice in the Boundaryless Workplace: The Tension Between Due Process and Public Policy*, 80 NOTRE DAME L. REV. 501 (2005).

ern law and economics scholarship. As we explain *infra*, Ian Ayres's recent article on altering rules comes the closest—indeed, in our view, bargaining procedures constitute a type of altering rule—but the question of which procedures are most appropriate in different situations has been largely unaddressed.

But there is a voluminous body of literature, both theoretical and empirical, that is specifically dedicated to evaluating which types of procedures efficiently facilitate optimal agreement in bilateral trade bargaining under asymmetric information.[23] The law and economics literature has considered similar ideas when evaluating the "incomplete contracting" field. Incomplete contracting scholarship develops formal models for designing effective ex ante procedures within contractual terms to ensure efficient performance.[24] Price terms, for example, might be intentionally omitted from the contract, replaced by a procedure for determining the price in real time.

This slightly differs from our notion of a bargaining mechanism. We ask whether regulators should institute mandatory procedures for the sake of reducing transaction costs and improving social welfare, not whether parties would find it optimal to agree to use a procedure on their own initiative. Interestingly, the incomplete contracting literature has been criticized for failing to reflect actual contracting and for assuming that people have unrealistic cognitive abilities.[25] Of course, current practice does not resolve

23. *See infra* Section II.A.

24. *See, e.g.*, PATRICK BOLTON & MATHIAS DEWATRIPONT, CONTRACT THEORY (2005); OLIVER D. HART, FIRMS, CONTRACTS, AND FINANCIAL STRUCTURES 34 (1995); Oliver Hart & John Moore, *Incomplete Contracts and Renegotiation*, 56 ECONOMETRICA 755 (1988); Ilya Segal, *Complexity and Renegotiation: A Foundation for Incomplete Contracts*, 66 REV. ECON. STUD. 57, 72–73 (1999).

25. *See* Eric A. Posner, *Economic Analysis of Contract Law After Three Decades: Success or Failure?*, 112 YALE L.J. 829, 859 (2003) (citing Karen Eggleston, Eric A. Posner & Richard Zeckhauser, *The Design and Interpretation of Contracts: Why Complexity Matters*, 95 NW. U. L. REV. 91, 122–25 (2000)). *See also* George J. Mailath, *Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory*, 36 J. ECON. LIT. 1347 (1998) ("The contracts that the models predict do not exist in the world. Instead, we see simple fixed price contracts or contracts that are conditional on a relatively small number of real world contingencies. Intuitively, the problem with the predicted contracts is that they are too complex for parties to design. To write such contracts, parties would need to imagine their bargaining position if a breach should occur, and then work their way via backward induction to the optimal terms of the contract. People are not very good at backward induction."). *But see* George S. Geis, *Automating Contract Law*, 83 N.Y.U. L. REV. 450, 489 (2008) ("I suspect that [empirical analysis of historical contracts], while complicated, would reveal some areas where parties have restructured their contracts in procedural terms."). However, Geis subsequently observes: "Yet even a quick glance through

the normative question of whether bargaining procedures would be more efficient than fixed terms.[26] More fundamentally, even if cognitive limitations are implied by the empirical underutilization of bargaining procedures, this merely suggests that the law should not force contracting parties to invent such procedures in an ad hoc manner, but instead should offer a repertoire of procedures from which the parties can choose the one best suited to resolving the dispute they face.

We show later that regulators can establish simple, easy-to-follow bargaining procedures for situations in which unstructured negotiation would likely lead to inefficiently high transaction costs. Since neither set of procedures involves long chains of reasoning, concerns with citizens' ability to engage in backward induction are misplaced. Indeed, the reason why parties will be truthful about their reservation prices is intuitive, even if a mathematical argument is needed to prove this rigorously. By contrast, the opportunity cost of informal bargaining may be arduous and prolonged negotiation; if negotiation fails, a dispute may end up in costly litigation. Regulators would seem well-suited to institute mandatory procedures, possibly embodied in a website, that enjoy a greater level of public trust than those operated by private entities.[27]

The mandatory nature of bargaining procedures is justified under the two traditional rationales for mandatory rules in contract law: paternalism and externalities.[28] On one level, these procedures

CORI or other contract databases suggests that many agreements do indeed lack the sort of procedural bargaining mechanisms prescribed in the incomplete contracting literature. They are simple fixed price deals, or they focus more on substantive contingencies." *Id.* at 489. Geis has advanced ideas similar to ours by suggesting the automated analysis of historical contracts, *see id.*, and proposing that parties incorporate procedural mechanisms into contracts to facilitate optimal substantive outcomes. *See* George S. Geis, *Internal Poison Pills*, 84 N.Y.U. L. REV. 1169, 1209 (2009). Unlike Geis, we argue for the regulatory imposition of mechanisms to advance social goals. *See* discussion *infra* Section I.C.

26. *See* Ian Ayres, *Valuing Modern Contract Scholarship*, 112 YALE L.J. 881, 881 (2003) ("[T]he thought that efficiency analysis would provide a mechanism to predict the details of current doctrine is a serious misreading of the aims of modem scholarship."); *see also* Richard Craswell, *In That Case, What Is the Question? Economics and the Demands of Contract Theory*, 112 YALE L.J. 903 (2003) (supplying a general critique of Posner's argument).

27. That said, as noted *supra* note 12, at least one commercial website, FAIR OUTCOMES, INC., does not charge for the use of one of its patented procedures and escrow services ("Fair Buy-Sell"), which we alluded to earlier as an example of a procedure in which a party does not know whether it will be the buyer or the seller.

28. *See* Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 YALE L.J. 87, 88–89 (1989) ("[I]mmutable rules are

protect "parties within the contract"[29] by mitigating transaction costs: parties cannot be trusted with unstructured negotiation because, in the presence of asymmetric information, they have a natural incentive to distort their offers, which may prevent their reaching a mutually satisfactory agreement.[30] A bargaining procedure thus promotes efficiency by restricting the method by which negotiation can be conducted in order to advance what the parties would have wanted in the absence of transaction costs—a bargain that maximizes their respective utilities. On another level, bargaining procedures protect "parties outside the contract"[31] by constraining negotiation to ameliorate the social externalities of nonagreement and providing additional regulatory benefits, such as obtaining accurate information about reservation prices. Bargaining mechanisms are thus justified as mandatory constraints on parties' contractual freedom when they not only benefit the parties themselves but also minimize social costs, e.g., the costs imposed on the public by a strike if the parties cannot reach an agreement.

Despite their mandatory nature, however, bargaining procedures differ fundamentally from traditional mandatory rules, because they do not replace the negotiation of efficient substantive terms between the parties. Indeed, a bargaining rule merely prescribes the process by which an agreement is to be reached. Accordingly, it is best understood as an altering rule—albeit a mandatory altering rule—as we explain in the next Section.

B. Negotiation Mechanisms as Procedural Altering Rules

In this Section, we suggest that the bargaining procedures—*negotiation mechanisms* in our terminology[32]—can be understood under traditional contract theory as structured altering rules. As Ian Ayres explains in his recent article, altering rules are "the necessary and sufficient conditions for displacing a default legal treatment with some particular other legal treatment."[33] A classic example of altering rules in contract law is the U.C.C. requirement that a disclaimer of the warranty of merchantability must mention

justifiable if society wants to protect (1) parties within the contract, or (2) parties outside the contract. The former justification turns on paternalism; the latter on externalities.") (footnote omitted).

29. *Id.* at 88.

30. *See* Davey, *supra* note 5.

31. Ayres & Gertner, *supra* note 28, at 88.

32. *See* discussion *infra* Section I.C for a discussion of the field of mechanism design.

33. Ian Ayres, *Regulating Opt-Out: An Economic Theory of Altering Rules,* 121 YALE L.J. 2032, 2036 (2012).

"merchantability" to be effective.[34] Altering rules thus specify the procedural conditions under which agreement regarding contractual terms will be effective. Viewed in this light, bargaining procedures are simply a highly structured type of altering rule. The default substantive allocation of rights may be displaced by agreement between the contracting parties if and only if they comply with the specified procedure.

One might quibble with our characterization of bargaining procedures as altering rules because we are advocating the application of mechanisms to contract formation and not simply the displacement of defaults for specific terms once agreement has been reached. Yet we see little substance in this distinction. Terms of a contract that are essential to agreement are no less subject to default rules than nonessential terms. By making agreement on such terms necessary for contract formation, the default rule is simply a condition that altering the term is a necessary prerequisite to a binding agreement. Our proposal follows the line of reasoning implied in Ayres's suggestion to mandate disclosure of information concerning markups and comparable sales if contractors wish to displace the default rule that a price must be reasonable.[35] Presumably, any attempt to agree on a nonreasonable price without such disclosure would be ineffective, leading the price term to revert to the reasonable price default. We simply propose taking this one step further by conditioning the very formation of a contract upon compliance with an altering rule that formalizes the bargaining procedure. Doctrinally, this would mean replacing the reasonable-price default rule with no substantive default and a condition that agreement on price must be reached, as is currently the case with the quantity term.[36] The altering rule for these terms would then consist of the execution of a bargaining procedure.

Another analogy in Ayres's article further illuminates the role of bargaining procedures as effective altering rules: the use of software confirmations to ensure that users give sufficient thought to their actions.[37] Ayres discusses a "two-click altering rule" of clicking on an attachment and clicking on a button in a confirmation window in Microsoft Outlook to displace the default rule that attachments do not open upon opening an e-mail message.[38] He

34. U.C.C. § 2-316 (1977).

35. Ayres, *Regulating Opt-Out*, *supra* note 33 at 2107.

36. For more discussion on quantity versus price defaults, see Ayres & Gertner, *supra* note 28, at 95–97.

37. Ayres, *Regulating Opt-Out*, *supra* note 33, at 2040–41, 2063–66, 2068–69.

38. *Id.* at 2040.

rightly points out that this altering rule is itself a default, which can be "altered" by checking a box in the confirmation window.[39]

We propose to focus on the *procedural* content of this altering rule for opening e-mail attachments. Clicking on the attachment is just one possible process for altering the nonopening default. While the check box second-order altering rule reduces the first-order altering rule from two clicks to one, this still may not be the most efficient method of opening attachments. One could think of numerous alternative mechanisms, from opening documents "in place," within the e-mail message, to automatically downloading attachments to a folder on the computer. Our point is that the mechanism itself matters.

In the context of bargaining under bilateral monopoly with asymmetric information, bargaining theory has shown that certain procedures can minimize transaction costs and reduce social externalities by giving parties natural incentives to honestly disclose reservation prices. We suggest that policymakers should look to these mechanisms as a regulatory means of prescribing more efficient and socially beneficial altering rules.

Indeed, Ayres acknowledges that altering rules can reduce transaction costs and promote external social goals such as enhancing competition.[40] As we mentioned earlier, his proposal for an altering rule that conditions displacement of a reasonable-price default on certain types of disclosure is a good example.[41] Mechanisms are closest to this type of altering rule because they impede contractual freedom to reduce the transaction costs inherent in unstructured negotiations as well as promote socially beneficial goals, such as truthful disclosure of reservation prices.[42]

Unlike traditional mandatory rules, mechanisms do not suffer from the inefficiency of pooling all contractors at identical, prede-

39. *Id.*

40. *See Id.* at 2103.

41. *See Id.* at 2107.

42. In *Regulating Opt-Out*, Professor Ayres mentions a prior proposal he developed with Barry Nalebuff to impose a mechanism-like altering rule for credit card issuers wishing to unilaterally raise a cardholder's interest rates. The card issuer must first "put the existing account balance up for auction on a LendingTree-like service that would allow other credit card issuers to bid for a chance to issue a new card and take over the existing balance." *Id.* at 2108 (quoting Ian Ayres & Barry Nalebuff, *A Market Test for Credit Cards*, FORBES, July 13, 2009, at 96, *available at* http://www.forbes.com/forbes/2009/0713/opinions-market-credit-cards-why-not.html). Such an altering rule is similar to the type of mechanisms we envision, but our approach alters incentives in a more fundamental way: it induces the parties to reveal their reservation prices, or bottom lines, which are private information.

termined terms. One of the chief advantages of altering rules is that they can induce efficient separating equilibria.[43] Mechanisms permit parties to contract for optimal outcomes more efficiently, maximizing their utility by setting their own contracting terms. They operate by *structuring* incentives within the negotiation process, not eliminating them. Mechanisms are rules of bargaining, and as such they still permit contractors to realize the benefit of their bargain (e.g., the economic value obtained by agreement, which they might not reach on their own).

Our mechanisms offer an additional advantage beyond preserving individual efficiency: they provide an effective platform to regulate microeconomic transactions that may have detrimental macroeconomic outcomes. In the financial regulatory context, Ian Ayres and Joshua Mitts recently pointed out the potential for increased systemic risk with excessive clustering of home mortgages at low levels of equity.[44] This reflects a more general problem where individually rational microeconomic decisions contribute to macroeconomic risk by leading to excessive pooling equilibria.[45]

In these situations, regulation can reduce systemic risk by inducing contractors to choose beneficial separating equilibria. Mechanisms are particularly suited for this task, because the parties' freedom of contract is already constrained by the bargaining procedure. If, for example, as Ayres and Mitts propose, the law should impose a system of leverage licensing to enable regulators to directly control the distribution of home equity,[46] such licenses could be implemented more easily and cheaply if actors were already utilizing a structured mechanism to reach agreement. The mechanism would serve as a natural enforcement device by simply preventing agreement at unlicensed terms.

As we explain fully in Section III.B *infra*, the power of mechanistic altering rules is particularly evident when applied to the recent controversy over blockholder disclosure. While any altering rule permitting private ordering would be more efficient than the current ten-day mandatory rule, a mechanism could also induce the honest disclosure of reservation prices. This would bring independent social benefits, such as assisting the U.S. Securities and Exchange Commission ("SEC") in improving the mechanism (e.g., by

43. *See* Ayres, *Regulating Opt-Out*, *supra* note 33, at 2091.

44. *See* Ian Ayres & Joshua Mitts, *Three Proposals for Regulating the Distribution of Home Equity*, YALE J. ON REG. (forthcoming Feb. 2014) (manuscript at 13–20), *available at* http://ssrn.com/abstract=2161545.

45. *Id.* at 48.

46. *See id.* at 33–39 .

increasing penalties for opportunistic bargaining conducted in bad faith) and other aspects of securities regulatory policy. Finally, a mechanism could induce hedge funds to separate at different equilibria to prevent the negative social externality of pooling at an identical lengthy waiting period.

C. What is Mechanism Design? A Review of the Literature

In this Section, we introduce the field of mechanism design, which provides a link between the conceptual notion of a bargaining procedure and specific theoretical and empirical research on structured procedures that give actors incentives to reach certain desired outcomes. Mechanism design is a vast, sophisticated field encompassing economics, political science, and computer science.[47] It is "the art of designing the rules of the game (a.k.a. mechanism) so that a desirable outcome (according to a given objective) is reached despite the fact that each agent acts in his own self- interest."[48] In 2007, the Royal Swedish Academy of Sciences awarded the Nobel Prize in Economics to three scholars for "having laid the foundations of mechanism design theory," which addresses "the optimal mechanism to reach a certain goal, such as social welfare or private profit" in "transactions [that] do not take place in open markets but within firms, in bargaining between individuals or interest groups and under a host of other institutional arrangements."[49] Mechanism design, in the view of the Academy, "has greatly enhanced our understanding of the properties of optimal allocation mechanisms in such situations, accounting for individuals' incentives and private information."[50] Five years later, in 2012, the Nobel Prize in Economics was awarded to two scholars for

47. For an overview of mechanism design, see Eric S. Maskin, *Mechanism Design: How to Implement Social Goals*, 98 AM. ECON. REV. 567 (2008); Roger B. Myerson, *Perspectives on Mechanism Design in Economic Theory*, 98 AM. ECON. REV. 586, 586–87 (2008). Maskin, Myerson, and Leonid Hurwicz won the 2007 Nobel Prize in Economics. Recent books on mechanism design that reflect different approaches to the subject include Y. NARAHARI, DINESH GART, RAMASURI NARAYANAM & HASTAGIN PRAKASH, GAME THEORETIC PROBLEMS IN NETWORK ECONOMICS AND MECHANISM DESIGN SOLUTIONS (2009) and RAKESH V. VOOHRA, MECHANISM DESIGN: A LINEAR PROGRAMMING APPROACH (2011).

48. Tuomas Sandholm, *Automated Mechanism Design: A New Application Area for Search Algorithms*, *in* PRINCIPLES AND PRACTICE OF CONSTRAINT PROGRAMMING – CP 2003 19, 19(Francesca Rossi, ed. 2003).

49. Press Release, Royal Swedish Academy of Sciences, The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2007 (Oct. 15, 2007), *available at* http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2007/press.html.

50. *Id.*

mechanism design of a different sort—finding "stable matching[s]" of "new doctors with hospitals, students with schools, and organ donors with patients."[51]

Mechanism design arose out of the fundamental insight by Leonid Hurwicz that efficient outcomes could be obtained if procedures are designed to be *incentive-compatible*, giving parties an incentive to truthfully report their private information.[52] Roger Myerson subsequently articulated the "revelation principle" in its most general form, which shows that "[g]iven any feasible auction mechanism, there exists an equivalent feasible direct revelation mechanism which gives to the seller and all bidders the same expected utilities as in the given mechanism."[53] This insight led to substantial innovations in bargaining theory[54] and auction theory.[55] For example, scholars have used principles of mechanism design to propose many specialized forms of auctions, including combinatorial,[56] flexible double,[57] and simultaneous ascending auctions.[58]

51. Press Release, Royal Swedish Academy of Sciences, The Prize in Economic Sciences 2012 (Oct. 15, 2012), *available at* http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2012/press_02.pdf.

52. *See* Leonid Hurwicz, *On Informationally Decentralized Systems*, *reprinted in* STUDIES IN RESOURCE ALLOCATION PROCESSES 425, 430 (Kenneth J. Arrow & Leonid Hurwicz, eds. 1977) ("It is therefore natural to seek a feature of the mechanism that would in some sense equalize the chances of the participants, regardless of their initial endowments."). For an overview of the history of mechanism design, see generally PRIZE COMM. OF THE ROYAL SWEDISH ACAD. OF SCIS., SCIENTIFIC BACKGROUND ON THE SVERIGES RIKSBANK PRIZE IN ECONOMIC SCIENCES IN MEMORY OF ALFRED NOBEL 2007: MECHANISM DESIGN THEORY (2007), *available at* http://www.nobelprize.org/nobel_prizes/economics/laureates/2007/advanced-economicsciences2007.pdf.

53. Roger Myerson, *Optimal Auction Design*, 6 MATHEMATICS OF OPERATIONS RESEARCH 58, 62 (1981). These conclusions built on Myerson's earlier work. *See* Roger Myerson, *Incentive Compatibility and the Bargaining Problem*, 47 ECONOMETRICA 61 (1979).

54. *See* discussion *infra* Section II.A.

55. *See generally* Paul Milgrom, PUTTING AUCTION THEORY TO WORK 35, 37–38 (2004); Paul Klemperer, *Auction Theory: A Guide to the Literature*, 13 J. ECON. SURVEYS 227 (1999).

56. *See, e.g.*, Peter Cramton, Yoav Shoham & Richard Steinberg, *Introduction to Combinatorial Auctions*, *in* COMBINATORIAL AUCTIONS (Peter Cramton, Yoav Shoham, & Richard Steinberg, eds., 2006).

57. *See, e.g.*, Peter R. Wurman, William E. Walsh & Michael P. Wellman, *Flexible Double Auctions for Electronic Commerce: Theory and Implementation*, 24 DECISION SUPPORT SYSTEMS 17, 18 (1998).

58. *See, e.g.*, Paul Milgrom, *Putting Auction Theory To Work: The Simultaneous Ascending Auction*, 108 J. POL. ECON. 245, 246 (2000).

Mechanisms have been used in a wide range of political applications, including school choice and student assignment,[59] voting,[60] and the design of democratic political institutions.[61] Mechanism design has recently even taken hold in the computer science literature, where it is known as "algorithmic mechanism design."[62] This research seeks to design algorithms "where the participants cannot be assumed to follow the algorithm but rather their own self-interest."[63] Examples of applications of algorithmic mechanism design include preserving privacy,[64] real-time scheduling,[65] and even pricing wireless Internet access at Starbucks.[66] Algorithmic mechanism design is essentially a mirror image of our proposal: because computer-science applications cannot force actors to comply with the mechanism, algorithmic mechanism design focuses on designing procedures which achieve goals based on fulfilling actors' self-interest. On the other hand, because the law can compel compliance, we argue for the mandatory imposition of mechanisms—but only when execution of the procedure would be in parties' and society's interests (i.e., when transaction costs are high and nonagreement imposes negative externalities).

While most of the mechanism design literature is theoretical in nature, there are an increasing number of empirical studies as well.

59. *E.g.*, Atila Abdulkadiroglu & Tayfun Sönmez, *School Choice: A Mechanism Design Approach*, 93 AM. ECON. REV. 729, note 3 (2003); Parag A. Pathak, *The Mechanism Design Approach to Student Assignment*, 3 ANN. REV. ECON. 513 (2011).

60. *E.g.*, STEVEN J. BRAMS, MATHEMATICS AND DEMOCRACY: DESIGNING BETTER VOTING AND FAIR-DIVISION PROCEDURES (2008); Jens Großer, *Voting Mechanism Design: Modeling Institutions in Experiments*, *in* EXPERIMENTAL POLITICAL SCIENCE: PRINCIPLES AND PRACTICES 72 (Bernhard Kittel, Wolfgang J. Luhan & Rebecca B. Morton, eds., 2012).

61. *See* Emmanuelle Auriol & Robert J. Gary-Bobo, *On Robust Constitution Design*, 62 THEORY & DECISION 241 (2007).

62. *E.g.*, Noam Nisan & Amir Ronen, *Algorithmic Mechanism Design*, 35 GAMES & ECON. BEHAV. 166 (2001). Nisan and Ronen's paper has been highly influential, with over 1,274 citations as of Jan. 22, 2014 according to Google Scholar. For a general overview of algorithmic mechanism design, see NOAM NISAN ET AL., ALGORITHMIC GAME THEORY (2007).

63. Nisan & Ronen, *id.* at 166.

64. *See generally* Frank McSherry & Kunal Talwar, *Mechanism Design via Differential Privacy*, *in* 48TH ANNUAL IEEE SYMPOSIUM ON FOUNDATIONS OF COMPUTER SCIENCE 94 (2007).

65. *See generally* Ryan Porter, *Mechanism Design for Online Real-Time Scheduling*, *in* EC '04 PROCEEDINGS OF THE 5TH ACM CONFERENCE ON ELECTRONIC COMMERCE 61 (2004).

66. *See* Eric J. Friedman & David C. Parkes, *Pricing WiFi at Starbucks: Issues in Online Mechanism Design*, *in* EC '03 PROCEEDINGS OF THE 4TH ACM CONFERENCE ON ELECTRONIC COMMERCE 240 (2003).

For example, an experiment on house allocation mechanisms with seventy-eight subjects found that the theoretical advantages of the top-trading cycles mechanism over the random serial dictatorship with a squatting- rights mechanism held true in practice.[67] In the auction context, an experimental test of sealed-bid auctions with ambiguity showed that ambiguity leads to lower prices in first-price auctions, but first-price auctions enjoy higher revenue than second-price auctions regardless of whether ambiguity due to incomplete information is present.[68] Mechanisms are particularly suited to experimental testing because of their procedural nature, and experimental results can inform the design of effective mechanisms in practice.[69]

Surprisingly, relatively little legal scholarship has addressed mechanism design, despite its great potential to inform policymaking. A handful of articles have utilized mechanism design in a descriptive sense, i.e., to model existing legal rules. In particular, Eric Talley utilizes the theoretical advantages of mechanism design—namely, that the theory is applicable regardless of the precise bargaining game utilized by the parties—to demonstrate the efficiency of the liquidated damages penalty doctrine.[70] He shows that the penalty doctrine "reduces both parties' incentives and abilities to engage in deceptive behavior during renegotiation, and it thereby mitigates the inefficiencies that usually accompany bilateral monopoly."[71]

Despite the apparent similarities, Talley's approach differs fundamentally from ours. Talley employs mechanism design theory—particularly the revelation principle[72]—to model an existing legal rule and argue that his model demonstrates that the existing legal

67. *See generally* Yan Chen & Tayfun Sonmez, *An Experimental Study of House Allocation Mechanisms*, 83 ECON. LETTERS 137 (2004). The theoretical advantages of the top trading cycles mechanism were proven in Atila Abdulkadiroglu & Tayfun Sönmez, *House Allocation with Existing Tenants*, 88 J. ECON. THEORY 233, 249 (1999).

68. *See* Yan Chen, Peter Katušcák & Emre Ozdenoren, *Sealed Bid Auctions with Ambiguity: Theory and Experiments*, 136 J. ECON. THEORY 513, 514 (2007).

69. *See* Großer, *supra* note 60, at 72, 73–74 ("Game theory provides an effective toolbox for describing specific institutions and procedures, and revealing the players' strategic incentives in these mechanisms. Its mathematical structure makes it relatively easy to test the games' assumptions and predictions in experiments. In turn, laboratory results often inspire game theoretic modeling when unexpected behavior is observed, yielding more realistic assumptions about the players' behavior . . . or their motives . . . .").

70. Eric L. Talley, *Contract Renegotiation, Mechanism Design, and the Liquidated Damages Rule*, 46 STAN. L. REV. 1195, 1220–25 (1994).

71. *Id.* at 1198.

72. *See id.*at 1222–23.

rule can reduce transaction costs.[73] Our claim, however, is normative: we are not justifying existing doctrine but suggesting the imposition of altering rules based on principles of mechanism design—namely, the use of structured procedures—to reduce transaction costs and social externalities. This distinction also sets our proposal apart from other legal literature that utilizes mechanism design to model existing legal rules.[74]

With one exception, we could not find any literature using game-theoretic principles of mechanism design in a normative sense, i.e., to suggest designing legal rules consisting of mechanistic procedures.[75] In a 2009 article, George S. Geis utilized mechanism design to propose an "internal poison pill," consisting of embedded options that place minority shareholders under a "veil of ignorance," in order to "elicit honest valuations."[76] This proposal, however, does not argue for a new legal rule but rather for a "novel security," which would constitute a "tool for finessing the dual hazards of majority expropriation and minority holdout."[77] Nonetheless, Geis recognizes the appeal of a mechanism design approach in protecting entitlements:

73. *See id.* at 1225–49.

74. *See, e.g.*, Farrell, *supra* note 7, at 117–21 (describing the theoretical role of mechanism design in contractual settings); Jason Scott Johnson, *Default Rules/Mandatory Principles: A Game Theoretic Analysis of Good Faith and the Contract Modification Problem*, 3 S. CAL. INTERDISC. L.J. 335 (1993). Practice preceded theory in the case of final-offer arbitration—whereby an arbitrator cannot split the difference between the two final offers of each side but must choose one offer or the other—which is used to reach settlements in salary disputes in major league baseball and in some disputes between government agencies and their public-employee unions. While designed to induce the two sides to make offers that converge on a settlement, this is not the case in theory as well as practice. See Steven J. Brams & Samuel Merrill, III, *Equilibrium Strategies for Final-Offer Arbitration: There Is No Median Convergence*, 28 MANAGEMENT SCIENCE 927 (1983), and, at a less technical level, Steven J. Brams, D. Marc Kilgour & Samuel Merrill III, *Arbitration Procedures*, *in* NEGOTIATION ANALYSIS 47 (H. Peyton Young ed., 1991). Another example in which an empirical procedure (to match medical interns and residents with hospitals) was used before its theoretical properties were well understood is described in ALVIN E. ROTH & MARILDA A. OLIVEIRA SOTOMAYOR, TWO-SIDED MATCHING: A STUDY IN GAME-THEORETIC MODELING AND ANALYSIS 4–5 (1992).

75. However, the arbitration and two-sided matching procedures described in the preceding footnote, though tried out empirically before they were analyzed theoretically, are certainly examples of procedures grounded in game theory. So are two arbitration procedures (cited earlier) that induce the two sides to reveal their reservation prices. *See* BRAMS & MERRILL, *supra* note 22; ZENG ET AL., *supra* note 22.

76. George S. Geis, *Internal Poison Pills*, 84 N.Y.U. L. REV. 1169, 1209 (2009).

77. *Id.* at 1221.

> One of the more exciting developments in economic theory posits that incentive-molding rules can corral parties toward optimal social ends strictly by appealing to their rational self-interest. If these ideas can be put into practice, it may become possible for policymakers to craft intermediate legal entitlements—somewhere in between the property and liability rules of Calabresi and Melamed—that promote welfare-enhancing substantive outcomes at a streamlined administrative cost.[78]

We echo this observation, for it lies at the heart of our proposal to institute mechanistic altering rules. Yet relatively little legal scholarship has engaged with the mechanism design literature in a normative, rule-setting sense. In the following Part, we present a theoretical proposal for incorporating mechanism design in the contractual setting.

II.
APPLYING MECHANISM DESIGN TO CONTRACTUAL NEGOTIATIONS

This Part examines theoretical aspects of applying mechanism design to the law. It begins by discussing the problem of inducing honesty in bargaining in light of the Chatterjee-Samuelson procedure and the problem of collusion in light of the Bonus Procedure, which does induce honesty, absent collusion. Section II.B discusses the Two-Stage Procedure, which is robust against collusion but slightly less efficient than the Chatterjee-Samuelson procedure. Finally, Section II.C considers issues related to the implementation of mechanism design in legal applications.

A. Truth-Telling and Collusion in Bargaining: The Chatterjee-Samuelson and Bonus Procedures

As Ayres and Talley show, a major source of transaction costs in bargaining is the incentive to misrepresent offer prices.[79] This so-called honesty-in-bargaining problem has been considered extensively in the bargaining literature. Beginning with the seminal works by John von Neumann and Oskar Morgenstern[80] and John F. Nash,[81] economic and game-theoretic scholars have considered the circumstances under which bargaining will lead to different out-

78. *Id.*

79. Ayres & Talley, *supra* note 10, at 1030.

80. JOHN VON NEUMANN AND OSKAR MORGENSTERN, THEORY OF GAMES AND ECONOMIC BEHAVIOR 15–31 (2d ed. 1944).

81. John F. Nash, Jr., *The Bargaining Problem*, 18 ECONOMETRICA 155 (1950).

comes using cooperative game theory. The use of noncooperative game theory to study bargaining was pioneered by Kalyan Chatterjee and William Samuelson, who asked which bargaining offers by a buyer and a seller would constitute an equilibrium in a game of incomplete information.[82]

Their procedure consists of the sealed-bid submission of offers, b and s, by a buyer and a seller, respectively. If $b \geq s$, a transaction is consummated at a price equal to $kb + (1 - k)s$, where $0 \leq k \leq 1$. The Chatterjee-Samuelson procedure has been shown to be more efficient than any other procedure in maximizing the parties' expected profit in equilibrium.[83] However, the Chatterjee-Samuelson procedure has an infinite number of inefficient, asymmetric equilibria as well.[84] More importantly, at all of these equilibria, the parties have a natural incentive to exaggerate their reservation prices, except when the prices are extreme, in order to maximize their expected profit.[85]

Mechanisms to address the honesty-in-bargaining problem were proposed in a 1996 article by political scientist Steven J. Brams and mathematician D. Marc Kilgour[86] and in a 2012 article by Steven J. Brams, D. Marc Kilgour, and economist Todd R. Kaplan.[87]

82. Kalyan Chatterjee & William Samuelson, *Bargaining Under Incomplete Information*, 31 OPERATIONS RES. 835 (1983).

83. Roger B. Myerson & Mark A. Satterthwaite, *Efficient Mechanisms for Bilateral Trading*, 29 J. ECON. THEORY 265 (1983).

84. *See* W. Leininger, P. B. Linhart & R. Radner, *Equilibria of the Sealed-Bid Mechanism for Bargaining With Incomplete Information*, 48 J. ECON. THEORY 63, 66 (1989).

85. For a detailed explanation of the incentive to make dishonest offers in the Chatterjee-Samuelson procedure, and of how some of the bargaining procedures to be discussed can reveal true reservation prices if not induce truthful offers, see STEVEN J. BRAMS, NEGOTIATION GAMES: APPLYING GAME THEORY TO BARGAINING AND ARBITRATION 34–38 (rev. ed.2003). Note that Vickery-Clarke-Groves ("VCR") mechanisms also induce honest disclosure of reservation prices, but they are not ex post budget-balanced and individually rational. The original work of these researchers is given in Theodore Groves, *Incentives in Teams*, 41 ECONOMETRICA 617 (1973); Edward H. Clarke, *Multipart Pricing of Public Goods*, 11 PUBLIC CHOICE 17 (1971); William Vickrey, *Counterspeculation, Auctions, and Competitive Sealed Tenders*, 16 J. FIN. 8 (1961).

86. Steven J. Brams & D. Marc Kilgour, *Bargaining Procedures that Induce Honesty*, 5 GROUP DECISION & NEGOTIATION 239 (1996).

87. Steven J. Brams, Todd R. Kaplan & D. Marc Kilgour, *A Simple Bargaining Mechanism That Elicits Truthful Reservation Prices* (Feb. 2012) (unpublished manuscript), *available at* http://politics.as.nyu.edu/docs/IO/2578/Bargainin_Feb_2012.pdf. Another procedure to encourage "reasonable settlement offers" is described in Robert H. Gertner & Geoffrey P. Miller, *Settlement Escrows,* 24 J. LEGAL STUD. 87 (1995), but it does not render honesty a weakly dominant strategy in the way that the bargaining procedures we analyze later do.

Of course, these are not the only mechanisms found in the literature, but they are particularly relevant here because they give the parties a *weakly dominant strategy*—at least as good and sometimes better than any other strategy—of honestly revealing their reservation prices, though these prices are not necessarily those used in the settlement that occurs (for reasons to be discussed later). By utilizing these mechanisms, the law can ensure that when there is not a settlement, it will be known whether one is possible, even though it may not be revealed to the parties themselves.

As we discuss further in Section II.C, the accurate disclosure of reservation prices can serve valuable social goals. The incentive structure induced by these mechanisms makes such honest disclosure in the parties' best interests, independent of the behavior of an opponent (because truthful revelation is a weakly dominant strategy, which we will say more about later).

Brams and Kilgour present the Bonus Procedure as a solution to the honesty problem in single-offer bargaining.[88] The latter is epitomized by the procedure of Chatterjee and Samuelson (1983), which, as noted earlier, Roger B. Myerson and Mark A. Satterthaite showed leads to a greater expected profit in equilibrium than any other bargaining mechanism.[89] Generally speaking, however, it induces the parties to shade their offers (for the buyer, downward from its reservation price unless it is already low; for the seller, upward from its reservation price unless it is already high—details to follow) to try to ensure as favorable a settlement as possible.

More precisely, under the Chatterjee-Samuelson procedure, a transaction occurs at a price equal to the mean of the buyer's and the seller's offers if they overlap (i.e., $b \geq s$). However, the Chatterjee-Samuelson procedure generally leads buyers and sellers to exaggerate their reservation prices, denoted by B and S, respectively.

Assume that the parties' reservation prices are independent and uniformly distributed over [0,1]. Then there is a linear, symmetric Nash equilibrium in which the buyer will offer[90]

88. In legal applications, bargaining will often be a single offer in practice. Even if there are multiple offers, this procedure would be effective for each offer. In the "thin" two-person markets that are the subject of this analysis, structured negotiations help to capture the efficiency usually found only when there is competition among many players in large markets.

89. Myerson & Satterthwaite, *supra* note 83, at 266.

90. This analysis also assumes that the players believe each other's reservation price to be so distributed.

$$b = \begin{cases} B & \text{if } 0 \le B \le 1/4 \\ (2/3)B + 1/12 & \text{if } 1/4 < B \le 1 \end{cases}$$

and the seller will offer

$$s = \begin{cases} (2/3)S + 1/4 & \text{if } 0 \le S < 3/4 \\ S & \text{if } 3/4 \le S \le 1 \end{cases}$$

The Chatterjee-Samuelson procedure thus encourages both parties to exaggerate their reservation prices, except when $B \le 1/4$ or $S \ge 3/4$, in which case one party truthfully reports its reservation price. But in these cases, there will be no settlement because even though the buyer will be honest if its price is not greater than $1/4$, and the seller will be honest if its price is at least $3/4$, the other party will not be honest in this circumstance, precluding all settlements except in the intermediate range, $B > 1/4$ and $S < 3/4$, given $B \ge S$.

In fact, B must exceed S by at least $1/4$ in order for there to be a settlement. While other distributions lead to different exaggerated offers, the buyer generally benefits from understating B, and the seller from overstating S, when the parties make equilibrium offers, b and s, respectively.

Is there an antidote to exaggeration and posturing? The Bonus Procedure solves the honesty problem by paying a bonus to the buyer and seller when their offers overlap (i.e., $b \ge s$) and they settle at the mean of their offers, $m = (b + s)/2$.[91] Brams and Kilgour prove that in a game of incomplete information in which the buyer does not know the seller's reservation price but believes it to be distributed according to some probability density function (not necessarily the uniform density function, which we assumed for the Chatterjee-Samuelson procedure earlier), the parties' dominant strategies, which we star, are to bid $b^* = B$ and $s^* = S$ (i.e., to be truthful) if the bonus given to each party is half the difference between their offers, $(b^* - s^*)/2 = (B - S)/2$.

Assume $B \ge S$. Then the buyer's expected profit with such a bonus will be

$$P_B(B,S) = B - m + (B - S)/2 = B - S$$

because $m = (B + S)/2$. Similarly, the seller's expected profit will be

$$P_S(B,S) = m - S + (B - S)/2 = B - S$$

If and only if $B < S$ will there be no settlement, resulting in a profit of 0.

91. *See* Brams & Kilgour, *supra* note 86, at 239.

Notice that the parties benefit equally when their reservation prices overlap, each receiving an expected profit of $B - S$. Because bidding one's reservation price is optimal whatever the other party does, these strategies constitute a dominant-strategy Nash equilibrium. (However, it is *weakly* dominant, because it may not always give a better outcome than any other strategy but, instead, one that is at least as good and in at least one instance better.)

Brams and Kilgour explain why it is optimal for the buyer to bid its reservation price for any offer s of the seller:[92]

> The optimality of $b = B$ . . . is not difficult to understand intuitively. The buyer gains $B - s$ when there is an exchange, and 0 when there is not. Obviously, the buyer would prefer an exchange [if and only if] $B > s$. By picking $b < B$, the buyer gains $B - s$ when $s \leq b < B$, but misses out on some profitable exchanges when $b < s < B$. On the other hand, if $B > b$, the buyer effects every profitable exchange (when $s < B$), but these also include some of negative value (when $B < s < b$). Thus, moving b away from B in either direction costs the buyer, so $b = B$ is optimal.[93]

A similar argument shows why the seller will not deviate from being truthful and bid $s = S$.

Paying a bonus equal to half the difference between the buyer and seller's offers if they settle alters the parties' incentives such that dishonesty becomes suboptimal. Specifically, were the buyer to make an offer different from his or her reservation price, he or she would either fail to make a purchase or make a suboptimal purchase (i.e., by paying a price above his or her reservation price).

The Bonus Procedure structures the bargaining incentives so that it is in the parties' interest to bid their honest reservation prices. Inducing such honesty, however, necessitates paying the bonus. This cost could fall on a regulatory agency, as discussed further in Part III *infra*. To obviate budget concerns, Brams and Kilgour show that a tax can be assessed to recoup the cost of paying the bonus. This tax must be assessed prior to bargaining but remains lower than the two sides' combined expected profit. We discuss such a taxing mechanism to render the Bonus Procedure budget-balanced in Subsection II.C.1.

As Brams and Kilgour acknowledge, a drawback of the Bonus Procedure is that it is vulnerable to collusion, whereby the parties

92. We substitute our notation for the buyer's and seller's offers and reservation prices, and correct one inequality, in this quotation.

93. Brams & Kilgour, *supra* note 86, at 244.

increase the size of the bonus they receive by submitting offers farther apart than their reservation prices ($(b - s) > (B - S)$), resulting in their receiving a bigger bonus. However, as Brams and Kilgour show, a "collusion equilibrium" is highly unstable because a party has no incentive to select a collusion strategy given that the other party has chosen its collusion strategy. To be sure, legal penalties could be imposed to raise the cost of collusion, at least on an expected-value basis, but collusion may not be easy to detect.

The Two-Stage Procedure that we discuss next may be preferable if the legal regime in question is particularly concerned with the possibility of collusion. But such robustness comes at a price, because the Two-Stage Procedure is not as efficient as the Bonus Procedure since not all profitable settlements are implemented.

B. The Two-Stage Procedure: Overview, Visualization, and Implementation

1. Overview of the Mechanism

The Two-Stage Procedure proposed in Brams, Kaplan, and Kilgour (2012) is another mechanism that makes the honest disclosure of reservation prices a weakly dominant strategy.[94] The mechanism is conducted with the assistance of a referee, which in the legal setting might be a computer program operated by a regulatory agency.

In stage 1 of the mechanism, the buyer and seller submit their reserve prices, $\hat{B}$ and $\hat{S}$, to a referee, which may not be their truthful reservation prices. However, Brams, Kaplan, and Kilgour prove that the players have weakly dominant strategies of being truthful under the Two-Stage Procedure, so we henceforth presume these prices are the parties' true reservation prices, B and S. If $B < S$, then the buyer is unwilling to pay as much as the seller is asking, so there is no settlement and the procedure ends.

If $B \geq S$, on the other hand, then there is the *possibility* of a settlement, and we proceed to stage 2. In stage 2, the buyer and seller submit to the referee their offers, b and s, respectively. If both offers are within the reservation price window $[S,B]$—that is, both are at least equal to S and do not exceed B—a transaction is consummated at the mean of the offers, $m = (s + b)/2$. If only one offer is in this window, the referee picks one of the two parties at random. If the chosen party's offer is the one in the window, a settlement occurs at that offer (b or s). Otherwise, there is no settlement.

94. Brams, Kaplan & Kilgour, *supra* note 87, at *5–11. We will briefly discuss a third honesty-inducing mechanism later.

As with the Chatterjee-Samuelson procedure and the Bonus Procedure, we assume under the Two-Stage Procedure that each party knows its own reservation price and believes that the other party's reservation price is distributed over some interval, which we suppose to be [0,1]. In the case of the Chatterjee-Samuelson procedure, we assumed the distribution to be uniform, making it equally likely that the other party's reservation price is at any point in [0,1]. If we make the same assumption for the Two-Stage Procedure, then the optimal offers of the buyer and seller in stage 2, which we star, are as follows:

$$b^* = \frac{B}{2}; \quad s^* = \frac{1 - S}{2}$$

These choices in stage 2, together with each party's truthful revelation of its reservation price, B or S, in stage 1, constitute the Nash-equilibrium strategies of the players under the Two-Stage procedure.

By (truthfully) choosing B and S in stage 1, the parties maximize the width of the interval, $[S,B]$, if any, in which a settlement occurs (at m, b, or s) without either party incurring a loss. But it is in stage 2, when $B \geq S$, that the parties can maximize their expected profit:

- by the buyer's selecting b so as to shade downward its reservation price (from B to $B/2$) by choosing the midpoint of the interval, $[0,B]$, in which S can fall; and
- by the seller's selecting s so as to shade upward its reservation price (from S to $(1 + S)/2$) by choosing the midpoint of the interval, $[S,1]$, in which B can fall.

Unlike the Chatterjee-Samuelson procedure, the offers need not overlap (i.e., $b \geq s$) in stage 2, but one or both must fall in $[S,B]$. The exaggeration of the reservation prices that occurs at stage 2—when the players report b^* and s^*—"affords" the parties the opportunity to be honest at stage 1 by truthfully reporting B and S.

The Two-Stage Procedure can be modified in various ways without affecting its honesty-inducing nature. First, the order of the stages can be reversed or even implemented simultaneously.[95] The Two-Stage Procedure may be combined with the Chatterjee-Samuelson procedure to improve the efficiency of the mechanism but reduce the gains from truth-telling.[96] From an implementation perspective, the referee, after receiving the reservation price of one

95. *Id.* at *15.
96. *Id.* at *15–16.

party, can initiate the use of the procedure by disclosing to the other party that it received such information and asking for a response. We discuss initiation of the procedure by a government regulator in the context of enforcement and its mandatory use in Part III.

Perhaps the most compelling property of the Two-Stage Procedure is the incentive it gives the parties to reveal their reservation prices honestly in stage 1, regardless of whether an agreement is reached in stage 2. Indeed, the mere use of the mechanism over time can provide valuable information to regulators or other actors in the legal system regarding the distribution of reservation prices and the reason why a settlement was or was not reached. For example, learning how many agreements failed in stage 1 (i.e., because there was no overlap in reservation prices) as opposed to stage 2, in which one or both parties' offers did not fall within the reservation price window, can shed light on whether reservation prices (in stage 1) or offers (in stage 2) more frequently lead to bargaining failures.

It is worth noting that the most efficient of the three procedures we have discussed so far is the Bonus Procedure, which leads to a settlement whenever $B \geq S$ and so captures the maximum possible expected profit of $1/6 \approx 0.167$ when the parties' reservation prices are uniformly distributed over [0,1]. (Later we show how this figure is obtained.) Its drawbacks are that a third party must pay each of the bargainers a bonus and that it is vulnerable to collusion. The next most efficient is the Chatterjee-Samuelson procedure, which provides an expected profit of $9/64 \approx 0.141$, but is not honesty inducing. Finally, the Two-Stage Procedure provides an expected profit of $1/8 = 0.125$ which makes it $(1/8)/(9/64) = 8/9 \approx 0.89$, or almost 90% as efficient as the Chatterjee-Samuelson procedure and, like the Bonus Procedure, honesty-inducing.[97]

We next show how the Two-Stage Procedure can be made more transparent to the parties via a visualization of the different possible outcomes that can occur under it. We then discuss ques-

97. *See id.* at *14. Another honesty-inducing procedure, called the "Penalty Procedure," makes the probability of implementing a settlement, given $B \geq S$, to be a function of the *amount* of overlap of B and S (i.e., $B - S$), but its expected profit is only $1/12 \approx 0.083$, which makes it $(1/12)/(9/64) = 16/27 \approx 0.59$, or less than 60%, as efficient as the Chatterjee-Samuelson procedure. *See* Brams & Kilgour, *supra* note 86, at 248. For a more detailed comparison of the three procedures, see D. Marc Kilgour, Steven J. Brams & Todd R. Kaplan, *Three Procedures for Inducing Honesty in Bargaining*, Proceedings of the 13th Conference of Theoretical Aspects of Rationality 170 (2011).

tions of implementation and possible extensions of this procedure. Later, in Subsection II.C.2, we describe more fully how regulatory policy can benefit from obtaining accurate information on reservation prices, where we also describe how the parties might "transcend" the limitations of the procedure.

2. Visualization

Because the Two-Stage Procedure is not as straightforward as the Chatterjee-Samuelson procedure or the Bonus Procedure, it is helpful to illustrate the conditions under which a settlement is or is not reached:

Stage 1. The buyer and seller submit their *reservation prices*, B and S, to a referee. If these prices do not overlap (i.e., $B < S$), there is no settlement, and the procedure ends:

$$0\_\_\_\_\_B\_\_\_\_\_S\_\_\_\_\_1$$

If $B \geq S$, there is an *overlap interval*, $[S,B]$, and the procedure goes to stage 2:

$$0\_\_\_\_\_S\_\_\_\_\_B\_\_\_\_\_1$$

Stage 2. The buyer and seller submit their *offers*, b and s, to the referee, which can produce a settlement in three different ways:

(i) If both b and s fall in the overlap interval, whether they do not crisscross because $s > b$ (first diagram) or do because $b > s$ (second diagram), the settlement price is the mean $m = (b + s)/2$:

$$0\_\_\_\_\_S\_\_b\_\_\_\_\_m\_\_\_\_\_s\_\_\_\_\_B\_\_\_\_\_1$$

$$0\_\_\_\_\_S\_\_s\_\_\_\_\_m\_\_\_\_\_b\_\_\_\_\_B\_\_\_\_\_1$$

(ii) If only b is in the overlap interval, then the settlement price is b with probability $1/2$:

$$0\_\_\_\_\_S\_\_\_b\_\_\_\_\_B\_\_\_s\_\_\_\_\_1$$

(iii) If only s is in the overlap interval, then the settlement price is s with probability $1/2$:

$$0\_\_\_\_\_b\_\_\_S\_\_\_\_\_s\_\_\_B\_\_\_\_\_1$$

In both cases (ii) and (iii), there is no settlement with probability $1/2$, even though one party's offer is inside the overlap interval (the mean of the parties' offers, m, may or may not be inside). In addition, there is no settlement with certainty if both b and s fall outside the overlap interval, even though m may fall inside (not shown).

As indicated earlier, Brams, Kaplan, and Kilgour (2012) prove that this mechanism renders it optimal for the parties to be truthful about their reservation prices, B and S, in stage 1, independent of

their beliefs about the other party's reservation price. However, the parties' optimal offers, b and s, in stage 2 do depend on these beliefs (defined by a probability distribution over the other party's reservation price) and do not simply duplicate B and S.

In fact, because the settlement price uses one or both of b and s, the bargainers have an obvious incentive to exaggerate them: The buyer will always choose $b \leq B$, and the seller will always choose $s \geq S$, as we indicated earlier when the parties' beliefs are given by a uniform distribution over the other party's reservation price. Consequently, one or both of the parties' offers may fall outside the overlap interval.

If exactly one of b or s falls inside, then there will be a settlement only with probability $1/2$, not certainty. This uncertainty, for which one inside offer is necessary but not sufficient to produce a settlement, helps to induce the bargainers to be truthful about B and S. Moreover, even when the mechanism fails to produce a settlement, it does reveal—if it continues to stage 2 because the reservation prices overlap—that these prices *allow* for a mutually profitable settlement.

3. Implementation

The implementation of the Two-Stage Procedure is straightforward. Although the referee could be a person, his or her task is completely mechanical. Hence, we propose that implementation be by a computer program, to which the parties input, separately and independently, B and S. The mechanism then determines if $B \geq S$ in stage 1. If not, there is no settlement, and the procedure ends.

If there is overlap, the players input, separately and independently, b and s. Provided that there is either *double overlap* (both b and s are in $[S,B]$) or *single overlap* (only one of b or s is in $[S,B]$), a price is determined according to the rules of stage 2.

It is worth noting that the mechanism could be initiated by just one party (say, the buyer), who would input B and then invite the seller to use the mechanism—using e-mail or some other form of communication—to input S. If the seller agrees, the mechanism would proceed as already discussed.

If the seller refuses, the buyer would be sent a confidential and dated statement, e.g., in the form of an affidavit, that he or she proposed B, which could then be used as evidence (e.g., in a judicial proceeding) that he or she made a good-faith offer to try to reach a settlement. We believe that the willingness of one party to input his or her reservation price, and possibly use it later as evi-

dence of his or her commitment to a settlement, might well induce the other party to follow suit and use the mechanism.

As noted earlier, the order of stages 1 and 2 can be reversed without changing the incentive of the mechanism to induce the parties to be truthful about their reservation prices. Here is how it would work: In stage 1, the parties would submit their offers; in stage 2, they would submit their reservation prices. If the offers crisscross in stage 1 ($b \geq s$), the referee would announce that there is a settlement price—the mean of the offers, $m = (b + s)/2$—and the procedure would end. If the offers do not overlap, each party would be asked in stage 2 to submit his or her reservation price without knowledge of the other party's stage 1 offer. The settlement, or lack thereof, would then be exactly the same as that in which the submission of the reservation prices precedes the submission of offers.

If the offers are made first (i.e., in stage 1), they can be thought of as "posted prices." If the offers do not overlap at this stage, in stage 2 each party would have an incentive to be truthful about his or her reservation price to ensure, insofar as possible, that it overlaps the other party's offer (i.e., posted price), because a party's reservation price will not be the settlement—the overlapped offer (with probability $1/2$) will be if there is single overlap, or the mean of the two offers will be if there is double overlap. Of course, if the initial offers crisscross in stage 1, there will also be double overlap in stage 2, which is why there is no need to proceed to stage 2.

Because the two stages can be reversed without changing the incentives of the players to be truthful about their reservation prices, their order does not matter. Therefore, we can as well assume that the parties submit their offers and reservation prices simultaneously, as we indicated earlier.

Practically speaking, however, the bargainers will probably prefer to proceed in stages. Whether they submit their (i) reservation prices first or (ii) their offers first, the rules allow for the procedure to terminate in stage 1 if either the reservation prices *do not* overlap in (i), or the offers *do* cross in (ii). Thereby, going in stages renders the mechanism simpler, possibly needing only one stage, without strategic consequences. Because it is not evident whether the parties will prefer (i) or (ii), we recommend that the choice be up to them, unless, of course, they prefer the simultaneous submission of both their offers and reservation prices.

Assume that the parties choose (i), so they submit their reservation prices first. Then if stage 2 is reached, they know that there is

an overlap interval and, therefore, that there is the *potential* for a mutually profitable settlement. If the mechanism fails to produce a settlement in stage 2, we recommend that the parties be told why it failed—either because both parties' offers, *b* and *s*, were outside the overlap interval, or one offer was inside but it was not implemented, which occurs with probability 1/2 (*which* party's offer was inside may or may not be revealed).

If the mechanism fails in stage 2, the parties might still try to find a settlement by other means, such as informal bargaining, mediation, etc. We stress, however, that under the mechanism, the parties must assign probability zero to the possibility that they could benefit further; otherwise, their incentive to be truthful in stage 1 will be compromised. Thus, for example, they might be told that the procedure cannot be used again for six months, or some other time period that signals that a settlement that they hoped for is "dead in the water" for a significant time—the implication being that they should take the procedure seriously when it is first tried.

But there is nothing in the mechanism, after it has been *unsuccessfully* tried, that prevents the parties from continuing to negotiate with each other—in effect, to transcend the limitations of the mechanism. So we ask: Can they do anything to assuage their dissatisfaction, and possibly escape the failure of the mechanism in stage 2, when it is known that their reservation prices overlap?

We suggested earlier that the parties be told whether the mechanism's failure in stage 2 was because both their offers were outside the overlap interval, or only one party's offer was inside and it was not selected with probability 1/2. If both offers were outside, there would appear to be not much more that can be done except exhort the parties to try harder next time—if there even is a next time (but see below for a possible resolution to even this unpromising scenario).

More promising, it seems, is the situation in which exactly one party's offer is inside. Then, if both parties are agreeable, there are two plausible ways in which their dispute can be resolved:

- Make the inside offer the settlement with certainty; or
- Make the settlement the inside offer averaged with the other party's reservation price.

In either case, the settlement will be inside the overlap interval, with the latter more favorable to the party that made the inside offer.

Both "solutions," of course, would alter our mechanism and, in particular, undermine the incentives of the parties to be truthful about their reservation prices. Hence, we do not recommend ap-

pending either to the mechanism in a possible stage 3, but instead suggest that if the mechanism fails in stage 2, the parties be asked whether either option to resolve their dispute would be acceptable to them.

Only if both parties agree would an option be used, which gives each party a veto on continuation. Presumably, the inside party would seek the latter solution and the outside party would push for the former (a compromise would be that the average of these two solutions be implemented).

After being told the settlement price, the parties may or may not be given the option of backing out of the settlement, which a party may want to do if the settlement price is very close to its reservation price. The option of backing out should make the parties more willing to try to "rescue" a settlement that is mutually profitable but which the mechanism failed to produce.

Now consider how a resolution might be achieved if *both* offers are outside the overlap interval in stage 2. The parties might be allowed to make new offers, in successive rounds, until at least one party's offer goes inside the interval. If both offers go inside on the same round, the average would be the settlement; otherwise, the settlement would be the single offer that goes inside first.

In making successive offers, an optimal strategy is *not* to inch very slowly toward the other party's reservation price, because the other party could "beat you to the punch" and go inside first—and still, on average, be quite close to your reservation price and far from its own. While the idea of making successively better offers to try to converge on a settlement is the way real-life bargaining often occurs, it does not always get the bargainers to a settlement. By contrast, the aforementioned extensions of the Two-Stage Procedure, which would require both parties' acceptance to be implemented, would do just that.

Because, as noted earlier, these "fixes" to a failure in stage 2 *force* a settlement, they will, if implemented, affect how truthful the parties will be in reporting their reservation prices. Accordingly, we do not propose them as "add-ons" to our mechanism but, instead, as separate procedures that both parties can, if they wish, decide to use if they fail to reach a settlement in stage 2.

But, we emphasize again, if these fixes are anticipated, they alter the parties' incentives to be truthful, knowing that they might have the possibility of escaping the failure of the mechanism. We mention them only to make the point that the theoretical conditions that induce the parties to be truthful might, *in practice*, be renegotiated, especially if the parties are desperate for a settlement

that they know, from the overlap of their reservation prices in stage 1, is within their grasp.

C. Extensions of the Procedures for Legal Applications

We next describe two extensions of the Bonus and Two-Stage Procedures to increase their appeal for legal applications. First, we show how an incentive-neutral tax can be incorporated into the Bonus Procedure to render it budget-balanced. Second, we show how the truthful information elicited under both the Bonus Procedure and the Two-Stage Procedure can be used to improve regulatory policy.

1. A Budget-Balanced Bonus: Taxing Under the Bonus Procedure

Both the Bonus and Two-Stage Procedures make honesty a weakly dominant strategy, or part of one, thereby revealing the bottom lines of the parties. However, the Bonus Procedure requires paying a bonus to induce the honest disclosure of reservation prices. Surprisingly, perhaps, the bonus-giver can completely recover its bonus through taxation without altering the incentive of the parties to be truthful. This means that employing the Bonus Procedure could be entirely budget-balanced and revenue-neutral for a regulatory agency.

To understand how such a tax would work, recall that the Bonus Procedure yields the maximum possible expected profit, or *surplus*, of 1/6 to the parties when their reservation prices are uniformly distributed over [0,1]. This surplus is the expected difference between B and S, integrated over the region in which $B \geq S$:[98]

$$\int_0^1 \int_S^1 (B - S)\,dB\,dS = \int_0^1 \left[\frac{B^2}{2} - BS\right]_S^1 ds$$

$$= \int_0^1 \left[\left(\frac{1}{2} - S\right) - \left(\frac{S^2}{2} - S^2\right)\right] dS$$

$$= \int_0^1 \left[\frac{1}{2} - S + \frac{S^2}{2}\right] dS$$

98. For those with little or no knowledge of calculus, the definite double integral shown below can be interpreted as follows: It first "sums" B from S (when $B = S$) to 1; it then "sums" S from 0 to 1, thereby picking up the surplus, $B - S$, over the entire region in which $B = S$.

This surplus is what the two parties *in total* receive when their reservation prices overlap and they are truthful. Because the Bonus Procedure doubles this amount, *each party* receives 1/6. By paying back one-half of this amount, or 1/12 each, to the bonus-giver, the parties can repay the total of 1/6 that the bonus-giver gave them to be truthful and still receive a surplus of 1/12 each. Consequently, the parties will have an incentive to be honest in reporting their reservation prices, netting, on average, all the gains from what honesty allows them to share. Evaluation of the previous integral yields:

$$= \frac{S}{2} - \frac{S^2}{2} + \frac{S^3}{6}\Big]_0^1 = \frac{1}{6}$$

However, one must be careful in interpreting this result.[99] The bonus-giver, on average, will recover its bonus and so break even. As for the parties, they must pay the tax even if there is no settlement. In fact, the tax must be charged before the parties report their reservation prices, B and S, because otherwise there will be some occasions (low values of B or high values of S) when the buyer and seller will be virtually certain that their profit at the honesty equilibrium will be less than the tax.

Consequently, they will have no motivation to pay the tax at these times. For this reason, the Bonus Procedure does not satisfy Myerson and Satterthwaite's *interim* individual-rational condition, though it does satisfy their *a priori* individual-rational condition.[100] That is, while it is in each party's interest to "play the game" over a long series of trials—for example, in repeated negotiations between labor and management—in any single trial a party may find it unprofitable because the likely profit it will receive from a settlement will be low or nonexistent.

The rationality of paying a tax, then, rests on an expectation over a series of plays, on some of which the parties will benefit but on others they will not. In fact, even if $B < S$, and there is therefore no settlement and zero profit to the parties, the tax still must be paid. Yet the strategy of honesty remains weakly dominant even after the tax is charged, because a constant tax of 1/12 subtracted from $P_B(B,S)$ and $P_S(B,S)$ does not affect their maximization.

It is not hard to generalize the tax calculation under the Bonus Procedure to prior distributions other than the uniform. With an

99. Most of the discussion in this Section is adapted from STEVEN J. BRAMS, NEGOTIATION GAMES: APPLYING GAME THEORY TO BARGAINING AND ARBITRATION 44–45 (1st ed. 1990).

100. *See* Myerson & Satterthwaite, *supra* note 83, at 268.

appropriate tax, the bonus-giver can always break even in the long run, provided that there is no collusion by the parties.

Even without collusion, however, there may be some practical difficulties in determining an appropriate tax and implementing the Bonus Procedure. For example, the probability distribution(s) over the parties' reservation prices, assumed as common knowledge in our model, may not be known or agreed to by the parties. In this situation, a trial-and-error taxing procedure could be used to determine an appropriate tax. Presumably, what the bonus-giver takes in by taxing—minus its costs of administering the Bonus Procedure—should be paid back in bonuses to the parties.

Adjustments in the tax may have to be made from time to time, especially if there is collusion between the buyer and the seller. Given some experience with a series of settlements, for example, the bonus-giver might decide it needs to raise the future tax to break even. The parties, knowing this possibility might arise—and perhaps facing severe sanctions if caught colluding—would presumably have little or no incentive to try to deceive the bonus-giver, at least in principle. The tax may raise enforcement challenges, such as preventing parties from attempting a transaction outside of the required procedure. However, to the extent that the tax discourages transactions from occurring at all, it can serve as a policy lever to induce the economically optimal level of transactions of this type.

Although bonuses are not needed in the Two-Stage Procedure to induce honesty, this procedure does not always produce settlements when the parties' reservation prices overlap, rendering it less efficient. Nonetheless, both the Bonus Procedure and the Two-Stage Procedure offer substantial advantages over the Chatterjee-Samuelson procedure and over unstructured bargaining by encouraging the truthful disclosure of reservation prices. Such information, as we next show, can be utilized to improve regulatory policy.

2. Utilizing Truthful Information to Improve Regulatory Policy

The Bonus Procedure and the Two-Step Procedure each provide an important social benefit—namely, eliciting truthful information that may be utilized to improve regulatory policy. The Two-Stage Procedure is especially well-suited to this information-providing function, because once stage 1 is completed, the parties know immediately whether there is the basis for a settlement.

Indeed, some parties may seek to engage in the Two-Stage Procedure simply to discover whether there is overlap in the their reservation prices prior to making an actual offer. Because there is no

bonus-giver, a regulator need not worry about possible collusion between the buyer and seller.

Assume that the Two-Stage Procedure advances to stage 2. While it tells the parties that a profitable settlement is possible, it provides the referee with more information. Specifically, if the referee is a regulator, he or she will learn whether a failure at stage 2 is because the parties' offers were outside the reservation window or because only one offer was in the reservation window but the random selection process caused termination of the procedure.

Regulators may also obtain descriptive data from each stage, such as the mean and distribution of reservation prices and offers, which they could use in administering the procedure in the future. An analysis of such data could provide, for example, a better understanding of why bargains fail at each stage. Thus, repeated failures at stage 1 might indicate that regulation is improperly encouraging negotiation in situations when it is premature, and regulators should instead discourage negotiation until an impasse is imminent or has already occurred.

At stage 1, repeated failures might suggest that the parties are engaging in opportunistic bargaining rather than being completely truthful. This information could then be used to improve the implementation of the mechanism (i.e., by increasing legal penalties for bargaining in bad faith). A regulator might also encourage good-faith bargaining by providing data on the reservation prices that led to settlements in similar situations in the past.[101] This information may send the message that extreme deviations from previous successful uses of the Two-Stage Procedure will be viewed suspiciously.

Information on the reservation prices may produce other benefits. Suppose, for example, that under the Two-Stage Procedure such information shows that some socioeconomic groups systematically report a small overlap between their reservation prices, but the gap between their reservation prices and their offers is large. Consequently, even though the parties' reservation prices overlap, their offers tend to fall outside the narrow window for a settlement pro-

101. In principle this should not be necessary: Because the parties have weakly dominant strategies, they cannot do better than be truthful, whatever the past record is. Still, it may helpful, especially for inexperienced negotiators, to know the record of success of different strategies. For example, if a union president is not sure how much he or she can compromise without losing the support of rank-and-file union members—and possibly being voted out of office—the record may serve as a guide to his or her determining a reservation price that otherwise might be difficult to estimate.

vided by the reservation prices. In addition, if one party is shading much more than the other, this might suggest power disparities that are inhibiting successful negotiation. Given such information, regulators would be better armed to take corrective measures.

In the case of the Bonus Procedure, regulators are more likely to observe a different kind of pattern—namely, that the parties' reservation prices overlap to an unusually great extent, rendering $B - S$ unexpectedly large. This would suggest collusion in order to "soak" the bonus-giver. If the bonus-giver is the government, then citizens might rightfully complain that their tax dollars are being wasted to promote a settlement. In such a situation, the public interest might be better served if the parties use the Two-Step Procedure, even though it is less efficient.

A possible solution to the opposite biases of the Two-Stage Procedure and the Bonus Procedure might be for a regulator not to specify in advance which honesty-inducing procedure he or she will use. To protect themselves whichever procedure is used, the parties would be well advised neither to understate nor overstate B and S—that is, to be *entirely* truthful—and not exaggerate too much their offers if the Two-Stage Procedure is used. Although truth-telling is a weakly dominant strategy under both procedures, keeping the parties in the dark before revealing the procedure to be used could, in effect, reinforce their incentive to tell the truth about their reservation prices and then, under the Two-Stage Procedure, make reasonable offers.

III.
APPLYING MECHANISM DESIGN TO SETTLEMENT NEGOTIATIONS AND SECURITIES REGULATION

This Part applies mechanism design in two legal contexts: settlement negotiations and securities regulation. Intuitively, the former seems well-suited for bargaining procedures, because traditional settlement negotiation is often slow, arduous, and unsuccessful. Indeed, the existence of mandatory pretrial settlement negotiations strongly demonstrates that excessive litigation imposes a substantial cost on society. Moreover, certain contexts (e.g., labor disputes) impose additional harm on third parties. Mandatory bargaining mechanisms to reduce transaction costs are plainly justified

to reduce negative externalities and to paternalistically facilitate agreement.[102]

Securities regulation might seem an unusual arena in which to apply mechanism design. Yet mechanism design is useful whenever a legal regime imposes a suboptimal mandatory rule because of the high transaction costs arising from unstructured negotiation with asymmetric information. Mechanism design permits obtaining contractual equilibria that are more efficient than a mandatory rule, while retaining low transaction costs and the positive externalities of informing regulators regarding parties' reservation prices and the conduct of negotiations.

A. Settlement Negotiations

1. An Overview of Mandatory Pretrial Settlement Negotiations

Pretrial settlement has long been considered a better outcome for dispute resolution than the litigation process.[103] The Civil Justice Reform Act of 1990 and the Federal Rules of Civil Procedure have mandated pretrial settlement conferences in the federal courts,[104] and nearly every state has provided for a similar mechanism.[105] There is a vast body of alternative dispute resolution schol-

102. For a general discussion of civil procedure viewed through the lens of bargaining, negotiation, and contract law, see generally Judith Resnik, *Procedure as Contract*, 80 NOTRE DAME L. REV. 593 (2005).

103. *See, e.g.*, Stephen McG. Bundy, *The Policy in Favor of Settlement in an Adversary System*, 44 HASTINGS L.J. 1 (1992); Robert D. Cooter & Daniel L. Rubinfeld, *Economic Analysis of Legal Disputes and Their Resolution*, 27 J. ECON. LITERATURE. 1067, 1074–76 (1989); Samuel R. Gross & Kent D. Syverud, *Don't Try: Civil Jury Verdicts in a System Geared to Settlement*, 44 UCLA L. REV. 1, 3 (1996); Kent D. Syverud, *The Duty to Settle*, 76 VA. L. REV. 1113 (1990); David M. Trubek et al., *The Costs of Ordinary Litigation*, 31 UCLA L. REV. 72, 122 (1983). *But see* Owen Fiss, *Against Settlement*, 93 YALE L.J. 1073, 1075 (1983) ("Settlement is for me the civil analogue of plea bargaining . . . although dockets are trimmed, justice may not be done."); Judith Resnik, *Managerial Judges*, 96 HARV. L. REV. 374, 424–25 (1982).

104. *See* Civil Justice Reform Act of 1990, 28 U.S.C. § 471, 473(b)(5) (2012) (authorizing district courts to require that "representatives of the parties with authority to bind them in settlement discussions be present or available by telephone during any settlement conference"); *see also* G. Heileman Brewing Co. v. Joseph Oat Corp., 871 F.2d 648, 652 (7th Cir. 1989) (holding that district courts may "order represented parties to appear at pretrial settlement conferences"); *cf.* FED. R. CIV. P. 16(a)(5) ("In any action, the court may order the attorneys and any unrepresented parties to appear for one or more pretrial conferences for such purposes as . . . facilitating settlement.").

105. *See, e.g.*, ALA. R. CIV. P. 16(a)(5); CAL. R. CT. 3.1380(a); CONN. R. SUPER. CT. 14-13; FLA R. CIV. P. 1.200; ILL. SUP. CT. R. 218; MASS. R. CIV. P. 16; MINN. R. PRAC. ANN. 305.02; NEV. R. CIV. P. 16.1(b); 22 N.Y. COMP. CODES R. & REGS. Tit. 22 § 202.12(a), (c)(5); PA. R. CIV. P. 212.3(b); WA. R. SUPER. CT. CIV. CR. 16.

arship examining the "art" of settlement negotiations in light of the "psychological, sociological, and communicational principles that influence other interpersonal relations."[106] This literature provides a helpful set of practical recommendations to facilitate settlement in unstructured negotiation.[107]

Interestingly, economic analysis has been employed, primarily in a descriptive sense, to model parties' incentives throughout the litigation and settlement process.[108] Law and economics scholarship has yet to address the normative question of whether structured procedures might be mandated to fundamentally alter parties' incentives to reduce transaction costs in settlement negotiations. Alternative dispute resolution scholars have compared the

106. CHARLES B. CRAVER, EFFECTIVE LEGAL NEGOTIATION AND SETTLEMENT § 1.01 (4th ed. 2001); *see also*, Richard Birke & Craig R. Fox, *Psychological Principles in Negotiating Civil Settlements*, 4 HARV. NEGOT. L. REV. 1 (1999); Russell B. Korobkin, *Psychological Impediments to Mediation Success: Theory and Practice*, 21 OHIO ST. J. ON DISP. RESOL. 281 (2006); Russell Korobkin & Chris Guthrie, *Psychological Barriers to Litigation Settlement: An Experimental Approach*, 93 MICH. L. REV. 107, 109 (1994) ("[P]sychological processes create barriers that preclude out-of-court settlements in some cases."); Tristin Wayte et al., *Psychological Issues in Civil Law*, *in* 14 PERSPECTIVES IN LAW & PSYCHOLOGY 323 (James R. P. Ogloff ed., 2004).

107. *E.g.*, CRAVER, *supra* note 106, at § 5.01–§ 9.03 (discussing techniques for effective unstructured negotiation such as "assessing negotiator personalities," "establishing negotiation tone," and "questioning"). Similar literature exists for negotiating corporate transactions. *E.g.*, Fred Tannenbaum, *The Second Half of Smart: How to Temper Your Intelligence and Become a More Effective Deal Lawyer*, PRACTICAL LAWYER Oct. 2006, at 25.

108. The literature on this topic is vast. The primary strand of analysis utilizes the so-called standard model, which models settlement as deriving from agreement on the expected value of a lawsuit, "defined as the probability of liability multiplied by the expected judgment amount." Robert J. Rhee, *The Effect of Risk on Legal Valuation*, 78 U. COLO. L. REV. 193, 194 (2007) (citing John P. Gould, *The Economics of Legal Conflicts,* 2 J. LEGAL STUD. 279 (1973); William M. Landes, *An Economic Analysis of the Courts*, 14 J.L. & ECON. 61 (1971); Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399 (1973); George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation,* 13 J. LEGAL STUD. 1, 6–30 (1984); Alan E. Friedman, Note, *An Analysis of Settlement,* 22 STAN. L. REV. 67 (1969)). Accordingly, litigation is considered a mere transaction cost that should be eliminated under the Coasean ideal of frictionless bargaining. Rhee, *supra*, at 194, 200 (crediting Posner and Landes with formulating the standard model of litigation settlement). More recently, a *real-options approach*, which incorporates the variance of expected outcomes into the settlement decision, has been proposed to incorporate the parties' "ability to adapt to new information into the model itself." Joseph A. Grundfest & Peter H. Huang, *The Unexpected Value of Litigation: A Real Options Perspective*, 58 STAN. L. REV. 1267, 1273–76 (2006). *But see* Rhee, *supra*, at 212–13 (critiquing Grundfest and Huang for assuming perfect information, misconstruing variance in the litigation context, and ignoring risk preferences).

strengths and weaknesses of different procedures, such as arbitration and mediation.[109] But this literature takes the distribution of incentives as given and does not rigorously examine whether a legal mandate to utilize a certain "negotiation procedure" might inherently improve the likelihood of reaching agreement.

We suggest that mandatory bargaining procedures for settlement negotiations are justified under the twin rationales of paternalism and externalities.[110] In this context, the social externality of nonagreement is particularly easy to see: lack of settlement imposes substantial costs on society in the form of maintaining the court system.[111] Moreover, in certain contexts, such as labor negotiations, third parties suffer direct injury when the parties cannot reach agreement. Any mechanism that encourages settlement between striking schoolteachers and the administration benefits the schoolchildren who suffer the detrimental effects of a suspended education because of the dispute.[112]

Indeed, the existence of statutory mandates to conduct pretrial settlement conferences implies legislative recognition of this social cost. Courts are also increasingly utilizing alternative dispute resolution, further indicating that the legal system is searching for ways to reduce the mounting costs of excessive litigation.[113] In effect, our proposal to utilize mandatory bargaining procedures can be viewed as simply one type of pretrial mediation, albeit with an approach based on altering incentives to reduce transaction costs inherent in direct negotiation and ultimately to foster agreement.

The economically rigorous nature of our proposed bargaining procedures relates to a paternalistic justification as well. Scholars have shown that a substantial source of transaction costs in settlement negotiations arises from the presence of asymmetric informa-

109. *E.g.*, Lela P. Love & Kimberlee K. Kovach, *ADR: An Eclectic Array of Processes, Rather than One Eclectic Process*, 2000 J. DISP. RESOL. 295; Frank E. A. Sander & Stephen B. Goldberg, *Fitting the Forum to the Fuss: A User-Friendly Guide to Selecting an ADR Procedure*, 10 NEGOTIATION J. 49 (1994).

110. *See* Ayres & Gertner, *supra* note 28, at 88–89.

111. *See generally* COMMITTEE FOR ECONOMIC DEVELOPMENT, BREAKING THE LITIGATION HABIT: ECONOMIC INCENTIVES FOR LEGAL REFORM 9–10 (2000), *available at* http://www.ced.org/pdf/Breaking-the-Litigation-Habit.pdf; A. Leo Levin & Denise D. Colliers, *Containing the Cost of Litigation*, 37 RUTGERS L. REV. 219 (1985) (discussing the rising cost of litigation). *But see* Charles Silver, *Does Civil Justice Cost Too Much?*, 80 TEX. L. REV. 2073, 2112 (2002) (concluding that "[s]tudies of procedural reform proposals consistently find that reforms save few dollars").

112. *See, e.g.*, Lane, *supra* note 5; Davey, *supra* note 5, at A1.

113. *See, e.g.*, Thomas J. Stipanowich, *ADR and the "Vanishing Trial": The Growth and Impact of "Alternative Dispute Resolution,"* 1 J. EMPIRICAL LEGAL STUD. 843 (2004).

tion.[114] The *inherent* incentives to posture and exaggerate are precisely those that can be mitigated very effectively by mechanism design.

To be sure, our mechanisms do not solve all problems. One knotty problem will be for the two parties to agree on the range in which B and S can fall, which we assumed earlier to be the interval [0,1]. Our model also assumes that each party has a distribution over the other party's reservation price in this interval. Then, given that $B = S$, the overlap interval, $[S,B]$, will be a subinterval in this range.

Presumably, the buyer will want the seller to think that the interval is bounded from below by a very low price (from which the buyer benefits), and the seller will want the buyer to think that the interval is bounded from above by a very high price (from which the seller benefits). Notwithstanding these incentives, the fact that litigants make high-low agreements in lawsuits, which fix the maximum and minimum amounts that the defendant will pay the plaintiff, suggests that reaching a consensus on a range, wherein B and S must lie, is not an impossible task. Given such a range, and assuming that each party has a distribution over it, the calculations we described earlier, which yield the parties' optimal strategies, can be made.

We are not naively suggesting that bargaining procedures can overcome every obstacle to settlement. In particular, a party that is very likely to emerge victorious at trial may have little incentive to negotiate. But many settlement negotiations are conducted against a backdrop of great uncertainty regarding the eventual outcome at trial. This is especially likely to be the case with labor disputes and complex commercial disputes, where there may be significant legal or factual ambiguity. Holding all else equal, reducing transaction costs in the bargaining process through a mandatory procedure can facilitate more effective agreement. To the extent that one side may have a greater likelihood of prevailing at trial, this information will be incorporated into both sides' reservation prices for the procedures that we propose.

114. *E.g.*, Lucian Ayre Bebchuk, *Litigation and Settlement Under Imperfect Information*, 15 RAND J. ECON. 404 (1984); Robert D. Cooter & Daniel L. Rubinfeld, *Economic Analysis of Legal Disputes and Their Resolution*, 27 J. ECON. LIT. 1067 (1989); John Kennan & Robert Wilson, *Bargaining with Private Information*, 31 J. ECON. LIT. 45 (1993).

2. Applying the Bonus and Two-Stage Procedures to Settlement Negotiations

The operation of the Bonus and Two-Stage Procedures in the context of settlement negotiations is fairly straightforward. The procedures could be conducted via a website or computer program operated by the court itself. Alternatively, the services of a private firm may be employed if it could implement the procedures more cheaply than the court.[115] The operator of the procedure would simply request the parties' offers and possibly tax them (if the Bonus Procedure were used), or request reservation prices and offers (if the Two-Stage Procedure were used), apply the algorithms as described in Part II *supra*, and report the settlement if the application of the procedures is successful. If not, extensions of the Two-Stage Procedure discussed in Subsection II.B.3 could be tried.

A more complex question is whether courts could compel the parties to engage in either of the procedures under current law. Numerous state statutes explicitly permit judges to order the parties to engage in court-supervised mediation,[116] and bargaining mechanisms would likely qualify as a form of mediation.[117] Congress has empowered federal district courts to compel mediation, including services provided by private-sector firms, by adopting local rules.[118] Moreover, the First Circuit has held that district courts have inherent power to compel the parties to engage in court-supervised me-

115. Fair Outcomes, Inc. provides game-theoretic dispute-resolution services, albeit using a different set of procedures. FAIR OUTCOMES, INC.: GAME THEORETIC SOLUTIONS FOR DISPUTES AND NEGOTIATIONS, http://www.fairoutcomes.com (last visited Jan. 24, 2014).

116. *See, e.g.*, FLA. STAT. § 44.102(2)(b) (2013) ("A court . . . [m]ay refer to mediation all or any part of a filed civil action for which mediation is not required under this section."); *see also* Holly A. Streeter-Schaefer, *A Look at Court Mandated Civil Mediation*, 49 DRAKE L. REV. 367, 373–77 (2001) (listing state statutes that permit mandatory mediation).

117. *See, e.g.*, FLA. STAT. § 44.1011(2) (2012) ("'Mediation' means a process whereby a neutral third person called a mediator acts to encourage and facilitate the resolution of a dispute between two or more parties."). Bargaining mechanisms would certainly constitute a neutral process. The only difficulty might lie in the definition of the term "person." Human supervision of an automated bargaining procedure might be sufficient to bring it within the definition of mediation under the Florida statute.

118. *See* 28 U.S.C. §§ 651–653 (2006). The definition of an "alternative dispute resolution process" under the federal statute is "any process or procedure, other than an adjudication by a presiding judge, in which a neutral third party participates to assist in the resolution of issues in controversy." § 651(a). The statute specifically authorizes "professional neutrals from the private sector" to provide ADR services. § 653(b). However, any ADR procedure must be adopted by local rule. § 651(b).

diation even absent a local rule: "[I]n the absence of a contrary statute or rule, it is perfectly acceptable for the district court to appoint a qualified and neutral private party as a mediator."[119] However, the court emphasized, "a mediation order must contain procedural and substantive safeguards to ensure fairness to all parties involved."[120]

Compelling the parties to engage in a bargaining procedure would therefore seem permitted in many instances. However, one caveat is in order. Mediation (unlike arbitration) is typically considered to be a nonbinding process. Indeed, the First Circuit justified its holding by emphasizing that "[i]n the context of non-binding mediation, the mediator does not decide the merits of the case and has no authority to coerce settlement."[121] Accordingly, while existing law may permit a court to compel the parties to *engage* in a bargaining procedure, it seems unlikely that a court could coerce the parties into *accepting* the transaction outcome as a binding judgment.

In our view, this does not present any difficulty. If the parties truthfully report their reservation prices, as would be in their best interest under both the Bonus and Two-Stage Procedures, each party's next-best alternative (i.e., proceeding to litigation) will be reflected in his or her reservation price. Accordingly, it would be in the parties' best interest to voluntarily accept the transaction if it can be effected by the bargaining mechanism. Indeed, one of the most powerful aspects of mechanism design is its capability to make the basis of win-win solutions—truthful revelation of reservation prices—at least part of a dominant strategy for both parties.[122]

3. Social Benefits of Honesty-Inducing Procedures for Settlement Negotiation

There are substantial informational advantages to utilizing an honesty-inducing procedure for settlement negotiation. One of the greatest benefits of the Two-Stage Procedure is that it permits the discovery of the parties' truthful reservation prices, and surmises

119. *In re* Atl. Pipe Corp., 304 F.3d 135, 146 (1st Cir. 2002) (citing *Ex parte* Peterson, 253 U.S. 300, 312 (1920)).

120. *Id.* at 147.

121. *Id.* at 146.

122. *Cf.* STEVEN J. BRAMS & ALAN D. TAYLOR, THE WIN-WIN SOLUTION: GUARANTEEING FAIR SHARES TO EVERYBODY 13 (2000) (discussing the notion of a win-win solution). For more technical details on fair-division algorithms, see STEVEN J. BRAMS & ALAN D. TAYLOR, FAIR DIVISION: FROM CAKE-CUTTING TO DISPUTE RESOLUTION (1996).

the distribution of such prices from related cases, even when settlements are not obtained. This information could be used in several ways.

Data on reservation prices could be utilized by lawmakers or regulators who exercise legislative authority over the substantive legal regime forming the subject of the dispute. This data could provide valuable feedback on whether individuals actually behave as the lawmaker or regulatory body expects. It could also show the economic conditions (i.e., combinations of reservation prices and substantive law) under which disputes arise. This could facilitate improving the law or regulatory regime constituting the subject of the dispute.

As a concrete example, consider a tort lawsuit over negligence where the alleged damages exceed the defendant's insurance coverage. Data regarding the prices at which the parties are *willing* to settle—even if no settlement is actually reached—could lead lawmakers or a regulatory body to set more accurate minimum insurance coverage requirements in the future. Or take a securities lawsuit over alleged fraud or misrepresentation. Obtaining aggregate data regarding the distribution of the parties' reservation values would indicate what the firm and plaintiffs were willing to pay (or receive in settlement). Any overlap in reservation prices—even if no settlement were actually reached—would indicate that agreement is possible in principle. This suggests that some system of compensation, such as given by the Bonus Procedure, might be justified to ensure agreement.

In these situations, the Two-Stage Procedure provides the additional benefit of discovering precisely *where* negotiations broke down. Failure in stage 1 means that agreement was not feasible, rendering adjudication appropriate. But failure in stage 2 indicates that agreement was possible, even though the Two-Stage Procedure failed to find it. While it may have been optimal for each party to shade its offer, we suggest that the law should consider ways to minimize the incentive to exaggerate (e.g., via the Bonus Procedure). It may also suggest suboptimal opportunistic negotiation, which could lead to penalties ex post. For example, the discovery of manipulative behavior in stage 2 offers in settlement negotiations might lead to the imposition of sanctions.[123]

123. *See* Chambers v. NASCO, Inc., 501 U.S. 32, 46–51 (1991) (finding that courts have inherent power to impose sanctions for "bad-faith conduct"). However, such conduct is not synonymous with exaggerated offers, which will in general be optimal. This means that judging when bargaining is in bad faith, unnecessarily opportunistic, or purely manipulative—not just strategic to optimize value for one-

In addition, aggregate data regarding the conduct of settlement negotiations could be reported to an oversight body, which could consider whether there might be procedural defects with the negotiation process. For example, voluntary demographic data could be obtained from litigants and examined to determine whether certain social groups are consistently underutilizing the bargaining mechanism. This finding might suggest that these users should receive special assistance and guidance when utilizing the procedures. Indeed, scholars have recently raised concerns regarding the lack of representation in settlement negotiations, and alternative dispute resolution more generally.[124] In addition to easily providing guidance for pro se litigants, another advantage of a bargaining procedure is that the negotiations could be conducted asynchronously, free from the pressure of intimidation and other traditional tactics of unstructured negotiation. They could even be conducted remotely, reducing the cost of physically attending negotiations.

B. Securities Regulation: Reconsidering Mandatory Blockholder Disclosure

1. Overview, Private Ordering, and Efficient Trade

At the core of the United States securities laws lies the fundamental principle of mandatory disclosure. When enacting the Securities Act of 1933 and the Securities Exchange Act of 1934, Congress emphasized the necessity of mandatory disclosure by pointing to the need to "provid[e] all [market] participants the opportunity to make informed investment judgments."[125] Academic scholars have also justified mandatory disclosure generally under principles of efficiency.[126]

Nonetheless, scholars have long questioned whether *every* disclosure requirement under the securities laws should be

self—will not be straightforward. Put another way, one cannot ask the parties to ignore, to their detriment, their self-interest.

124. *See generally, e.g.*, Stephan Landsman, *Nothing for Something? Denying Legal Assistance to Those Compelled to Participate in ADR Proceedings*, 37 FORDHAM URB. L.J. 273, 280–81 (2010).

125. Eric D. Roiter, *Illegal Corporate Practices and the Disclosure Requirements of the Federal Securities Laws*, 50 FORDHAM L. REV. 781, 783 (1982) (citing SEC v. Texas Gulf Sulphur Co., 401 F.2d 833, 849 (2d Cir. 1968) (en banc); H.R. REP. NO. 73-1383, at 5, 13 (1934)).

126. *E.g.*, John C. Coffee, Jr., *Market Failure and the Economic Case for a Mandatory Disclosure System*, 70 VA. L. REV. 717, 747–51 (1984).

mandatory.[127] Jonathan Macey and Jeffry Netter, for example, have argued that the mandatory disclosure of a blockholder's plans and purposes under Item 4 of schedule 13D is inefficient and should be replaced by a private ordering system where individual firms can opt in to require disclosure on a firm-specific basis.[128] Larry Ribstein has even advocated rendering the whole of affirmative disclosure under the securities laws optional, subject to a variety of exceptions to protect unseasoned investors.[129] These arguments follow from the basic economic principle that mandatory rules are generally less efficient than bargained-for terms.[130]

The latest debate over mandatory disclosure involves section 13(d) of the Securities Exchange Act of 1934. Section 13(d) requires investors who acquired over 5% of the beneficial ownership of a reporting company to disclose their equity stake within ten days.[131] As with most disclosure requirements under the securities law, blockholder disclosure was justified under a fairness rationale. When introducing the bill that would later become the Williams Act of 1968, Senator Williams condemned the contemporary practice of corporate raiders being able to operate under a "cloak of secrecy . . . while obtaining the shares needed to put him on the road to successful capture of a company."[132] The ten-day disclosure

127. *E.g.*, Frank H. Easterbrook & Daniel R. Fischel, *Mandatory Disclosure and the Protection of Investors*, 70 VA. L. REV. 669, 714 (1984) ("We have not constructed a compelling case for regulation of any sort, let alone for the particular regulations the SEC uses."); Henry Manne, *Economic Aspects of Required Disclosure under Federal Securities Laws*, *in* WALL STREET IN TRANSITION 23, 31–40 (H. Manne & E. Solomon eds., 1974).

128. Jonathan R. Macey & Jeffry M. Netter, *Regulation 13D and the Regulatory Process*, 65 WASH. U.L.Q. 131, 154 (1987) ("Interestingly, no one has ever explained why target firms could not themselves provide incentives for bidders to disclose the information required by the Williams Act if such disclosure would benefit shareholders. If shareholders of potential target firms find such information of value, they could make appropriate adjustments in their firms' articles of incorporation that would require the disclosure.").

129. Larry E. Ribstein, *Private Ordering and the Securities Laws: The Case of General Partnerships*, 42 CASE W. RES. L. REV. 1, 26–30 (1992).

130. *See* Ayres & Gertner, *supra* note 28, at 89, 89 n.15 (citing Anthony T. Kronman, *Specific Performance*, 45 U. CHI. L. REV. 351, 370 (1978) ("[E]x ante arguments for the efficiency of a particular legal rule assume that individuals remain free to contract around that rule, and a legal system that denies private parties the right to vary rules in this way will tend to be less efficient than a system that adopts the same rules but permits contractual variation.")).

131. Securities Exchange Act of 1934 § 13(d), 15 U.S.C.A. § 78m(d)(1) (2012).

132. 111 Cong. Rec. 28,258 (Oct. 22, 1965).

window was introduced after the SEC argued that preacquisition disclosure would be impracticable.[133]

Section 929R of the Dodd-Frank Wall Street Reform and Consumer Protection Act empowered the SEC to shorten this ten-day period by rule.[134] In a recent petition, the law firm Wachtell, Lipton, Rosen & Katz ("Wachtell Lipton") requested that the SEC exercise its rulemaking authority under section 929R and shorten the disclosure window.[135] Wachtell Lipton noted that activist hedge funds have utilized the ten-day window to acquire massive blocks of ownership far exceeding the 5% disclosure threshold.[136] In Wachtell Lipton's view, such stealth acquisitions contravene the purpose of the disclosure requirement enacted by the Williams Act.[137]

Law professors Lucian Bebchuk and Robert Jackson replied to Wachtell Lipton in an academic article, arguing that the ten-day window should be preserved because it provides an essential incentive for activist hedge funds to intervene in target companies.[138] Since hedge funds do not typically acquire controlling blocks of ownership, they must share the benefit of their activism pro rata with other investors. To make intervention worthwhile, hedge funds need to acquire shares during the ten-day window prior to disclosing their holdings.[139] These shares are cheaper than postdisclosure, because investors bid up the company's stock upon a schedule 13D announcement of hedge fund intervention. As fi-

133. *See* 112 Cong. Rec. 19,004 (Aug. 11, 1966).

134. Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. 111-203, § 929R, 124 Stat. 1866 (2010).

135. Letter from Wachtell, Lipton, Rosen & Katz to Elizabeth M. Murphy, Sec'y, U.S. Sec. & Exch. Comm'n (Mar. 7, 2011), *available at* http://www.sec.gov/rules/petitions/2011/petn4-624.pdf.

136. *Id.* at 6 (citing Maxwell Murphy, *How Bill Ackman Stalked J.C. Penney,* WALL ST. J. BLOG, (Oct. 8, 2010, 3:50 PM), http://blogs.wsj.com/deals/2010/10/08/how-bill-ackman-stalked-jc-penney/; Joann S. Lublin & Karen Talley, *Big Shoppers Bag 26% of J.C. Penney,* WALL ST. J., (Oct. 9, 2010), http://online.wsj.com/news/articles/SB10001424052748704657304575539880781136228).

137. *Id.* at 2–3.

138. Lucian A. Bebchuk & Robert J. Jackson Jr., *The Law and Economics of Blockholder Disclosure*, 2 HARV. BUS. L. REV. 39, 50 (2012). *But see* Adam O. Emmerich, Theodore N. Mirvis, Eric S. Robinson & William Savitt, *Fair Markets and Fair Disclosure: Some Thoughts on the Law and Economics of Blockholder Disclosure, and the Use and Abuse of Shareholder Power* 2 (Columb. L. & Econ. Working Paper No. 428, 2012), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2138945 (replying to Bebchuk and Jackson by reiterating the absence of "any explanation of how their position – their conception of how the Section 13(d) reporting rules should operate – is consistent with the clear purpose of the statute.").

139. *Id.* at *16–19.

nance research shows that hedge fund activism is beneficial for target companies, the SEC should retain the ten-day window as a form of compensation for hedge funds to monitor and discipline management.[140]

In a recent paper, Joshua Mitts takes a different approach.[141] Mitts argues that it is essential to consider the costs as well as the benefits of hedge fund activism. Empirical research in the management and accounting disciplines has shown that, with their extreme short-term orientation, hedge funds often exacerbate the pervasive problem of short-termism in corporate governance.[142] A lengthy blockholder disclosure window may encourage detrimentally excessive activism.[143] Moreover, delayed disclosure facilitates trading on asymmetric information, which imposes economic and noneconomic social costs.[144] To find the optimal duration for the blockholder disclosure window, Mitts proposes a private ordering solution akin to the approach taken in the proxy access context whereby the length of the disclosure window is set on a firm-specific basis by a shareholder amendment to the corporate bylaws.[145]

This is an important first step toward incorporating mechanism design into securities regulation. The current mandatory ten-day window likely encourages an inordinate level of hedge fund activism and trading on asymmetric information that is harmful to firms and society as a whole. But this doesn't go far enough. It is probably impossible to find a single, universal duration that would be optimal for every firm because the distribution of costs and benefits may vary dramatically.[146]

However, individual firms may be willing to accept a financial payment in exchange for delaying disclosure and hedge funds may be willing to pay for such a delay to keep their actions secret for as long as possible. Imposing a mandatory rule, even at a firm-specific level, may preclude more tailored solutions that better satisfy the needs of individual firms and their hedge-fund suitors, enhancing the possibility of welfare-enhancing settlements for both sides.[147]

140. *Id.*

141. Joshua Mitts, *A Private Ordering Solution to Blockholder Disclosure*, 35 N.C. CENT. L. REV. 203 (2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2180939.

142. *Id.* (manuscript at 27–30).

143. *Id.*

144. *Id.* (manuscript at 31–33, 35–37).

145. *Id.* (manuscript at 43–44).

146. *Id.* (manuscript at 39–40).

147. In the following discussion, we consider the efficiency of blockholder disclosure from the perspective of hedge funds' willingness to pay. Later, we relate

Admittedly, viewing disclosure as a tradable "good" is counterintuitive in the context of a system of mandatory rules such as securities regulation. But it is quite natural from an economic perspective, inasmuch as we consider solely the two "sides" to the transaction, i.e., hedge funds and the target firm. (We will consider society and the investing public in Subsection III.B.3.). Delayed disclosure imposes a certain cost on a firm by encouraging intervention by activist hedge funds that pursue an agenda of short-term profit maximization. Delayed disclosure also has value to hedge funds, because it permits the acquisition of shares at a below-market price, i.e., below the level that the market would have paid if it had the information regarding the hedge fund's accumulation.

Bebchuk and Jackson correctly point out that this discount is necessary to make activism worth it for institutional investors who take a noncontrolling stake, such as hedge funds. Yet it does not follow that they should be permitted to obtain this discount at no cost. Activism would still be profitable if the price of obtaining the discount were less than the gross profit from intervention. Consider the following two examples:

Hedge Fund	Buy @ (Pre-Disclosure)	Sell @ (Post-Disclosure)[148]	Cost of Activism	Gross Profit	Cost of Delayed Disclosure[149]	Net Profit/Loss
A	$ 25	$ 50	$ 20	$ 5	$ 10	$ (5)
B	$ 25	$ 50	$ 10	$ 15	$ 10	$ 5

this analysis to the costs and benefits of hedge fund activism to target firms and society as a whole.

148. Technically, this column should reflect the price at which the hedge fund sells the shares after intervention, i.e., at the conclusion of its holding period. However, this price would be similar to that of the postdisclosure "pop" if markets are at least temporally efficient over the short run, i.e., the share price incorporates immediately the effect of hedge fund intervention *in the short run*. Hedge fund activism may cost firms over the long run, but the finance literature strongly suggests that it brings short-term benefits, and these are reflected in the share price "pop" upon the schedule 13D filing, which discloses the blockholder acquisition. *See* Alon Brav et al., *Hedge Fund Activism, Corporate Governance, and Firm Performance*, 63 J. FIN. 1729, 1730 (2008) ("[T]he market reacts favorably to activism [immediately after disclosure], consistent with the view that it creates value. . . . We find that the positive returns at announcement are not reversed over time."); April Klein & Emanuel Zur, *Entrepreneurial Shareholder Activism: Hedge Funds and Other Private Investors*, 64 J. FIN. 187, 188 (2009) ("[H]edge fund targets earn 10.2% average stock returns during the period surrounding the initial Schedule 13D [disclosure]. . . . Furthermore, our target abnormal returns do not dissipate in the 1-year period following the initial Schedule 13D.").

149. This includes transaction costs associated with the purchase of delayed disclosure and reflects the total cost of delayed disclosure for the duration that would permit the acquisition of the number of shares such that the average price

For hedge fund A, imposing a requirement to purchase delayed disclosure at the price of $10 makes activism no longer profitable. But for hedge fund B, activism remains profitable even at this price.

This example illustrates that requiring hedge funds to purchase delayed disclosure would not necessarily eliminate hedge fund activism. But it is also unnecessary to take a static approach to pricing delayed disclosure. Why should every hedge fund pay the same price to delay disclosure? In the example above, it would be profitable for hedge fund A to intervene if the price of delayed disclosure was in the range of $[0,5]. An identical, fixed price for delayed disclosure across all hedge funds is plainly less efficient than pricing that varies with hedge funds' and firms' willingness to pay.

Of course, the current regulatory regime does not charge a fixed price to obtain a given level of delayed disclosure. It simply imposes a mandatory rule that disclosure is required after ten days. Yet this mandatory rule is essentially identical to charging a fixed cost of disclosure for those firms that desire to purchase shares after the ten-day window but are unable to do so. For them, the ten-day window imposes a cost on any purchase subsequent to the expiration of the window equal to the difference between the pre- and postdisclosure prices. When averaged with the shares that were purchased at discount during the ten-day window, this is equivalent to any other "price" the firm must pay to reach its desired level of ownership. Under Wachtell Lipton's proposal to shorten the disclosure window to one day,[150] this "cost of delayed disclosure" would jump dramatically for most hedge funds. However, the possibility of acquiring more than 5% ownership on that first day means that activism might still be profitable for the few firms that are able to do so and thereby reduce their "cost of delayed disclosure" sufficiently to obtain a profit.

This discussion demonstrates just how inefficient a mandatory rule would be when applied to all firms. But the same structural inefficiency would exist with firm-specific fixed disclosure durations. In that case, each firm would impose a different "cost of

equals the figure in the second column. A hedge fund could not acquire an unlimited number of shares because the average price eventually would rise above the profitable level. For the sake of simplicity, this analysis assumes that each hedge fund already has determined the duration of disclosure that would permit the optimal acquisition of shares in light of the average cost and other idiosyncratic constraints, such as portfolio diversification and total assets under management.

150. Letter from Wachtell, Lipton, Rosen & Katz, *supra* note 135, at 5. **R**

delayed disclosure" through different fixed disclosure durations under the same analysis described *supra* with a single universal disclosure window. Nonetheless, such an approach would still prevent some welfare-enhancing trade from occurring. Many hedge funds might be willing to intervene if the cost of delayed disclosure were slightly lower and firms might be willing to accept this price. The mere presence of different fixed durations does not imply efficient competition or even a "marketplace" of disclosure windows. This is particularly true in this type of thin market where target firms are not necessarily substitutes for each other, and changes to the disclosure duration would require undertaking the cumbersome process of amending corporate bylaws.

Moreover, while a fixed duration may have a similar effect as purchasing delayed disclosure for hedge funds, it is fundamentally different for target firms. Under a fixed disclosure window, the cost of delayed disclosure inherently imposed by being forced to purchase postdisclosure is paid to other shareholders selling to the hedge fund on the secondary market. However, if delayed disclosure were conducted as a sale transaction between hedge funds and target firms, this price would be paid directly to firms rather than other shareholders. This would fundamentally alter firms' cost-benefit calculation regarding hedge fund activism. Firms might conclude that the possibility of encouraging detrimental short-termism is acceptable if they will receive a monetary payment in exchange for this risk.[151] Under a cost-benefit analysis, the ability to sell delayed disclosure offsets the potential cost of hedge fund activism.

Viewed in this light, even a firm-specific mandatory disclosure rule is suboptimally inflexible. The costs and benefits of hedge fund activism are likely to vary from firm to firm, but they may also vary *across time* within a single firm. Unless we unrealistically assume that firms can alter their bylaws rapidly in response to changing conditions, there could be many Pareto-optimal transactions prevented by a fixed disclosure duration.

We have suggested throughout this discussion that the suboptimal nature of a mandatory rule derives from its inability to reflect hedge funds' and firms' reservation prices (i.e., willingness to pay and to sell at a given price) for *specific transactions* of delaying disclosure. In our view, this is a fundamental paradigm shift that could greatly improve the efficiency of blockholder disclosure and has profound implications for securities regulation as a whole. Once a

151. We discuss *infra* the question of who should conduct the sale of disclosure from within target firms and how to reduce agency costs.

mandatory disclosure rule is understood as a suboptimal replacement for investors' and firms' willingness to pay for disclosure, it is possible to reconceive of the entire issue as a bargaining problem. Delayed blockholder disclosure has value to hedge funds and firms, but the mandatory rule inhibits efficient trade. As the Coase theorem implies, the law should permit the two parties to bargain to the efficient solution.

However, the mandatory nature of the blockholder disclosure rule is not the *only* source of inefficiency. Unstructured negotiation between firms and hedge funds regarding the duration of the blockholder disclosure window would likely lead to severe transaction costs. And thus we come full circle to the problem with which we opened this Article: the Coasean ideal of bargaining to the efficient outcome applies only in a world of zero transaction costs. Negotiations between a hedge fund and a target firm are a classic case of bilateral monopoly: this is a thin market characterized by little-to-no competition between hedge funds and firms, such that both sides are "stuck with each other" and have an incentive to hold out for the best possible price.[152] Moreover, because hedge fund activism threatens management and the status quo, even simply receiving a signal that a hedge fund is accumulating a block of shares might lead management to take the preemptive action of announcing such information, thus undermining any chance of hedge fund activism succeeding.

Nonetheless, this shows that the fundamental challenge with implementing negotiated transactions for blockholder disclosure is not *whether* such an approach would be optimal but rather *how* to facilitate such negotiations effectively (i.e., with minimal transaction costs). In other words, this is a question of how to structure the rules of bargaining—a question that is suited quite nicely for mechanism design. Society would be better off if the parties could negotiate to an optimal outcome, but the nature of the bilateral monopoly context and the problems of incomplete information make unstructured bargaining impracticable. This is a situation calling for a bargaining mechanism.

We will show shortly how the Bonus and Two-Stage Procedures could be applied to blockholder disclosure to attenuate these transaction costs and facilitate efficient trade. But before doing so, it is necessary to distinguish between two different sources of potential transaction costs that would arise with a negotiated solution in this context: agency costs versus bargaining incentives. Up to this point,

152. *See* Rose, *supra* note 3, at 2183.

we have referred to the "firm" as an entity capable of conducting negotiations and maximizing "its" utility when bargaining with hedge funds. Of course, firms are artificial entitles created by the law. Regardless of the procedure utilized, negotiation must be actually performed by individuals acting on behalf of the firm. It is essential to consider potential agency costs that might arise when negotiating with hedge funds for delayed disclosure.

As we noted previously, management and its supporters on the board of directors have a strong incentive to preserve the status quo and oppose hedge fund activism even if the firm's shareholders would benefit from the intervention. This is the rationale behind a proposal to permit private ordering at the firm-specific level through an amendment to the corporate bylaws. Shareholders—particularly long-term shareholders who might enact such an amendment—have interests most closely aligned with those of the firm itself. The benefit to the firm from hedge fund activism is reducing managerial slack. It thus seems that giving management the authority to negotiate with the very entities tasked with disciplining and monitoring it might lead to significant agency costs.

For example, as we mentioned, management might simply undermine any attempt at hedge fund intervention by preemptively announcing to the market that a hedge fund is attempting to negotiate for delayed blockholder disclosure. Strictly speaking, this is not a transaction cost arising from unstructured negotiation but rather an *agency cost*: the firm might be better off if a hedge fund were to intervene, but management's own interests diverge with those of the firm. Similarly, management might hold out for a price far in excess of what the firm's existing shareholders would have accepted, were they so informed and able to make the decision instead.[153]

These agency costs may be ultimately diminished ex post through the instruments of corporate law. For example, shareholder litigation for violating the duty of loyalty might serve as an effective check against managerial self-interest in this context.[154] Upon discovering that management did not negotiate in good faith, either by disclosing the offer to the market or by insisting on

153. As noted *infra*, the "sale" of delayed disclosure brings revenue to the firm that would improve its financial position and thus offsets the potential cost of excessive hedge fund activism.

154. *See, e.g.*, Stone v. Ritter, 911 A.2d 362, 370 (Del. 2006) ("[A] director cannot act loyally towards the corporation unless she acts in the good faith belief that her actions are in the corporation's best interest.") (quoting Guttman v. Huang, 823 A.2d 492, 506 n.34 (Del. Ch. 2003)).

an unreasonably high price, an injured shareholder might claim that the transaction failed the entire fairness test, which would arguably apply because of the inherent conflict of interest in this context.[155]

Yet mechanism design can assist with reducing these agency costs as well. As noted *supra*, one of the most powerful benefits of applying mechanism design in a legal setting is the ability to convey valuable information to regulators and other interested parties.[156] Simply channeling negotiations through a structured procedure, operated by a regulator on a website, would permit tracking attempted offers and management's responses. These data could be disclosed to shareholders periodically, providing reliable evidence that would facilitate a shareholder's lawsuit in the event management did not conduct negotiations in good faith. Total disregard for shareholders' interests (e.g., by disclosing offers preemptively) could even be viewed as a form of market manipulation subject to civil and criminal penalties.[157] Indeed, the very threat of disclosure to the principal (i.e., shareholders) would likely compel the agent (i.e., management) to negotiate in the best interests of shareholders.

Mechanism design can make a greater contribution to reducing agency cost than simply facilitating ex-post enforcement. It is possible to apply the foregoing procedures to negotiations between management and shareholders such that the firm's ultimate position when negotiating with hedge funds is itself the product of a negotiated transaction. We return to this point when considering the operation of the procedures *infra*.

This example nicely demonstrates how mechanism design can align management's and shareholders' interests. Yet the most powerful aspect of mechanism design is its ability to structure bargaining incentives to induce the honest disclosure of reservation prices. Even if management's interests are perfectly aligned with those of shareholders, incentives remain in unstructured negotiation to bargain inefficiently by distorting offers and holding out for the best possible price.[158] By structuring the bargaining process, mechanism design can alter those incentives and make it in hedge funds' and firms' best interests to honestly disclose reservation prices and

155. *See* Weinberger v. UOP, Inc., 457 A.2d 701, 711 (Del. 1983).

156. *See* discussion *supra* Subsection II.C.2.

157. *See* discussion *infra* Subsection III.B for further examples of how information obtained through bargaining procedures can facilitate effective enforcement of the securities laws.

158. *See* discussion *supra* Subsection I.A.

thereby reduce the transaction costs inherent in unstructured negotiation.

2. Applying the Bonus and Two-Stage Procedures to Negotiating Delayed Blockholder Disclosure

We propose that Congress enact a statutory reform to blockholder disclosure that would facilitate bargained-for transactions by utilizing principles of mechanism design. In particular, the Bonus and Two-Stage Procedures each could facilitate negotiated delayed disclosure, albeit with different strengths and weaknesses. Under both approaches, negotiation ideally would be conducted through an automated website, operated by the SEC, that uses one of these procedures. We discuss the operation of each procedure, the choice between the procedures, and implementation issues of timing.

Before diving into the mechanics of the procedures, it is necessary to clarify what precisely would be negotiated. In the prior Section we referred generally to a price that hedge funds would pay for delayed disclosure. Of course, the parties are interested not only in the price but also the duration of the delay. In our view, it is sufficient to permit negotiation over a per-day price alone, thus giving hedge funds the freedom to delay disclosure for any duration provided that the appropriate price is paid. This proposal is not without challenges. In particular, firms might balk at the notion of giving hedge funds free rein to delay blockholder disclosure without any absolute durational limit. In our view, however, this concern is best addressed by setting the correct daily price for delayed disclosure. Hedge funds do not have unlimited resources. At the right price level, delayed disclosure beyond a certain duration will no longer be profitable, because the hedge fund cannot acquire a sufficient number of shares in that additional day to offset the marginal cost of one more day of delayed disclosure.[159]

We now consider the operation of the procedures. If the SEC were to apply the Bonus Procedure, it would begin by prompting

159. It is possible to extend our proposal to permit simultaneous negotiation over the duration and the price of disclosure, but the negotiation process becomes much more complicated. Such an approach is largely unnecessary, because the per-day price would be effective the vast majority of the time. If lawmakers are particularly concerned with the potential for stealth acquisitions for an exceedingly long period, an intermediate solution might involve permitting each firm to elect a "growth function" in its corporate bylaws whereby each additional day of disclosure results in an increased price based on some function, e.g., an exponential increase with a constant base set in the bylaws.

the firm and hedge fund to input their offers. This would not necessarily be synchronous—we discuss issues of timing *infra*. If the hedge fund's offer was equal to or exceeded that of the firm's, the SEC would pay a bonus equal to one-half of the difference of their offers and would record a delayed disclosure transaction at a price equal to the mean of their offers. If the hedge fund's offer was less than that of the firm, no transaction would be consummated. Recording a transaction would impose a legal obligation on the hedge fund to pay the firm for the period of time that elapses until the schedule 13D blockholder disclosure form is filed.[160]

As schedule 13D forms are filed with the SEC, tracking this duration should not be particularly difficult. The SEC could easily send an automated "bill" based on the duration that elapses between the delayed disclosure transaction and the schedule 13D filing and could collect payment from hedge funds to target firms. By conducting the process through the SEC's website, the hedge fund could remain anonymous to target firms until the moment of schedule 13D disclosure. This anonymity is essential for the transaction to function effectively. We return to this point when discussing the timing of the procedures *infra*.

If the SEC were to apply the Two-Stage Procedure, it would operate in a similar manner, albeit with slight differences. It would begin by prompting the firm and the hedge fund to name their reservation prices and offers. If their reservation prices do not overlap, the SEC would terminate the negotiation. If they overlap, the SEC would consider the relative position of the offers. If both offers are in the overlap interval of the reservation prices, a delayed disclosure transaction would be recorded at the mean of their offers. If only one of the two offers is in the overlap interval, the SEC would record a delayed disclosure transaction at that offer price with probability equal to $1/2$, using a random device.

Choosing between these procedures requires consideration of their respective strengths and weaknesses. As we showed in Part II, both of these procedures make honesty a weakly dominant strategy, reducing transaction costs resulting from exaggeration and puffery. The strengths of the Bonus Procedure are its simplicity and complete transactional efficiency. Every optimal transaction is implemented. The primary downside of the Bonus Procedure is that it requires paying a bonus, which imposes a fiscal burden on a bonus

160. This assumes that hedge funds initiate the process upon acquiring 5% beneficial ownership. We discuss issues of timing *infra*.

payer, and its vulnerability to collusion.[161] The fiscal impact may be ameliorated by the taxing procedure we presented in Subsection II.C.1 to render the application of the Bonus Procedure budget-balanced or by drawing the funds from the general treasury. The Two-Stage Procedure is less efficient because of exaggerated offers and the probabilistic implementation of a transaction, but it has the advantage of obtaining the honest disclosure of parties' reservation values even if one or both offers do not fall in the overlap interval. This can reduce agency costs and provide valuable information to securities regulators regarding the value of blockholder disclosure to firms and hedge funds.

Certainly, the greatest challenge with implementing these procedures in the context of blockholder disclosure is determining the precise timing and nature of the interaction with the SEC's website. Because it is essential not to "tip off" management that a specific hedge fund is accumulating shares, we suggest that negotiations be conducted anonymously, i.e., relayed through the SEC to the parties without disclosing their identity. It does not follow, however, that the process must be conducted synchronously or in "real time." Indeed, simply knowing that a hedge fund is interested in purchasing delayed disclosure at a given moment in time might be sufficient to permit the firm to undermine any attempted hedge fund activism.

To that end, we suggest that the SEC periodically request reservation prices and offers from target firms, perhaps when they submit their quarterly 10-Q filings. It is possible that a quarterly interval may be insufficient, i.e., if the firm's reservation prices and corresponding offers change within the span of three months. We see no reason why firms need be restricted to certain intervals when updating their reservation price and offers, provided that the mechanism is binding with respect to the firm's existing reservation price and offers at the time a hedge fund purchases delayed disclosure. The key requirement is that the firm's preferences be set ex ante.

Of course, unlike the public filings, these reservation prices and offers would be submitted privately to the SEC and remain confidential. A hedge fund interested in purchasing delayed disclosure would initiate the process on the SEC's website, and the SEC would

161. One of us (Brams) has argued that the United States, in effect, paid large bonuses in the form of financial aid, military equipment, and security guarantees to engineer the 1978 Camp David agreement between Egypt and Israel, which was formalized by Anwar Sadat and Menachem Begin's signing of a peace treaty in 1979 under the auspices of President Jimmy Carter. BRAMS, *supra* note 85, at 57–60.

simply utilize the data it had already obtained from the target firm to apply the procedures and determine whether to record a transaction. There is no need for target firms to know the identity of the hedge fund before setting their reservation prices and offers. Ideally, firms would supply these reservation prices and offers based on a thorough evaluation of their financial position and determination of the price the firm would be willing to receive in exchange for an additional day of delayed blockholder disclosure. This would permit management to consult with shareholders and arrive at this conclusion in a cooperative, reasoned fashion.

As noted previously, these procedures may also be utilized within the firm to resolve differing views between management and shareholders regarding the value of blockholder disclosure. For example, the Two-Stage Procedure could be used by management and shareholders to determine the firm's reservation price and offer *prior* to applying it in negotiations with the hedge fund.[162] The informational advantages of the Two-Stage Procedure would permit the SEC (again, likely through an automated algorithm) to evaluate whether management negotiated in good faith with the shareholders.

In addition to enforcement penalties, the absence of good faith on management's part might simply cause shareholders' position to represent that of the firm in negotiations with the hedge fund. Alternatively, the Bonus Procedure could be applied with a modification: the entirety of the bonus could be paid to management to agree on a price in its best interest. This could be viewed as a type of "severance pay," in recognition of the negative impact that the acquisition of shares would have on management (e.g., if they are laid off). We present these possibilities only as examples of ways in which regulators may respond to potential bad faith or to the inherent conflict of interest in negotiations with an activist shareholder seeking to oust existing management. The notion of reducing the discretion typically afforded to management in cases of potential conflicts of interest underlies the imposition of a higher standard in allegations of a violation of the duty of loyalty, where

162. The collective bargaining position of management and shareholders when multiple individuals are involved could be determined by having each choose its ideal price and amalgamating these prices according to their weights (e.g., shareholders' proportions of equity ownership). We propose that the final position of management and the shareholders be a *weighted median* of their positions—such that half the weight is on one side and half on the other—because the median, as opposed to the mean, is relatively invulnerable to manipulation.

the business-judgment presumption is replaced by the more exacting standard of fairness.[163]

The ability to conduct these negotiations in an asynchronous manner shows the power of mechanism design. An algorithmic procedure for negotiation enables this type of reasoned, careful evaluation of the value of hedge fund intervention to a firm in an atmosphere free of the tension and pressure that would arise in an unstructured negotiation. It is even conceivable that allowance would be made to permit boards (and shareholders) to set reservation prices for specific hedge funds in anticipation of the value that certain suitors might bring to the firm, or to specify a more nuanced pricing structure. We leave these further refinements of our proposal for another day. But we emphasize that unlike the present mandatory rule under section 13(d), mechanism design permits facilitating efficient transactions, i.e., those in which the value to hedge funds of an additional day of disclosure exceeds the value to the firm.

3. Regulating Social and Macroeconomic Effects

In addition to reducing the social externality of nonagreement, mechanism design permits addressing additional social and macroeconomic effects that arise specifically in the context of blockholder disclosure. As we noted previously, simply conducting negotiations through a website operated by a regulator could reduce agency costs by facilitating disclosure to shareholders in the event that management fails to conduct negotiations in good faith. Similarly, by retaining data regarding reservation prices and offers submitted through the website, regulatory agencies could detect outliers who might abuse the system rather than convey genuine information in good faith. A centralized clearinghouse for negotiations would decrease the costs of regulatory enforcement to ensure that these transactions are conducted appropriately.

Delayed blockholder disclosure imposes a cost on society because it permits trading on asymmetric information.[164] Mitts's proposal for a delayed disclosure fee fits nicely into the procedural framework we have articulated thus far. The SEC could simply add the fee into the negotiation process, increasing the charge to the hedge fund by a certain percentage to reflect the social cost of delayed disclosure at that price.

163. *See, e.g.*, Stone v. Ritter, 911 A.2d 362 (Del. 2006).

164. Mitts, *supra* note 141, at 227. R

The great advantage of a mechanism-design approach is that by obtaining honest disclosure of reservation prices (and offers if the Two-Stage Procedure is used), the SEC can obtain a great deal of information regarding the distribution of the costs and benefits of delayed blockholder disclosure. This information can facilitate a more efficient response to the social cost of trading on asymmetric information as a result of delayed disclosure.[165] In the case of the Bonus Procedure, the SEC could utilize a fee to break even when paying out bonuses for reaching agreements—being flexible about lowering or raising the fee as experience dictates—and closely monitoring the information it receives for signs of possible collusion.

In a similar manner, a mechanism-design approach would permit regulation of the macroeconomic effects of the microeconomic transactions occurring here to prevent excessive pooling or separating equilibria.[166] For example, there may be a higher social cost to clustering of delayed disclosure at certain durations. By examining the data of delayed disclosure transactions along with the duration that passes until the schedule 13D form is filed, the SEC's algorithm could deduce if a certain combination of reservation prices/offers are leading to excessive clustering at certain disclosure durations.[167] The algorithm could then discourage such clustering by imposing a very high tax on future agreements at that given distribution of reservation values or offers.

The more general point is that mechanism design permits nimble and adaptable regulation to specific transactions by channeling negotiations through algorithmic procedures. This permits more intelligent responses to the macroeconomic effects of microeconomic contracting terms. By inducing the honest disclosure of reservation values, regulators can employ algorithms that can respond more accurately to the parties' actual incentives and thereby shape socially desirable outcomes more effectively. Indeed, at the very least, regulators may utilize aggregate information regarding the distribution of reservation prices, offers, and transaction prices to determine whether securities regulation is facilitating

165. *See id.*

166. For a discussion of this problem in the context of low-equity mortgage lending, see generally Ayres & Mitts, *supra* note 44.

167. In order to determine whether these combinations of transaction terms would indeed lead to excessive clustering, the algorithm might consider additional data, such as the share price and the anticipated "pop" in value upon the schedule 13D disclosure. This could indicate the likely profit a hedge fund would receive. The cost of intervention would remain difficult to estimate, but it could likely be inferred indirectly from empirical data regarding the distribution of reservation values.

optimal intervention of prospective blockholders in target firms. More generally, a mechanism-design approach can be utilized in other contexts that presently employ mandatory disclosure to facilitate efficient transactions and informed regulatory policymaking.

CONCLUSION

In this Article, we have shown that mechanism design can reduce inefficient transaction costs arising from unstructured negotiations in bilateral monopoly with asymmetric information. Applying mechanism design to settlement negotiations has a straightforward rationale—failure to reach agreement because of strategic bargaining imposes a direct cost on society. It is a small step from mandatory mediation to structured negotiation procedures.

We believe that the example of blockholder disclosure demonstrates the power of viewing legal and regulatory regimes through the lens of procedural altering rules. There are many situations wherein the law presents an ultimatum: reach a deal through unstructured negotiation or accept a forced sale. For example, in corporate law, minority shareholders who oppose a merger are forced to sell their shares at the deal price. The only way to retain the contractual freedom to decide the price at which they will sell their shares is to convince the majority not to accept the deal. Even if the law gives minority shareholders appraisal rights, the choice again is between unstructured negotiation and a judicial determination of a "fair" sale price.

Mechanism design shows that this is a false dichotomy. Bargaining procedures can facilitate more efficient agreement between parties. Moreover, mechanism design can provide innovative solutions to the political aspects of corporate governance. Our suggestion to utilize a type of voting procedure for shareholder-management disputes could be extended to any type of decision facing a shareholder vote. One of us (Brams) has suggested voting procedures that might be used to give minority shareholders better representation—indirectly through majority positions they support, or directly by electing their own representatives—in matters pending a shareholder vote.[168]

168. *E.g.*, Steven J. Brams & Peter C. Fishburn, *Approval Voting*, 72 AM. POL. SCI. REV. 831 (1978). Additional information on different voting procedures, and comparisons among them, can be found in STEVEN J. BRAMS & PETER C. FISHBURN, APPROVAL VOTING (2d ed. 2007) and in STEVEN J. BRAMS, MATHEMATICS AND DEMOCRACY: DESIGNING BETTER VOTING AND FAIR-DIVISION PROCEDURES (2008).

Finally, the use of algorithmic procedures in mechanism design suggests new responses to the macroeconomic effects of microeconomic contracting. By directing contractual transactions through automated procedures, regulators could effectively shape the aggregate macroeconomic outcome of individual agreements. As we mentioned, this might take the form of "smart" taxation or licensing that responds in real-time to changing macroeconomic conditions (e.g., excessive leverage clustering in a particular region). More fundamentally, mechanism design allows viewing contract law as not merely setting the bounds of private agreements, but also regulating the structural conditions under which such agreement takes place.

AN APPEALING CHOICE: AN ANALYSIS OF AND A PROPOSAL FOR CERTIFICATES OF APPEALABILITY IN "PROCEDURAL" HABEAS APPEALS

*DAVID GOODWIN**

Introduction: AEDPA and Its Discontents ..... 792 R
I. Defining Terms: "Procedural" and "Merits" ..... 801 R
II. The Pre-AEDPA World and Appellate Review of Procedural Issues ..... 804 R
III. The Enactment of AEDPA and the New § 2253 ..... 807 R
IV. Post-AEDPA, Pre-*Slack*: The Struggle ..... 810 R
V. The Supreme Court Weighs in: *Slack* and Its Progeny ..... 813 R
A. *Slack* Itself ..... 813 R
B. Post-*Slack* Supreme Court Decisions ..... 816 R
VI. The Circuits Respond ..... 819 R
A. The First School: Permissive Scrutiny, Cognizability, and the "Quick Look" ..... 820 R
B. The Second School: The Strict Standard and Fealty to the Language of § 2253(c)(2) ..... 822 R
C. The *Pabon* Problem: The Third Circuit's Third Way ..... 825 R
VII. The Best Approach Is the Most Permissive ..... 829 R
VIII. What Does a "Compliant" COA Order Look Like in 2013? ..... 840 R
Conclusion: The Modest Proposal ..... 840 R

* J.D., New York University School of Law, 2010. I would like to thank the brilliant supervisors and attorneys at the Legal Division of the Third Circuit for their three years (!) of support and patience; Dr. Van C. Tran of Columbia University for his editorial assistance and encouragement; and the pleasant staff at the 4th Street Philadelphia Java Co./Milk & Honey, who never seemed to mind that I would spend entire weekends in their establishment. Of course, I would also like to extend my thanks to Alex Gorman, Ilyssa Coghlan, Julia Pilcer, and everyone else at the *N.Y.U. Annual Survey of American Law* who helped massage this piece into its final form. This Article and the views expressed herein reflect my own opinions, which should not be imputed to any entity, whether public or private, for whom I have worked.

INTRODUCTION: AEDPA AND ITS DISCONTENTS

One of the few universal truths in the law is this: the Antiterrorism and Effective Death Penalty Act of 1996 ("AEDPA")[1] is not well loved. As enacted, AEDPA departed from earlier efforts to reform the federal postconviction process[2] by implementing strict new procedural and substantive barriers to successful federal habeas corpus relief (as well as to relief under 28 U.S.C. § 2255, which is the primary pathway to collateral relief for federal prisoners).[3] AEDPA was intended, in part, to streamline and limit collateral attacks on con-

1. Pub. L. No. 104-132, 110 Stat. 1214 (1996).

2. Earlier efforts took a markedly different approach to solving the problem of protracted postconviction processes. For example, an AEDPA predecessor statute would have "eliminate[d] certain procedural barriers to consideration of the merits" because "[p]rocedural barriers have not only created new issues for litigation that cause unnecessary and artificial delays, but they have also barred the Federal courts from correcting constitutional errors." H.R. REP. NO. 103-470, at 3 (1994); *see also* Marianne L. Bell, Note, *The Option Not Taken: A Progressive Report on Chapter 154 of the Anti-Terrorism and Effective Death Penalty Act*, 9 CORNELL J.L. & PUB. POL'Y 607, 609–12 (2000) (discussing the history of the Powell Committee); David J. Garrow, *The Rehnquist Reins*, N.Y. TIMES MAGAZINE, Oct. 6, 1996, § 6, at 65.

3. In federal court, postconviction attacks on convictions and sentences, which generally occur after the direct appeal process has concluded, take two primary forms: petitions for a writ of habeas corpus under 28 U.S.C. § 2254 (filed by state prisoners) and motions to vacate a sentence under 28 U.S.C. § 2255 (filed by federal prisoners). The latter are not technically habeas corpus proceedings. *See, e.g.*, Swain v. Pressley, 430 U.S. 372, 377–78 (1977) (discussing briefly the distinction between § 2255 and habeas relief itself); Williams v. Carlson, No. 86-5503, 1987 U.S. App. LEXIS 14207, at *4–5 (D.C. Cir. Aug. 14, 1987) ("An attack on the imposition of sentence or the underlying conviction, as opposed to an attack on the execution of the sentence, is more properly within the ambit of a motion to vacate under 28 U.S.C. § 2255 which superseded habeas corpus, 28 U.S.C. § 2241, and provides nearly the exclusive remedy for such an attack by a federal prisoner."); 1 RANDY HERTZ & JAMES S. LIEBMAN, FEDERAL HABEAS CORPUS PRACTICE AND PROCEDURE § 41.1 (6th ed. 2012) [hereinafter FHCPP]; RONALD P. SOKOL, FEDERAL HABEAS CORPUS: A SECOND AND REVISED EDITION OF A HANDBOOK OF FEDERAL HABEAS CORPUS § 24 (1969) ("Section 2255 . . . is a statutorily compelled substitute for habeas corpus applicable to federal prisoners only."). It has been my experience that sticklers for accuracy in language dislike when a 28 U.S.C. § 2255 motion is referred to as a "petition for habeas corpus," while others fail to see the big deal with lumping everything under the umbrella of "petitions." *See, e.g.*, Matias v. Artuz, 8 F. App'x 9, 10 n.1 (2d Cir. 2001) (order denying certificate of appealability) ("[W]e use 'petition' to refer to the document seeking collateral relief, whether filed pursuant to 28 U.S.C. § 2254 or 28 U.S.C. § 2255."). While this Article does, on occasion, distinguish between habeas petitions and § 2255 motions in specific instances, I will risk the wrath of the majority by using "petition" and "petitioner" as shorthand in other contexts, and will invoke "habeas corpus" to signify the realm of postconviction remedies available to state and federal prisoners to challenge their convictions and sentences.

victions and sentences with the goals of "curb[ing] the abuse of the statutory writ of habeas corpus" and "address[ing] the acute problems of unnecessary delay and abuse in capital cases."[4] However, AEDPA succeeded instead in perplexing the judiciary, practitioners, and prisoners alike. Partly to blame was the statute's adherence to a labyrinthine process of review. But equally at fault, if not more so, was AEDPA's broad, ambiguous, and inconsistent language. Unlike many popular legislative bêtes noires, AEDPA draws condemnation on both its substance *and* the quality of its drafting, a sign of its quick construction and hasty passage.[5] The ultimate task of its interpretation has been placed back with the federal judiciary, the very entity that it was designed to constrain. And in defining AEDPA's contours through more than a decade of opinions, the judiciary has demonstrated inconsistent fealty to both the text of and the intent behind the statute.[6]

Warning signs were there from the beginning, to be sure. In his signing statement, President Clinton went so far as to rely on *Marbury v. Madison*,[7] assuring skeptical onlookers that "the courts, following their usual practice of construing ambiguous statutes to avoid constitutional problems,"[8] would make lemonade from lemons. But the verdict on AEDPA, fifteen or so years on, has not been kind. Courts are still struggling to make sense of the statute's most

4. H.R. REP. NO. 104-518, at 111 (1996) (Conf. Rep.); *see also* Hohn v. United States, 524 U.S. 236, 264–65 (1998) (Scalia, J., dissenting) ("The purpose of AEDPA is not obscure. It was to eliminate the interminable delays in the execution of state and federal criminal sentences, and the shameful overloading of our federal criminal justice system, produced by various aspects of this Court's habeas corpus jurisprudence."). An earlier committee report on a previous version of AEDPA announced similar goals. *See* H.R. REP. NO. 104-23, at 9–10 (1995). Of course, this path toward habeas reform drew strongly worded dissent as well. *See, e.g.*, *id.* at 34 ("[The bill] sacrifices the last hope of the falsely accused and the wrongly convicted—the Great Writ of Habeas Corpus—to a facile expediency driven by misguided passion for 'finality.'"); *see also* 141 CONG. REC. 15,018, 15,096 (1995) ("Our membership consists of both supporters and opponents of the death penalty, Republicans and Democrats, united in the belief that the federal habeas corpus process can be dramatically streamlined without jeopardizing its constitutional core.").

5. *See* Larry W. Yackle, *State Convicts and Federal Courts: Reopening the Habeas Corpus Debate*, 91 CORNELL L. REV. 541, 546 (2006) (discussing briefly the background of AEDPA's origins and passage).

6. *See, e.g.*, FHCPP, *supra* note 3, § 3.2 (describing the Supreme Court's give and take on AEDPA interpretation); *Supreme Court Upholds Limits on Appeals*, SEATTLE POST-INTELLIGENCER, June 29, 1996, at A8.

7. 5 U.S. (1 Cranch) 137 (1803).

8. Presidential Statement on Signing the Antiterrorism and Effective Death Penalty Act of 1996, 32 WEEKLY COMP. PRES. DOC. 719, 721 (Apr. 24, 1996).

fundamental provisions—for example, what "State . . . collateral review" means[9]—and a welter of onlookers has weighed in on a range of defects in the statute, including its poor drafting, its questionable underlying assumptions, and its efficacy at accomplishing its goals.[10]

Habeas corpus has long implicated doctrinally difficult areas, like retroactivity, and fascinating questions of federalism, which have simmered since the very day that the great writ "launched . . . on its brilliant career as a post-conviction remedy."[11] Understandably, a great quantity of post-AEDPA habeas scholarship has focused on the "big concepts" of the statute. Possibly the most debated and discussed portion of the statute is the amended 28 U.S.C.

9. *See, e.g.*, Wall v. Kholi, 131 S. Ct. 1278, 1287 (2011) (holding that "a motion to reduce sentence under Rhode Island law is an application for 'collateral review' that triggers AEDPA's tolling provision").

10. *See, e.g.*, Lindh v. Murphy, 521 U.S. 320, 336 (1997) ("[I]n a world of silk purses and pigs' ears, the Act is not a silk purse of the art of statutory drafting."); Bell, *supra* note 2, at 630 ("The AEDPA is redundant and Chapter 154 is ineffective, but more important, Chapter 154 is simply not well written."); Christopher Q. Cutler, *Friendly Habeas Reform—Reconsidering a District Court's Threshold Role in the Appellate Habeas Process*, 43 WILLAMETTE L. REV. 281, 282 (2007) ("The proverbial army of chimps pounding on typewriters could repeatedly recreate the AEDPA's shoddy language before reproducing even one melodious Shakespearean sonnet."); Andrew Hammel, *Diabolical Federalism: A Functional Critique and Proposed Reconstruction of Death Penalty Federal Habeas*, 39 AM. CRIM. L. REV. 1, 75 (2002) (describing AEDPA as a "new and poorly drafted statute that has required almost agonizingly intense interpretation by the judiciary"); Randal S. Jeffrey, *Successive Habeas Corpus Petitions and Section 2255 Motions After the Antiterrorism and Effective Death Penalty Act of 1996: Emerging Procedural and Substantive Issues*, 84 MARQ. L. REV. 43, 46–47 (2000); Emily Garcia Uhrig, *The Sacrifice of Unarmed Prisoners to Gladiators: The Post-AEDPA Access-to-the-Courts Demand for a Constitutional Right to Counsel in Federal Habeas Corpus*, 14 U. PA. J. CONST. L. 1219, 1228 (2012) (explaining that, in the wake of AEDPA, "[t]he resulting body of law is inordinately complex and vexing to even the most experienced of jurists"); Larry W. Yackle, *The Figure in the Carpet*, 78 TEX. L. REV. 1731, 1741 n.62 (2000) (lamenting AEDPA's "incomprehensible substantive standards"); *see also* John H. Blume, *AEDPA: The "Hype" and the "Bite"*, 91 CORNELL L. REV. 259, 263 (2006) (remarking that AEDPA's poor drafting significantly contributes to its not having had the effects its proponents envisioned). My favorite indictment of AEDPA would have to be Professor Anthony Amsterdam's foreword to the current edition of FHCPP, in which he describes AEDPA as an "atomic bomb" that "shatter[ed] the preexisting structure of habeas corpus law," with federal courts as the unfortunate toilers who are "still trying to reconstruct a bit of order" in the "blind killing ground" of post-AEDPA federal habeas jurisprudence. Anthony G. Amsterdam, *Foreword to the Sixth Edition* of RANDY HERTZ & JAMES S. LIEBMAN, FEDERAL HABEAS CORPUS PRACTICE AND PROCEDURE (2012). Unsurprisingly, in Professor Amsterdam's list of four primary AEDPA defects, its "elaborate and obscure" text occupies pole position. *Id.*

11. GRAHAM HUGHES, THE DECLINE OF HABEAS CORPUS 2 (1990).

§ 2254(d), which greatly enhanced the deference due by federal courts to state-court decisions. Despite arguably being the centerpiece of the post-AEDPA habeas regime, 28 U.S.C. § 2254(d) is oblique even in the context of AEDPA's general standard of legislative bunts.[12] The section's ambiguity, amplified by its daily relevance to habeas petitions arising out of state court convictions and sentences, has made it a justifiably popular topic for both scholars[13] and the Supreme Court, which continues to correct courts that fail to show the proper deference.[14]

This Article, by contrast, examines a more obscure part of AEDPA that is of equal daily importance to the federal courts: the certificate of appealability ("COA") requirement contained in 28 U.S.C. § 2253(c).[15] It may surprise the uninitiated to learn that habeas petitioners do not, strictly speaking, have an appeal of right in the federal courts, and must obtain COAs before a federal court

12. *Compare* 28 U.S.C. § 2254(d) (1994) (establishing, with multiple caveats, that "a determination after a hearing on the merits of a factual issue, made by a State court of competent jurisdiction . . . shall be presumed to be correct."), *with* 28 U.S.C. § 2254(d)(1)–(2) (2012) (restricting granting writ of habeas corpus to claims where the state court decision "resulted in a decision that was contrary to, or involved an unreasonable application of, clearly established Federal law, as determined by the Supreme Court of the United States" or "resulted in a decision that was based on an unreasonable determination of the facts in light of the evidence presented in the State court proceeding").

13. *See, e.g.*, Scott Dodson, *Habeas Review of Perfunctory State Court Decisions on the Merits*, 29 AM. J. CRIM. L. 223 (2002); Marshall J. Hartman & Jeanette Nyden, *Habeas Corpus and the New Federalism After the Anti-Terrorism and Effective Death Penalty Act of 1996*, 30 J. MARSHALL L. REV. 337 (1997); Allan Ides, *Habeas Standards of Review Under 28 U.S.C. § 2254(d)(1): A Commentary on Statutory Text and Supreme Court Precedent*, 60 WASH. & LEE L. REV. 677 (2003); Evan Tsen Lee, *Section 2254(d) of the New Habeas Statute: An (Opinionated) User's Manual*, 51 VAND. L. REV. 103 (1998).

14. These reversals often come in the form of *per curiam* opinions which draw the occasional dissent. *See, e.g.*, Coleman v. Johnson, 132 S. Ct. 2060 (2012) (per curiam); Wetzel v. Lambert, 132 S. Ct. 1195 (2012) (per curiam); Hardy v. Cross, 132 S. Ct. 490 (2011) (per curiam); Cavazos v. Smith, 132 S. Ct. 2 (2011) (per curiam).

15. During the pendency of this Article's creation, Jonah Horowitz published an excellent piece addressing another narrow question arising out of the COA requirement: "What happens when a federal district court judge rejects a federal magistrate judge's recommendation to grant habeas relief and then rules on a [COA]?" Jonah J. Horowitz, *Certifiable: Certificates of Appealability, Habeas Corpus, and the Perils of Self-Judging*, 17 ROGER WILLIAMS U. L. REV. 695, 698 (2012). Mr. Horowitz intended his article to be a "scholarly exploration of the lower courts' everyday work-product," *id.*, and I intend this Article to do the same.

of appeals may hear their appeals.[16] The process, stated simply, requires the district court to decide, in the first instance, whether it will grant a COA on any, a few, or all issues. Pursuant to a timely-filed notice of appeal and/or a separate application for a COA, the petitioner can request a COA or an expansion of the ambit of a COA from the court of appeals, which, in turn, is free to develop its own procedures for expanding, constricting, granting, and denying COA applications.[17] But the *substantive* requirements for obtaining a COA, which are set out by § 2253(c), are anything but clear, and courts have struggled to articulate what a prisoner must do to earn a COA. This Article explores one component of this test: to what extent federal appellate courts, in deciding whether to grant a COA,[18] should scrutinize the "merits" of a habeas corpus petition—the strength of a petitioner's substantive claims—when the district court rests its decision on a procedural ground that does not implicate the merits of the petition.

This issue, while undoubtedly specific, is neither as arcane nor as technical as one might assume and has split the circuits even after a Supreme Court case ostensibly settled the issue in 2000. The problem arises from the multi-stage, restrictive habeas appeals process. Before AEDPA, and since the beginning of the twentieth century, federal law required that habeas petitioners make a showing of merit—a demonstration of "probable cause"—before engaging in a full appeal. Supplanting the prior "probable cause" regime, AEDPA purported to adopt, but instead altered, a key component of the previous standard, replacing the word "federal" with the word "constitutional." This small change immediately caused a great deal of confusion over whether appeals of nonconstitutional matters, such as basic procedural questions, were still possible. Since AEDPA's slate of new procedural pitfalls—the one-year statute of limitations, for example—meant that more and more petitions would be denied on grounds other than on the merits, solving AEDPA's appellate riddle was a pressing task. Increasing the urgency of the affair

16. *See* Abney v. United States, 431 U.S. 651, 656 (1977) ("[T]here is no constitutional right to an appeal."); United States v. Brooks, 245 F.3d 291, 293 (3d Cir. 2001) (noting that certain provisions of 28 U.S.C. § 2253 denying a right to an appeal did not violate the Suspension Clause of the United States Constitution, U.S. CONST. art. I, § 9, cl. 2).

17. *See* FHCPP, *supra* note 3, § 35.4(b)(ii); *see also* Slack v. McDaniel, 529 U.S. 473, 483–85 (2000) (discussing procedures for treating a notice of appeal as a COA application).

18. Technically, courts grant an "application" for a COA, not a COA itself. For substantially the same reasons discussed above, *supra* note 3, I will employ the less-convoluted shorthand in this piece.

was the fact that appellate review would be necessary to clarify some of AEDPA's more opaque provisions.

Further complicating matters was the development and implementation of the actual COA-issuing procedure at the appellate level. The text of AEDPA suggested (but did not mandate) a two-step inquiry: First, does the petition or motion warrant granting a COA; and second, should the judgment of the district court be altered? As Federal Rule of Appellate Procedure 22(b) (as amended by AEDPA § 103) establishes, when a district judge has not issued a COA, an appellant may seek one from "a circuit judge or judges, as the court prescribes."[19] Rule 22(b) emphasizes the role a COA plays in the "triage" of habeas appeals, as does the fact that most merits cases in the circuit courts are decided by three-judge panels.[20] Because many litigants are pro se, and because full briefing is generally not required at this stage,[21] the COA requirement can easily become a gatekeeper of enhanced importance. For example, the Third Circuit appoints counsel as a matter of course to those pro se appellants who earn COAs.[22] Granting a COA, then, becomes a source of added expense and additional delay—an inefficient, and thus unappealing, proposition with regard to those appellants whose claims appear marginal at best.

Moreover, and in part because the COA stage precedes full appellate review, COA determinations are often brief orders or short, summary decisions. Many of these dispositions are not provided to legal publishers and, as a result, are not available on LexisNexis or Westlaw, are not easily searched (being available only on PACER), and may not become part of the relevant "case history" in online databases. Since COA orders are generally unreported, they rarely carry the weight of precedent, and provide few indications as to the extent of the court's reasoning, although subsequent merits decisions may shed light on the rationale underlying those orders or contain additional discussion of COA-related issues.[23] Indeed, as of 2013, most of the relevant federal circuit courts routinely issued or-

19. FED. R. APP. P. 22(b)(2).

20. Arising out of 28 U.S.C. § 46(c) (2013).

21. *See, e.g.*, 10TH CIR. R. 22.1 (2013) (requiring a brief from the appellant, but not the appellee, before a COA is granted).

22. *See* 3D CIR. I.O.P. 10.3.2 (2010) ("When a certificate of appealability is granted on behalf of an indigent appellant pursuant to 28 U.S.C. § 2254 or § 2255, the clerk appoints counsel for the appellant unless the court instructs otherwise."); *see also* Cutler, *supra* note 10, at 348 (articulating the "three tracks" upon which habeas appeals proceed).

23. *See, e.g.*, Darden v. Sobina, 477 F. App'x 192, 916–17 (3d Cir. 2012) (discussing attempt to expand COA in a later merits decision).

ders of varying length that were not indexed by legal databases and, with certain exceptions, contained little reasoning.[24]

24. This is readily apparent from a survey of the underlying district court decisions. *See, e.g.*, United States v. Pujols-Tineo, CR No. 05-137-04S, 2011 U.S. Dist. LEXIS 25112, at *19 (D.R.I. Mar. 11, 2011), *COA denied*, No. 11-2002 (1st Cir. Oct. 24, 2011); Platt v. Ercole, No. 05-CV-6050(JS), 2008 U.S. Dist. LEXIS 101368, at *12–13 (E.D.N.Y. Dec. 12, 2008), *COA denied*, No. 08-6263 (2d Cir. June 11, 2009); Rucker v. Curley, No. 11-82, 2011 U.S. Dist. LEXIS 97223, at *12 (W.D. Pa. Aug. 30, 2011), *COA denied*, C.A. No. 11-4617 (3d Cir. Mar. 9, 2012); Copeland v. Quarterman, CV-07-77, 2008 U.S. Dist. LEXIS 75219, at *11–13 (S.D. Tex. Sept. 28, 2008), *COA denied*, No. 08-41154 (5th Cir. July 30, 2009); Barnes v. Lafler, No. 2:07-CV-10782, 2010 U.S. Dist. LEXIS 39077, at *27–29 (E.D. Mich. Apr. 21, 2010), *COA denied sub nom.* Barnes v. Howes, No. 10-1613 (6th Cir. Oct. 13, 2010); Rodgers v. Pfister, No. 11-3120(SEM), 2011 U.S. Dist. LEXIS 94862, at *30–31 (C.D. Ill. Aug. 24, 2011), *COA denied*, No. 11-3145 (7th Cir. Mar. 29, 2012); Harris v. United States, No. 11-5081-CV-SW-RED, 2012 U.S. Dist. LEXIS 5792, at *6 (W.D. Mo. Jan. 18, 2012), *COA denied*, No. 12-1991 (8th Cir. June 7, 2012); Baldonado v. Ryan, No. CV-11-0090-PHX-FJM, 2011 U.S. Dist. LEXIS 146024, at *6–7 (D. Ariz. Dec. 16, 2011), *COA denied*, No 11-18081 (9th Cir. July 9, 2012); Reese v. United States, No. 6:11-cv-57, 2012 U.S. Dist. LEXIS 41364, at *1 (S.D. Ga. Mar. 26, 2012), *COA denied*, No. 12-11500-F (11th Cir. July 30, 2012). All of these decisions were procedural denials in the district court, and in each case a COA application was denied by the relevant court of appeals. Most of the orders were quite short; however, the Sixth Circuit and Eleventh Circuit issued lengthier memoranda that, nevertheless, were not available on either Lexis or Westlaw. The Seventh Circuit and Ninth Circuit utilized two-judge panels to issue their respective orders. *See* Santiago Salgado v. Garcia, 384 F.3d 769, 772–75 (9th Cir. 2004) (discussing differing composition of panels by circuit, while establishing the statutory basis for such a variance).

The Fifth Circuit makes some of its COA denials available to legal publishers. *See, e.g.*, Gates v. Thaler, 476 F. App'x 336 (5th Cir. 2012). Others are not available. It may be that COA orders deemed to be more "important" are submitted, whereas rote dispositions are not; or, alternatively, it simply may be the caprice of the various panels. The Sixth Circuit also makes some of its COA dispositions available. *See, e.g.*, Lathan v. Duffey, No. 10-3253, 2010 U.S. App. LEXIS 27415 (6th Cir. Nov. 3, 2010).

By contrast, the Fourth and Tenth Circuits make what appears to be the bulk of their COA decisions available to legal publishers. The Fourth Circuit's decisions are often rote and utilize standard language; in that sense, the decisions do not markedly differ from the summary orders of sister circuits. *See, e.g.*, Wilson v. United States, No. 5:09-CR-63-BR, 2011 U.S. Dist. LEXIS 79236, at *5 (E.D.N.C. July 20, 2011), *COA denied per curiam*, 459 F. App'x 240, 241 (4th Cir. 2011). However, the Tenth Circuit issues full opinions to dispose of its COA applications, and often these read like complete decisions on the merits. *See, e.g.*, Warren v. Milyard, No. 10-cv-02557-BNB, 2011 U.S. Dist. LEXIS 17986, at *9–10 (D. Colo. Feb. 9, 2011), *COA denied*, 427 F. App'x 670, 673 (10th Cir. 2011). As discussed further *infra*, the Tenth Circuit's approach can, at times, appear to contravene the Supreme Court's insistence that the COA stage is not the proper place to evaluate the full spectrum of the petitioner's or movant's arguments. But even in circumstances where the Tenth Circuit grants a COA, a separate stage (requiring additional briefing, *see supra* p. 797) precedes a full merits determination. *See, e.g.*, United States v.

These factors—triage and unavailability—collide when appellants seek review of procedural questions. In the aforementioned case from 2000, *Slack v. McDaniel*,[25] the Supreme Court clarified that federal appellate courts could review the procedural decisions of the district courts, even if those decisions themselves were not of "constitutional" dimension.[26] While *Slack* settled the broader question raised by the passage of AEDPA, it did little to solve an equally thorny concern: What attention should be paid to the merits of a petition when the district court resolved the matter on a procedural ground? *Slack* says that jurists of reason must, in part, find it debatable whether a petition "states a valid claim of the denial of a constitutional right,"[27] but this pronouncement does not explicitly track the language of 28 U.S.C. § 2253(c) and lacks an obvious antecedent in the Supreme Court's habeas corpus jurisprudence. Both before and after *Slack*, courts have struggled to set a workable standard for evaluating the merits of these petitions. But despite the regularity of procedural COA grants and denials, vanishingly little law has been made on the question.

In this Article, I analyze the various standards used by the federal appellate courts and demonstrate that an undemanding standard is ideal. My thesis is this: petitioners who comply with the procedural requirements of AEDPA, or whose petitions are otherwise unfairly dismissed for other procedural reasons, should not be denied their "one bite of the [postconviction] apple," as provided by AEDPA,[28] and should be afforded a full review of their claims on the merits. As I will demonstrate, the wide reach of § 2253(c), combined with the innumerable ways that a petition can fail other than

Roberts, 492 F. App'x 869, 873–74 (10th Cir. 2012) (granting COA and directing additional briefing); United States v. Roberts, 500 F. App'x 757, 759 (10th Cir. 2012) (per curiam) (issuing limited remand to the district court). Curiously, the odd order by the "nonpublisher" courts does end up on Lexis or West, although I imagine this to be due to operator error or the eccentricity of individual chambers' procedures. For example, *Wanger v. Hayman*, No. 11-1375, 2011 U.S. App. LEXIS 18466 (3d Cir. May 26, 2011), is a rote denial of a COA, but the Lexis reproduction of the order also contains the Third Circuit's boilerplate clerk order reflecting an entry of judgment, which surely was not intended for even summary publication. And, of course, courts may choose to publish the occasional COA determination, even if doing so is not the regular practice of the court. *See, e.g.*, Jefferson v. Welborn, 222 F.3d 286 (7th Cir. 2000).

25. 529 U.S. 473 (2000).

26. *Id.* at 484.

27. *Id.*

28. *See In re* Davis, 565 F.3d 810, 817–18 (11th Cir. 2009) (collecting congressional statements about the importance of a single round of federal habeas review); FHCPP, *supra* note 3, § 3.2.

on the merits, suggests that a more searching singular standard is simply unworkable, especially when the background of the case has not been developed and a record is not available. Adopting a single, lenient standard also provides numerous direct and indirect benefits, such as increased transparency and hastening of the development of case law. Even though more COAs would be issued, this need not have a deleterious effect on a court's ability to conserve resources. While AEDPA interposes the COA stage between initial appeal and a final merits determination, it does not affect an appellate court's power to take summary action without briefing or to affirm on an alternative basis not relied upon by the district court.[29]

Because COA grants on procedural grounds also implicate the actual *form* of a COA order itself, this Article also briefly discusses unresolved formal questions stemming from § 2253(c). Remarkably, what constitutes a fully § 2253(c)-compliant COA order is not yet a closed question in 2013. While a recent Supreme Court case, *Gonzalez v. Thaler*,[30] may have muddied the issue further, it also announced that certain defects would not deprive a court of appeals of jurisdiction[31]—and that, in any case, a court is free to revise a COA at any time in order to correct a defect. So as to forestall confusion, however, I argue that an order granting a COA on a procedural ground should indicate one or more specific claims in the underlying petition that satisfy subsection § 2253 (c)(2)'s "substantial claim" requirement. While not necessary as a matter of jurisdiction post-*Thaler*, such a practice is still "mandatory" under the statutory language of § 2253.

This Article is structured as follows. Part I establishes a working definition of "procedural," positing three "soft categories" that are useful for thinking about various kinds of procedural dismissals—

29. *See, e.g.*, Hemphill v. Hudson, 483 F. App'x 118, 120 (6th Cir. 2012) (per curiam) ("Because we can dispose of Hemphill's claims on the merits, however, we decline to address the district court's determination that Hemphill's claims are procedurally defaulted."). Indeed, as I argue further herein, the biggest problem with thinking of a COA as a strict gatekeeper is that it has the potential to add a needless stage to the quick reversal of simple mistakes by a lower court, such as computational errors. *See also* Barefoot v. Estelle, 463 U.S. 880, 894 (1983) ("[A] court of appeals may adopt expedited procedures in resolving the merits of habeas appeals, notwithstanding the issuance of a certificate of probable cause."); Garrison v. Patterson, 391 U.S. 464, 466 (1968) (per curiam) ("[W]here an appeal possesses sufficient merit to warrant a certificate, the appellant must be afforded adequate opportunity to address the merits, and that if a summary procedure is adopted the appellant must be informed, by rule or otherwise, that his opportunity will or may be limited.").

30. 132 S. Ct. 641 (2012).

31. *Id.* at 646.

both those unique to habeas and those common to other kinds of civil litigation. In Part II, I recite a brief history of appealability procedures pre-AEDPA in order to provide a basis for understanding the departure of (and confusion caused by) AEDPA's revisions. Part III looks at the statutory revisions to § 2253 and related provisions made by AEDPA, and discusses them from a "first impressions" perspective. Part IV examines the post-AEDPA, pre-*Slack* period, when the circuit courts attempted to square the new § 2253 with previous approaches to appealability, creating various circuit-specific tests. Part V tackles *Slack*, the primary Supreme Court decision on point, and also discusses subsequent decisions—most notably, *Miller-El v. Cockrell*[32]—that have addressed COAs. In Part VI, I show how *Slack* both replaced the individual tests marshaled by the circuit courts and caused new fractures, with courts (by this point well versed in parsing AEDPA's text) adopting markedly different interpretations of *Slack*. Part VII argues that the most lenient of these tests best pays fealty to *Slack* itself, as well as positing numerous legal, procedural, and prudential reasons for adopting a universal standard. In Part VIII, I describe what a "compliant" COA order should look like after *Thaler*.

So as to narrow the arena of discussion, I have generally restricted my analysis to the perspective of a court of appeals, because this Article contains greater implication for appellate courts than for trial courts. That is not to suggest that many of these questions do not also apply to a district court, which—in adjudicating close procedural issues—may be called upon to determine whether the merits of an uncertain or marginal petition are sufficient to warrant a COA grant on a complex (or simply opaque) procedural question.

I.
DEFINING TERMS: "PROCEDURAL" AND "MERITS"

Before starting a discussion on habeas corpus and procedural issues, it is helpful to first pin down a working definition of "procedural," as that term is extraordinarily broad and inconsistently used. Because this Article discusses "procedural" in the context of federal habeas corpus, the situations immediately evoked are those indelibly associated with habeas—for example, procedural default or AEDPA-related timeliness or successiveness concerns.[33] But fed-

32. 537 U.S. 322 (2003).

33. For example, a determination that a petition is time-barred under the AEDPA one-year statute of limitations is a procedural outcome for the purposes of

eral habeas cases are conducted like ordinary civil cases in many, but not all, respects.[34] Thus, they are subject to more "mundane" procedural devices as well, such as dismissals for failure to prosecute.[35]

Perhaps the easiest way to think about the term "procedural" is to contrast it with its counterpart: the "merits" of the claims contained in a habeas petition. Take, for example, a habeas petition in which a state prisoner is alleging ineffective assistance of trial counsel (a commonly raised collateral claim). To resolve the claim on the merits, the district court would determine whether trial counsel's performance passed the test established in *Strickland v. Washington*[36] or, in the context of a post-AEDPA petition, whether the state court resolved the claim within the wide range considered acceptable under the statute.[37] A disposition on procedural grounds, on the other hand, would not necessarily lead to a resolution of the substantive content of the petitioner's constitutional claim in federal court, but would instead look to whether the petition was timely filed, whether the petitioner properly exhausted his state-

granting a COA. *See, e.g.*, Holmes v. Spencer, 685 F.3d 51, 57–58 & n.3 (1st Cir. 2012) (applying procedural standard in COA deriving from timeliness issue). But these same time-bar issues render the previous petition disposed of "on the merits" if the petitioner later tries to file a second or successive action. *See, e.g.*, *In re* Rains, 659 F.3d 1274, 1275 (10th Cir. 2011) (per curiam); McNabb v. Yates, 576 F.3d 1028, 1029 (9th Cir. 2009) (per curiam); *see also* Quezada v. Smith, 624 F.3d 514, 518–19 (2d Cir. 2010) (discussing possible exceptions to the rule); Carter v. United States, 150 F.3d 202, 205 (2d Cir. 1998) ("[T]he denial of a first § 2254 petition for procedural default, which default is not overcome by a showing of cause and prejudice, 'must be regarded as a determination on the merits in examining whether a subsequent petition is successive.'").

34. *See* Day v. McDonough, 547 U.S. 198, 212 (2006) (Scalia, J., dissenting) (observing that the Federal Rules of Civil Procedure govern habeas cases unless they are inconsistent with the Habeas Corpus Rules); United States v. Fiorelli, 337 F.3d 282, 285–86 (3d Cir. 2003) (discussing the applicability of the Federal Rules of Civil Procedure to 28 U.S.C. § 2255 motions); *see also* FHCPP, *supra* note 3, § 2.2 ("For example, although the 'custody' prerequisite for habeas corpus jurisdiction usually limits review to the legality of criminal prosecutions and sentences, the Court typically describes the writ as a 'civil' remedy—one that, indeed, is partly governed by the Federal Rules of *Civil* Procedure.") (footnotes omitted); SOKOL, *supra* note 3, at § 18.

35. *See, e.g.*, United States v. Hyatt, No. 2:05-cr-216, 2012 U.S. Dist. LEXIS 105310, at *1–2 (E.D. Cal. July 27, 2012); *see also* Jeffrey, *supra* note 10, at 74–91 (discussing various kinds of procedural dismissals, ranging from the general to the habeas-specific, in the context of determining whether a later petition may be counted as "second or successive").

36. 466 U.S. 668, 687 (1984).

37. *See, e.g.*, Patel v. Dormire, 609 F. Supp. 2d 884, 887–90 (E.D. Mo. 2009).

court remedies before commencing a federal action, or whether it adhered to other filing and pleading requirements.[38]

In this Article, I will define as "procedural" all of those devices that lead to a case-dispositive order on grounds other than the merits. They can be separated into three broad, porous categories. In the first category are those procedural grounds that grow out of standard civil practice, and have nothing to do with habeas petitions *per se*. The aforementioned dismissal on the basis of failure to prosecute is one, but so would be failure to comply with *in forma pauperis* requirements, not following court orders, and so on. Many of these dismissals may also be ostensibly without prejudice,[39] although residual concerns might limit a petitioner's ability to correct certain errors in his filings. In the second category are grounds specific to habeas but not directly derived from AEDPA, such as dismissals for failure to exhaust state court remedies or dismissals based on the "in custody" requirement of the federal habeas statute.[40] In the third category are those dismissals based on the explicit statutory text of AEDPA, such as for failure to file within the one-year statute of limitations period. Each category involves a "procedural" dismissal that, despite being contained under that broad category, can involve markedly different scrutiny of the merits of a petition; indeed, for some prefiling dismissals, a petition may not even have been properly lodged in the district court.

A few words of caution: these categories are but a fiction, and actions can and do overlap. The categories are mostly a useful device for thinking about the development of the merits at the stage of dismissal. Furthermore, some "procedural" actions blur the lines between an evaluation of the merits and a procedural termination of a petition.[41]

38. *See, e.g.*, Rouse v. Iowa, 110 F. Supp. 2d 1117 (N.D. Iowa 2000).

39. *See, e.g.*, Patrick v. Erie Cnty. Court of Common Pleas, No. 12-156, 2012 U.S. Dist. LEXIS 142389, at *1–2 (W.D. Pa. Aug. 30, 2012), *report and recommendation adopted*, 2012 U.S. Dist. LEXIS 142390 (W.D. Pa. Oct. 2, 2012).

40. *See, e.g.*, Maleng v. Cook, 490 U.S. 488, 490–94 (1989) (per curiam).

41. In a common example, an analysis of whether to enforce a waiver of appellate or collateral-attack rights often includes a discussion of the merits of the underlying claims, because the test for enforcing the waiver considers whether limiting postconviction rights would work a "miscarriage of justice." *See, e.g.*, Sotirion v. United States, 617 F.3d 27, 36–39 (1st Cir. 2010).

II.
THE PRE-AEDPA WORLD AND APPELLATE REVIEW OF PROCEDURAL ISSUES

Any discussion of the changes wrought by AEDPA must begin with a quick overview of the law prior to its passage.[42] Appeals as of right in habeas proceedings arising out of state court detentions were restricted as early as the first years of the twentieth century, when Congress passed "An Act Restricting in certain cases the right of appeal to the Supreme Court in habeas corpus proceedings."[43] These restrictions did not apply to collateral review of federal convictions. Subsequent amendments to the law focused on the circuit courts of appeal, which were granted jurisdiction over habeas petitions in 1925.[44] The familiar "2253" designation came into being during the 1948 codification of the Judicial Code,[45] and, outside of a few mid-century modifications, the statute remained largely unchanged through 1996. The relevant pre-AEDPA text of § 2253, in its final amended form, was as follows:

> An appeal may not be taken to the court of appeals from the final order in a habeas corpus proceeding where the detention complained of arises out of process issued by a State court, unless the justice or judge who rendered the order or a circuit justice or judge issues a certificate of probable cause.[46]

As before, § 2253 applied to collateral attacks on state convictions only.[47] And nowhere did the text of this new section provide for an explicit limitation on the issues upon which appeal could be

42. For an in-depth look at the evolution of habeas appellate procedure, see Cutler, *supra* note 10, at 285–309.

43. Act of Mar. 10, 1908, ch. 76, 35 Stat. 40 (1908); *see also* Barefoot v. Estelle, 463 U.S. 880, 892 n.3 (1983) ("In 1908, concerned with the increasing number of frivolous habeas corpus petitions challenging capital sentences which delayed execution pending completion of the appellate process, Congress inserted the requirement that a prisoner first obtain a certificate of probable cause to appeal before being entitled to do so.").

44. *See* Ira Robbins, *The Habeas Corpus Certificate of Probable Cause*, 44 OHIO ST. L.J. 307, 313 & nn.36, 39 (1983).

45. *See* Act of June 25, 1948, ch. 646, 62 Stat. 967 (1948) (revising, codifying, and enacting into law title 28 of the U.S. Code entitled "Judicial Code and Judiciary").

46. 28 U.S.C. § 2253 (1994).

47. *See* Feldman v. Perrill, 902 F.2d 1445, 1449 n.2 (9th Cir. 1990) ("Certificates of probable cause are required only where the detention complained of arises out of process issued by a State court.") (internal quotation marks omitted).

taken if a certificate was granted, implying that no restriction was warranted.[48]

Jurisprudence on the showing required for a certificate of probable cause ("CPC") was itself slow to develop in the circuits. As late as 1979, the Second Circuit observed that it had "not directly passed on the question" and noted a dearth of "Supreme Court precedent clearly on point."[49] At the very least, however, CPCs granted on procedural questions were seen to be implicitly within the ambit of the statute.[50]

The Supreme Court addressed the CPC requirement head-on in a landmark opinion, *Barefoot v. Estelle*.[51] In *Barefoot*, the Court squarely aligned itself with "the weight of opinion in the Courts of Appeals that a certificate of probable cause requires petitioner to make a 'substantial showing of the denial of [a] federal right.' "[52] It also required "more than the absence of frivolity," which was a more exacting standard than one premised on mere "good faith."[53] Recognizing, too, that a "substantial showing of the denial of a federal right" might be otherwise hard to qualify, the Court offered up a now-familiar standard: a petitioner must simply show that the issues he raises "are debatable among jurists of reason; that a court could resolve the issues [in a different manner]; or that the questions are 'adequate to deserve encouragement to proceed further.' "[54] While *Barefoot* was decided in the context of a capital case, and habeas procedures in capital and noncapital cases are not al-

48. *See* Sherman v. Scott, 62 F.3d 136, 138–39 (5th Cir. 1995). As of 1995, the circuits were split on this question, with the Second Circuit departing from its brethren in allowing a court to limit the issues to be heard on appeal, with the proviso that (in accordance with later COA practices) a court could always expand the issues to be decided after the fact. *See id.* at 138 n.1. By contrast, other circuits viewed the certificate to apply to the entire appeal. *See* Tompkins v. Moore, 193 F.3d 1327, 1330 (11th Cir. 1999).

49. Alexander v. Harris, 595 F.2d 87, 89–90 (2d Cir. 1979) (per curiam); *see also id.* at 90–91 (observing that "federal courts have variously articulated the standard for issuance of a certificate of probable cause to appeal," citing several cases, and then holding that "the standard of probable cause to appeal requires the district court to find that the petition is not frivolous and that it presents some question deserving appellate review"); Horowitz, *supra* note 15, at 701–03 (detailing history of pre-AEDPA appealability requirements).

50. *See, e.g.*, Piercy v. Parratt, 579 F.2d 470, 471 (8th Cir. 1978) (granting CPC on failure to exhaust state remedies).

51. 463 U.S. 880 (1983).

52. *Id.* at 893 (alteration in original).

53. *Id.* (citations omitted).

54. *Id.* at 893 n.4 (alteration in original) (quoting Gordon v. Willis, 516 F. Supp. 911, 913 (N.D. Ga. 1980)).

ways perfectly alike,[55] the case nevertheless became an important standard-bearer for determining whether an appeal was warranted in habeas cases.[56]

Because the *Barefoot* standard requires only that prisoners demonstrate a denial of a federal right, review of procedural questions was largely uncontroversial. For example, courts readily exercised de novo review of exhaustion questions without otherwise addressing the merits of a petition. In *Story v. Kindt*,[57] the Third Circuit reversed a determination that state court remedies had not been exhausted and remanded for consideration by the district court of the petition's merits in the first instance.[58] In a nonprecedential opinion, the Ninth Circuit approved of a certificate of probable cause that explicitly disclaimed an analysis of the merits of the underlying habeas petition.[59] Much of the Supreme Court's "category 2" procedural habeas doctrine was established during this time, including landmark cases on procedural default,[60] exhaustion,[61] and abuse of the writ.[62]

55. *See, e.g.*, Rowell v. Dretke, 398 F.3d 370, 373 (5th Cir. 2005) ("In death penalty cases, doubts on whether a COA should issue are resolved in the petitioner's favor.").

56. *See, e.g.*, Lozada v. Deeds, 498 U.S. 430, 431–32 (1991) (applying *Barefoot* in the context of a state controlled-substances conviction); Hill v. Beyer, 62 F.3d 474, 480 n.5 (3d Cir. 1995); Flieger v. Delo, 16 F.3d 878, 882 (8th Cir. 1994); Cuppett v. Duckworth, 8 F.3d 1132, 1149 (7th Cir. 1993) (Ripple, J., concurring) (observing a "growing tendency in some of the district courts of this circuit to measure [certificate of probable cause] applications, either explicitly or implicitly, by an inappropriately high standard"); Gee v. Shillinger, 979 F.2d 176, 178 (10th Cir. 1992). Of course, because orders granting or denying CPCs were also generally not sent to legal publishers, it is difficult to gauge the exact rationale relied upon by the appellate courts in the interim, especially as many orders did not otherwise specify why they were granting a certificate. *See, e.g.*, Zilich v. Reid, Order Granting Request for Certificate of Probable Cause, C.A. No. 93-3459 (3d. Cir. Feb. 17, 1994).

57. 26 F.3d 402 (3d. Cir. 1994).

58. *See id.* at 407. The dissenting judge would have considered the merits of the petition on appeal in the first instance, although he otherwise joined the exhaustion finding of the majority. *See id.* at 407–12 (Cowen, J., dissenting).

59. Nilsen v. Borg, No. 94-15145, 1994 U.S. App. LEXIS 32933, at *1 n.1 (9th Cir. Nov. 18, 1994).

60. *See* Coleman v. Thompson, 501 U.S. 722, 750 (1991).

61. *See* Rose v. Lundy, 455 U.S. 509, 522 (1982); *see also* Castille v. Peoples, 489 U.S. 346, 351 (1989).

62. *See* McCleskey v. Zant, 499 U.S. 467, 490 (1991).

III.
THE ENACTMENT OF AEDPA AND THE NEW § 2253

In 1996, § 102 of AEDPA substantially amended 28 U.S.C. § 2253. The relevant part of the reworked statute, subsection (c), states the following:

> (1) Unless a circuit justice or judge issues a certificate of appealability, an appeal may not be taken to the court of appeals from—
>
> (A) the final order in a habeas corpus proceeding in which the detention complained of arises out of process issued by a State court; or
>
> (B) the final order in a proceeding under section 2255.
>
> (2) A certificate of appealability may issue under paragraph (1) only if the applicant has made a substantial showing of the denial of a constitutional right.
>
> (3) The certificate of appealability under paragraph (1) shall indicate which specific issue or issues satisfy the showing required by paragraph (2).[63]

Before examining the judicial construction of the new statute, it is helpful to take a step back and take in these substantial changes from an analysis of the statutory text itself.

The most obvious change was one of nomenclature: the revised § 2253(c) did away with the old "certificate of probable cause" designation, ushering in the era of the "certificate of appealability."[64] For those petitions filed after AEDPA's effective date, the COA requirement would supplant the CPC requirement.[65]

The COA's innovations, however, were more than semantic. The new § 2253 departed from its predecessor by extending its filtering mechanism to federal prisoners, eliminating their appeals as of right. As stated explicitly in § 2253(c)(1)(B), federal prisoners would now need to obtain a COA to appeal the denial of their 28

63. 28 U.S.C. § 2253(c) (2012). Section 2253 has not been amended since 1996.

64. This was a newish coinage. When the phrase appears in the federal judicial lexicon prior to AEDPA, it generally refers to interlocutory certification of appeals under 28 U.S.C. § 1292(b). *See, e.g.*, Sledge v. J. P. Stevens & Co., 585 F.2d 625, 633 n.14 (4th Cir. 1978); Rollins v. United States, 286 F.2d 761, 762 (9th Cir. 1961).

65. *See* United States v. Kunzman, 125 F.3d 1363, 1364 n.2 (10th Cir. 1997) (discussing the circuit split regarding petitions to which the COA requirement should apply).

U.S.C. § 2255 motions to vacate.[66] Hence, the new COA would be applied to a broad class of litigants that was previously exempt from the strictures of its predecessor.[67]

The new § 2253 also appeared to discard the "non-specification" regime of the CPC, and instead imposed something akin to the "questions presented" model of the Supreme Court: COAs would only be granted on certain issues; moreover, briefing and appellate decision making would then be confined to those issues, unless the ambit of the appeal was expanded by a court (either on motion or sua sponte).[68] Even here, however, the statute is maddeningly opaque on the specifics. For one, are the issues "indicated" in subsection (3) the same issues upon which a COA is to be

66. *Compare* United States v. Perez, 129 F.3d 255, 259 (2d Cir. 1997), Herrera v. United States, 96 F.3d 1010, 1011 (7th Cir. 1996), Thye v. United States, 96 F.3d 635, 636 (2d Cir. 1996) (per curiam), FHCPP, *supra* note 3, § 35.4, Brent E. Newton, *Applications for Certificates of Appealability and the Supreme Court's "Obligatory" Jurisdiction*, 5 J. App. Prac. & Process 177, 180 (2003), *and* Brian Serr, *Criminal Procedure*, 29 Tex. Tech L. Rev. 547, 578 (1998) (discussing recent Fifth Circuit developments in habeas procedure, and noting that "prior to the AEDPA's amendment of federal procedures for collaterally attacking a conviction, section 2255 petitioners possessed an automatic right to appeal"), *with* Payne v. United States, 539 F.2d 443, 445 (5th Cir. 1976), Fisher v. United States, 317 F.2d 352, 354 n.3 (4th Cir. 1963), *and In re* Marmol, 221 F.2d 565, 566 (9th Cir. 1955) (per curiam) ("Movant is a federal prisoner and a certificate of probable cause to appeal is unnecessary.").

67. Federal prisoners or detainees who brought proper challenges to their confinement under 28 U.S.C. § 2241 were still generally exempt from the COA requirement, however, as long as their detention was not the product of state court action. *See* Alaimalo v. United States, 645 F.3d 1042, 1047 (9th Cir. 2011); Padilla v. United States, 416 F.3d 424, 425 (5th Cir. 2005) (per curiam).

68. *See* Holmes v. Spencer, 685 F.3d 51, 58 (1st Cir. 2012); FHCPP, *supra* note 3, § 35.4(b)(i) & n.48; *see also Herrera*, 96 F.3d at 1012 ("The two certificates differ only in scope: a certificate of probable cause places the case before the court of appeals, but a certificate of appealability must identify each issue meeting the 'substantial showing' standard, see the amended § 2253(c)(3)."). AEDPA also amended Fed. R. App. P. 22, but much of the work at making coherent procedure out of AEDPA was the responsibility of the courts. *See, e.g.*, Harkins v. Roberts, 935 F. Supp. 871, 872 (S.D. Miss. 1996) ("[T]his Court will proceed under the assumption that it has the ability to grant a certificate of appealability."). In other words, the idea of the COA as a document of limitation has never quite come to pass. *See, e.g.*, United States v. Shipp, 589 F.3d 1084, 1087–88 (10th Cir. 2009) (collecting cases and local-rule citations about COA expansions); *see also* Coady v. Vaughn, 251 F.3d 480, 486 (3d Cir. 2001) ("Our conclusion that a certificate of appealability is required for this appeal to go forward does not compel dismissal. Because Coady filed a timely notice of appeal, we construe this notice as a request for a certificate of appealability . . . .").

granted?[69] Moreover, since the "certificate" itself is not a distinct, physical document, was any process required to amend it, and could it be amended by implication?[70] And was the COA limitation requirement intended to impose restrictions on an appellate court's review, or simply to function as triage to be disregarded if necessary?[71] In AEDPA's wake, courts would be left to determine the answers to these questions.

By far AEDPA's most notable alteration, however, was contained in some linguistic sleight of hand. In *Barefoot*, the Supreme Court had spoken of a petitioner's obligation to show a "substantial denial of a *federal* right" in order to obtain a COA. Suddenly, however, the new § 2253 required that a substantial showing be made of the denial of a *constitutional* right.[72] Given that what limited legisla-

69. Generally, no, which is intriguing given this part of the statute's ostensible origination of the requirement to specify the issues certified for appeal. *See, e.g.*, Satizabal v. Folino, 318 F. App'x 78, 81 (3d Cir. 2009) ("But if [the district court] issues a certificate of appealability, the Court should indicate the specific issue or issues on which Satizabal has made his substantive showing even if it issues the certificate of appealability solely on the procedural question involving equitable tolling."); Richardson v. Greene, 497 F.3d 212, 217 (2d Cir. 2007) ("Here, it is noteworthy that the district court granted the certificate of appealability on the constitutional merits issue, but not the procedural ground on which it based its decision . . . .").

70. Generally, yes. *See, e.g.*, Post v. Bradshaw, 621 F.3d 406, 413 (6th Cir. 2010) (noting that the court, by setting a briefing schedule including additional issues, "impliedly" expanded the COA to cover those issues); *see also* Thomas v. Crosby, 371 F.3d 782, 802 (11th Cir. 2004) ("There is no statutory or doctrinal prohibition against an appellate court issuing a COA *sua sponte* on issues not specified by a habeas petitioner.") (Tjoflat, J., specially concurring). *But see* Smaldone v. Senkowski, 273 F.3d 133, 139 (2d Cir. 2001) (per curiam) (finding lack of appellate jurisdiction when argued claim was not within amended COA). A court may also dismiss a COA as improperly granted, although the distinction between this device and summary affirmance is slight. *See, e.g.*, United States v. Sylvester, 258 F. App'x 411, 412 (3d Cir. 2007); Khaimov v. Crist, 297 F.3d 783, 786 & n.2 (8th Cir. 2002) (noting explicitly the similarity to the Supreme Court's practice of dismissing petitions for certiorari that have been improvidently granted); *see also* Cutler, *supra* note 10, at 325–26 (discussing how "[t]he circuits have divided into three separate camps concerning what procedure appellate courts must follow if the district court's certificate possibly fails to comply with section 2253(c)(2)'s substantial-showing-of-the-denial-of-a-constitutional-right standard").

71. Indeed, whether a COA order *must* conform with § 2253(2) and § 2253(3) as jurisdictional prerequisites was the subject of a circuit split until recently. *See* Ramunno v. United States, 264 F.3d 723, 725 (7th Cir. 2001). In *Gonzalez v. Thaler*, the Supreme Court determined that noncompliance with subsections 2 and 3 did not cost the reviewing court its jurisdiction over the appeal. 132 S. Ct. 641, 649 (2012). I will return to *Gonzalez* later in this piece.

72. *See* Santana v. United States, 98 F.3d 752, 757 (3d Cir. 1996) (noting the changed standard); Herman v. Johnson, 98 F.3d 171, 173 (5th Cir. 1996) (same).

tive history exists on § 2253 suggests that Congress, at least initially, intended to "enact[ ] the standard of *Barefoot*,"[73] this was a puzzling substitution. The punt of *Barefoot* was one by the judiciary for the judiciary, and had time to mature for years before AEDPA. This new language, a legislative creation, was—assuming that Congress meant what it said—untested. In the next few years, courts would struggle to determine the reach of the new § 2253—a matter of some urgency, given the multitudinous new (and often quite vague) restrictions contained in the rest of AEDPA.[74]

IV.
POST-AEDPA, PRE-*SLACK*: THE STRUGGLE

In the wake of the revisions to § 2253(c), the lower courts were confronted with a deceptively simple question with broad implications: Were there issues that were appealable under the old probable cause regime that were no longer so? Caught up in this inquiry were the first attempts to set in stone the now-familiar aspects of habeas practice post-AEDPA, particularly the new relationship between the district and appellate courts forged by the transition from CPC to COA.[75]

Courts first were required to confront the transition from "federal" to "constitutional," and with it the shift in substantive standards, under the new regime. In an early post-AEDPA decision, the Second Circuit concluded that the COA requirement was, at least for substantive questions presented by state prisoners, functionally identical to the old *Barefoot* standard, although it acknowledged that other courts were not all in agreement.[76] By contrast, while federal prisoners could still invoke nonconstitutional matters in

73. H.R. REP. NO. 104-23, at 9 (1995). As the Third Circuit would point out, the silent switch to "constitutional" has no clear explanation in the subsequent legislative history. *See* United States v. Cepero, 224 F.3d 256, 263–64 (3d Cir. 2000) (en banc), *overruled on other grounds by Gonzalez*, 132 S. Ct. at 649–50.

74. Some courts would, on occasion, give up entirely. *See* Miller-El v. Cockrell, 537 U.S. 322, 348–49 & n.* (2003) (Scalia, J., concurring) (indicating several cases where courts of appeal either applied flatly incorrect COA standards or ignored the COA prerequisite entirely).

75. It is easy to forget, these many years on, what an astonishing departure from precedent AEDPA was when it was first enacted. Given that the Illegal Immigration Reform and Immigrant Responsibility Act and the Prison Litigation Reform Act were passed around the same time, much of the basic jurisprudence associated with both detainees and prisoners changed in a few short years.

76. *See* Reyes v. Keane, 90 F.3d 676, 680 (2d Cir. 1996); *accord* Lennox v. Evans, 87 F.3d 431, 433 (10th Cir. 1996), *overruled on other grounds by* Lindh v. Murphy, 521 U.S. 320 (1997). The Ninth Circuit had described the new COA gateway as "more demanding" than the old standard, but had declined to elaborate on that

their § 2255 motions to vacate, they would no longer be able to seek appellate review of the district court's resolution of those claims—an odd outcome.[77]

Judicial effort was also spent determining how, if at all, the new COA standard affected an appellant's ability to seek review of procedural questions. In this regard, the Fifth Circuit issued a series of trailblazing decisions foreshadowing the direction eventually taken by the Supreme Court. In *Whitehead v. Johnson*,[78] the court held that when "the applicant for COA challenges the district court's dismissal for a reason not of constitutional dimension," he must first "make a credible showing that the district court erred." If he does so, the reviewing court must then determine "whether the petitioner has made a substantial showing of the denial of a constitutional right on one or more of his underlying claims," with doubts to be resolved in the applicant's favor.[79] But the Fifth Circuit also recognized that if a district court were to discuss only procedural questions, conducting an in-depth analysis of the merits of a petition at the COA stage would be premature, possibly prejudicial to the petitioner, and could run afoul of correct COA procedure.[80] Thus, despite language in earlier cases "suggest[ing] that we should proceed to examine the constitutional claims before granting a

interpretation in the particular opinion cited by the *Reyes* court. *See* Williams v. Calderon, 83 F.3d 281, 286 (9th Cir. 1996).

77. United States v. Gordon, 172 F.3d 753, 754 (10th Cir. 1999); Young v. United States, 124 F.3d 794, 799 (7th Cir. 1997); *see also* Mateo v. United States, 310 F.3d 39, 41 (1st Cir. 2002) ("The prospect of a constitutional argument is needed to permit the COA to be granted; but once back in district court Mateo is free—on a first section 2255 motion—to proffer non-constitutional claims. Section 2255, which governs federal habeas, extends beyond constitutional claims."). Intriguingly, the Seventh Circuit allows for the appeal of purely statutory issues as long as *one* of the claims upon which a COA is granted implicates a constitutional right. *See* Davis v. Borgen, 349 F.3d 1027, 1029 (7th Cir. 2003); *see also* Ramunno v. United States, 264 F.3d 723, 725 (7th Cir. 2001). No other circuit appears to have explicitly adopted this approach. *See, e.g.*, United States v. Hadden, 475 F.3d 652, 660 n.5 (4th Cir. 2007); United States v. Taylor, 454 F.3d 1075, 1078–79 (10th Cir. 2006). Of course, this debate may be much ado about nothing; "[d]espite the language of the statute, the right alleged violated on a [Section] 2255 motion virtually always must be a [c]onstitutional right" to have a chance of success. JOSEPHINE R. POTUTO, PRISONER COLLATERAL ATTACKS: FEDERAL HABEAS CORPUS AND FEDERAL PRISONER MOTION PRACTICE § 5.8, at 178 (1991).

78. 157 F.3d 384 (5th Cir. 1998).

79. *Id.* at 386. *Whitehead* followed other Fifth Circuit cases that dealt with specific procedural questions, such as the reviewability of procedural default. *See* Robison v. Johnson, 151 F.3d 256, 262 (5th Cir. 1998); Murphy v. Johnson, 110 F.3d 10, 11 (5th Cir. 1997).

80. *Whitehead*, 157 F.3d at 388.

COA," the court decided: "once we conclude that the district court erred in dismissing an application because of failure to exhaust, we vacate and remand to the district court to address the merits of the habeas claims in the first instance," thereby granting a COA and summarily vacating.[81]

The Fifth Circuit affirmed its approach in the later *Sonnier v. Johnson*.[82] There, the court determined that the appellant "ha[d] made a credible showing that the district court erred in dismissing his § 2254 application as barred by the post-AEDPA one-year statute of limitations."[83] But, given the state of the proceedings below, the court recognized that an analysis of the applicant's "substantial showing" would be premature, especially as the district court never reached the merits, again granting a COA and remanding in the same step.[84]

Other circuits also struggled to reconcile the AEDPA revisions with prior practices. The Eleventh, for example, faced head-on whether its prior holding that the pre- and post-AEDPA standards for adjudicating *merits* COA questions extended to its analysis of procedural questions. The court answered in the affirmative, determining that "the language of *Barefoot* provides the analytical framework for determining when COAs should be granted to permit such appeals." Furthermore, the court declined to decide whether it would deny a COA on procedural grounds "when the substantive claims are facially meritless," noting only that the petitioner's were not.[85] The Eighth Circuit, meanwhile, endorsed a more radical view. It determined that the statutory language addressing "constitutional" rights was directed only at:

> the sort of showing required for a petitioner to obtain appellate review of the merits of his or her claims for habeas corpus or § 2255 relief; [o]therwise, a final order entered by a district court based upon a question antecedent to the merits, if adverse to the petitioner, could never be reviewed on appeal.[86]

81. *Id.*

82. 161 F.3d 941 (5th Cir. 1998) (per curiam).

83. *Id.* at 945.

84. *Id.* at 945–46.

85. Henry v. Dep't of Corr., 197 F.3d 1361, 1364–66 & n.2 (11th Cir. 1999).

86. Nichols v. Bowersox, 172 F.3d 1068, 1070 n.2 (8th Cir. 1999) (en banc), *abrogated on other grounds by* Riddle v. Kemna, 523 F.3d 850, 856 (8th Cir. 2008). The dissenting Judge Arnold would have preferred a rule in which the prisoner's appeal would be contingent upon "some kind of abbreviated showing on the merits." *Id.* at 1078 (Arnold, J., dissenting).

Other circuits have observed that any analysis of the merits of a petition was not, apparently, a requirement of the COA process.[87]

The approaches may have differed, but all courts agreed on the basic idea: appeals from procedural dispositions were still possible. In *Morris v. Horn*,[88] a decision issued not long before *Slack*, the Third Circuit surveyed the landscape and agreed with the petitioner that (despite the Commonwealth's contention to the contrary) "courts, including this one, have granted certificates of appealability" on procedural questions that did not implicate the merits of the underlying petition.[89] The Third Circuit rejected the Eighth Circuit's approach, and instead adopted the Fifth Circuit's test as modified by *Sonnier* and *Whitehead*: a petitioner must make: "(1) a credible showing that the district court's procedural ruling was incorrect, and (2) a substantial showing that the underlying habeas petition *alleges a deprivation* of constitutional rights."[90]

It was into this breach that the Supreme Court finally tread. In doing so, it aimed to set a standard that would apply, universally, across the various circuits. However, as I will show, while its opinion would establish some uniformity—it would jettison the Eighth Circuit's approach, for example—it would lead to no greater consensus on what kind of merits showing was required by a prisoner seeking a COA.

V.
THE SUPREME COURT WEIGHS IN: *SLACK* AND ITS PROGENY

A. Slack Itself

In 2000, the issue of procedural appeals came before the Supreme Court, and Justice Kennedy's opinion in *Slack v. McDaniel*[91] allowed the lower courts to breathe a sigh of relief. Kennedy clarified that the "reasonable jurists" standard still held true and that appeals of both substantive and procedural questions were possible under AEDPA. But while Justice Kennedy's opinion made some headway in deciphering the requirement of a "substantial showing,"

87. *See, e.g.*, Gaskins v. Duval, 183 F.3d 8, 9 n.1 (1st Cir. 1999) (per curiam) (collecting cases).

88. 187 F.3d 333 (3d Cir. 1999).

89. *Id.* at 340 (collecting cases).

90. *Id.* at 340 & n.4 (emphasis added).

91. 529 U.S. 473 (2000). For an excellent, contemporaneous summary of the factual background of *Slack,* see Laurence A. Benner et al., *Criminal Justice in the Supreme Court: A Review of United States Supreme Court Criminal and Habeas Corpus Decisions (October 4, 1999 – October 1, 2000)*, 37 CAL. W. L. REV. 239, 273–79 (2001).

it introduced a new variable—the "valid claim"—into the equation, while arguably modifying sub silentio the actual statutory language.

In *Slack*, the Court was tasked with resolving whether an appeal may be taken if "the District Court relies on procedural grounds to dismiss the petition."[92] The State argued that "only constitutional rulings may be appealed."[93] Writing for the Court, Justice Kennedy disagreed. "The writ of habeas corpus plays a vital role in protecting constitutional rights," he explained, and "[i]n setting forth the preconditions for issuance of a COA under § 2253(c), Congress expressed no intention to allow trial court procedural error to bar vindication of substantial constitutional rights on appeal."[94] He first emphasized that AEDPA codified the language of *Barefoot*, with the substitution of "constitutional" for "federal."[95] He then applied the *Barefoot* standard—focusing on its "jurists of reason" verbiage—to both substantive and procedural COA processes, effectively setting the standard for each category.

Under Justice Kennedy's construction, the showing required to satisfy § 2253(c) was "straightforward" when a district court denied a petition on the merits: "[t]he petitioner must demonstrate that reasonable jurists would find the district court's assessment of the constitutional claims debatable or wrong."[96] Less "straightforward," and admittedly "somewhat more complicated," was the application of § 2253(c) to procedural denials made "without reaching the prisoner's underlying constitutional claim."[97] To this situation, Justice Kennedy applied a modification to the *Barefoot* standard:

> [A] COA should issue when the prisoner shows, at least, that jurists of reason would find it debatable whether the petition states a valid claim of the denial of a constitutional right and that jurists of reason would find it debatable whether the district court was correct in its procedural ruling.[98]

This construction for procedural appeals, according to Justice Kennedy, would "give[ ] meaning to Congress' requirement that a prisoner demonstrate substantial underlying constitutional claims and is in conformity with the meaning of the 'substantial showing' standard provided in *Barefoot*."[99] Kennedy further emphasized that this

92. *Slack*, 529 U.S. at 483.
93. *Id.*
94. *Id.*
95. *Id.*
96. *Id.* at 484.
97. *Id.*
98. *Id.*
99. *Id.*

test was a two-part inquiry; courts should feel free to "dispose of [a COA] application in a fair and prompt manner if it proceeds first to resolve the issue whose answer is more apparent from the record and arguments."[100]

At this juncture, it is helpful to examine what *Slack* purports to do in a procedural context. First, it is hard to escape the conclusion (despite Kennedy's statements to the contrary) that he has fundamentally, if not unjustifiably, re-written the poorly drafted § 2253(c)(2) to eliminate some of its ambiguity. In this interpretation, Kennedy's "state a valid claim/procedurally correct ruling" text *supplants* the "substantial showing" language in § 2253(c)(2). In other words, a petitioner is not required to satisfy both Kennedy's test and the "substantial showing" language. Rather, in doing the first, a petitioner has also accomplished the second. Alternatively, *Slack* can be read to modify "substantial showing" with "states a valid claim," while adding on the "procedurally correct ruling" language as an addendum to the statutory requirement. In this reading, a petitioner satisfies the gatekeeping requirement of *Barefoot* with his procedural argument, while fulfilling the "substantial showing" requirement by stating a valid constitutional claim.[101] Both of these approaches, while analytically distinct, arrive at the same destination: Kennedy's test, whether marshaled in tandem with or separately from § 2253(c)(2), is the operative guide for appealability of procedural issues.

Second, regardless of the path one chooses with regard to supplanting versus supplementing, Kennedy has replaced one vagary with another. What is required, the reader wonders, to "state a *valid claim*" of the denial of a constitutional right? Kennedy did not explicitly anchor his language in any prior Court precedent. "Valid"

100. *Id.* at 485.

101. The Seventh Circuit has apparently chosen the latter construction, which animates, in part, its reading that *Slack* allows statutory claims to piggyback on constitutional claims. *See* Ramunno v. United States, 264 F.3d 723, 725 (7th Cir. 2001). Further, I note that Justice Scalia saw no difference between "state a valid claim" and "make a substantial showing" in his *Miller-El* concurrence, characterizing it instead as an *additional burden* habeas petitioners must meet in order to obtain review of their procedural claims. *See* Miller-El v. Cockrell, 537 U.S. 322, 349–50 (2003) (Scalia, J., concurring). Justice Scalia takes pains in his *Miller-El* concurrence to emphasize that § 2253(c)(2) contains a necessary, but not sufficient, requirement for a COA; courts are free to demand more. *Id.* at 349. The idea that § 2253(c)(2) sets a floor is at odds with the general tenor of the Supreme Court's habeas jurisprudence as developed, as with the basic idea that § 2253(c)(2) was intended to incorporate, in some form, the prevailing *Barefoot* standard.

could mean meritorious or genuine,[102] yet "stating a claim" is indelibly associated (in federal practice, at least) with a minimal proffer—with satisfying a pleading standard that requires little to no demonstration of likely success.[103] Combining the two possible meanings leads to some dissonance. The Supreme Court had never invoked this exact language of a "valid claim" in the context of habeas, and its use in any other context was scanty. Indeed, *Slack* left the precise nature of the showing completely undefined.

Finally, a subordinate aside in the opinion raised other questions regarding the *minimum* showing required to gain a COA on a procedural question. Kennedy presents the "jurists of reason" test, but qualifies it with a cautionary "at least."[104] A petitioner could, of course, far exceed the minimum requisite showing and still obtain a COA. As we will see, applications of *Slack* would struggle to define the minimum showing required for success.

B. Post-Slack Supreme Court Decisions

Slack is, at the time of writing, the final Supreme Court case directly on point. Although the Court has revisited COAs as a category, it has not precisely revisited the procedural question that gave rise to its formulation in that case. But subsequent Supreme Court opinions *do* emphasize the Court's view that the COA requirement, far from being an impassable gatekeeper, is not meant to allow only the extraordinary and exceptional cases.

By far the most notable post-*Slack* case on the topic of COAs is *Miller-El v. Cockrell*,[105] the other cornerstone of modern COA juris-

102. *See, e.g.*, Logan v. Zimmerman Brush Co., 455 U.S. 422, 444 (1982) (Powell, J., concurring) (contrasting "valid" claims with "frivolous" claims).

103. *See, e.g.*, Conley v. Gibson, 355 U.S. 41, 45–46 (1957). At the time *Slack* was decided, the seismic shift of *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544 (2007), was still several years in the future.

104. *Slack*, 529 U.S at 484.

105. 537 U.S. 322 (2003). Like so many path-setting habeas cases, *Miller-El* is a death penalty case, and its holding, while obviously resilient, may be subject to a certain degree of a "death is different" distinction. Certainly, courts often err on the side of caution in death-penalty cases. *See, e.g.*, Pippin v. Dretke, 434 F.3d 782, 787 (5th Cir. 2005) ("[A]ny doubt as to whether a COA should issue in a death-penalty case must be resolved in favor of the petitioner."). I take it, as have the circuit courts generally, at face value. *See, e.g.*, McLaughlin v. Shannon, 454 F. App'x 83, 85 & n.2 (3d Cir. 2011) (per curiam) (using *Miller-El* in a noncapital case); *cf.* Porterfield v. Bell, 258 F.3d 484, 487 (6th Cir. 2001) ("While we do not necessarily disagree with the view that trial courts should err on the side of caution when it comes to the certification of claims that arguably have merit, there is nothing to suggest that *Slack* does not apply with equal force in capital cases."). For an

prudence. Unlike *Slack*, *Miller-El* is firmly rooted in the simpler of Justice Kennedy's two COA scenarios, a denial on the merits:

> The United States District Court for the Northern District of Texas, after reviewing the evidence before the state trial court, determined that petitioner failed to establish a constitutional violation warranting habeas relief. The Court of Appeals for the Fifth Circuit, concluding there was insufficient merit to the case, denied a [COA] from the District Court's determination. The COA denial is the subject of our decision.[106]

Again writing for the Court, Justice Kennedy reemphasized that the COA determination is distinct from a full analysis of the merits of a petitioner's claims. Instead, it is a "threshold inquiry," one that "requires an overview of the claims in the habeas petition and a general assessment of their merits."[107] In fact, Justice Kennedy wrote that a full merits analysis is explicitly forbidden by the statute, because a circuit court that races to a merits disposition is essentially "deciding an appeal without jurisdiction."[108] Invoking language that stretched back to *Barefoot*, Justice Kennedy stressed that "[a] petitioner satisfies this standard by demonstrating that jurists of reason could disagree with the district court's resolution of his constitutional claims or that jurists could conclude the issues presented are adequate to deserve encouragement to proceed further."[109] In summary:

> A prisoner seeking a COA must prove "something more than the absence of frivolity" or the existence of mere "good faith" on his or her part. We do not require petitioner to prove, before the issuance of a COA, that some jurists would grant the petition for habeas corpus. Indeed, a claim can be debatable even though every jurist of reason might agree, after the COA has been granted and the case has received full consideration, that petitioner will not prevail.[110]

Turning to the specific matter on appeal—an alleged violation of *Batson*[111]—Justice Kennedy took the chance to apply the abstract *Slack* standard to a distinct constitutional fact pattern, concluding that a COA should be granted if there was "any evidence demon-

overview of *Miller-El* itself, see Kristy Bowling, Note, Miller-El v. Cockrell*: Procedural Rules to Protect Prisoners' Rights*, 35 U. TOL. L. REV. 723 (2004).

106. Miller-El v. Cockrell, 537 U.S. 322, 326–27 (2003).

107. *Id.* at 336.

108. *Id.* at 336–37.

109. *Id.* at 327.

110. *Id.* at 338 (citations omitted).

111. Batson v. Kentucky, 476 U.S. 79 (1986).

strating that, despite the neutral explanation of the prosecution, the peremptory strikes in the final analysis were race based."[112] But Kennedy emphasized that, from the circuit's perspective, the relevant inquiry was not whether or not it could agree or disagree with the state court's analysis in the first instance—even though that might be the thrust of a final merits evaluation—but whether it could agree with *the district court's decision*. Here, Kennedy decided, the circuit court should have found plenty to quibble with, because "the District Court did not give full consideration to the substantial evidence petitioner put forth in support of the prima facie case;"[113] in doing the same thing, the circuit compounded the lower court's error. The proper question, as Kennedy stressed in his closing, was whether "the District Court's decision was debatable." Determining that it was, the Supreme Court remanded to the Fifth Circuit.[114]

Miller-El is hardly a skeleton key to the mysteries of COAs; like *Slack* before it, it raises almost as many questions as it answers, and, as a capital case, it prompts reasonable concerns about whether the Supreme Court expected adherence to its bipartite level of scrutiny in noncapital matters.[115] But Justice Kennedy states rather firmly that the appellate court's gaze should be focused on the district court's decision and should not leapfrog to a full analysis of AEDPA deference, bypassing the district court entirely; *Miller-El* clearly forbids a functionally de novo, premature appellate review of the state court's decision, AEDPA deference notwithstanding. Nor should the firm impression that the petitioner simply will not succeed stand as a bar to a COA.

Supreme Court cases since may not have expanded on *Miller-El* or *Slack*, but they do show that the Supreme Court is at least facially serious about adhering to its pronouncements in both. In *Tennard v. Dretke*,[116] for example, the Court once again affirmed the line of demarcation between: (1) the COA standard, focusing on the district court's performance; and (2) the later question of AEDPA deference at the merits stage, focusing on the state court's

112. *Miller-El*, 537 U.S. at 340.

113. *Id.* at 341.

114. *Id.* at 348. For a critical look at this outcome, see Cutler, *supra* note 10, at 341 (calling *Miller-El* "intellectually unsatisfying" because of its inconsistency with pre-AEDPA jurisprudence on CPCs).

115. *See* Jordon T. Stanley, Comment, *"Deference Does Not Imply Abandonment or Abdication of Judicial Review": The Evolution of Habeas Jurisprudence under AEDPA and the Rehnquist Court*, 72 UMKC L. Rev. 739, 762–64 (discussing "meaning" of *Miller-El*).

116. 542 U.S. 274 (2004).

determination as viewed through 28 U.S.C. § 2254(d).[117] Ultimately, *Miller-El* and *Slack* have been the last word.[118] The baton has been passed to the lower federal courts to put the Supreme Court's new guidance to the test.

VI.
THE CIRCUITS RESPOND

Slack largely replaced the individual tests the circuits had employed in their COA determinations; from this point onward, both *Slack* and *Miller-El* would become the go-to references in opinions and orders.[119] But while the Supreme Court had handed the courts of appeals a single unified standard, they would diverge markedly on how to interpret that standard, especially in the context of procedural COAs and Justice Kennedy's "state a valid claim" pronouncement.

Two major interpretations have since developed. The first, which I will call the "permissive" standard, sees in Justice Kennedy's "valid claim" language a basic "screening" function. In this interpretation, a petitioner's proffer on the merits in a procedural appeal would need to pass only the most cursory scrutiny. The second, which I will call the "strict" standard, focuses on the original statutory requirement of a substantial showing of the denial of a constitutional right, often applying *Miller-El*'s language about merits denials to the intrinsically distinct realm of procedural denials—in some cases, seeming to combine the two. In this more demanding model, an underlying claim that lacks merit, and/or that has been definitively denied by a state court (implicating AEDPA deference),

117. *Id.* at 282.

118. With one exception: *Harbison v. Bell*, 556 U.S. 180 (2009), a small case that may be anything from transformative to utterly inconsequential. For more on *Harbison*, see *infra* note 182.

119. There are numerous exceptions to this, of course. For example, in *Satizabal v. Folino*, the Third Circuit relied on both *Slack* and its pre-*Slack* opinion in *Morris v. Horn*, observing that "a certificate of appealability may issue only if the petitioner makes both a credible showing that the procedural ruling was incorrect and a substantive showing that the underlying habeas corpus petition alleges a deprivation of constitutional rights." *Satizabal v. Folino*, 318 F. App'x 78, 80–81 (3d Cir. 2009) (citing *Morris v. Horn*, 187 F.3d 333, 340–41 (3d Cir. 1999)). It is notable, however, that the Third Circuit has apparently not relied on *Morris* again in quite this way, although it has relied, on occasion, on opinions that were near-contemporaneous with *Slack*, such as *United States v. Brooks*, 230 F.3d 643 (3d Cir. 2000), where it held that it could not grant a COA "unless the issue is procedural and the underlying petition raises a substantial constitutional question." *Id.* at 646; *see also* Hubley v. Superintendent, SCI Camp Hill, 57 F. App'x 927, 929–30 (3d Cir. 2003) (citing *Brooks*, 230 F.3d at 646).

can be cause for denying a COA. For example, a court might invoke the "valid claim" language, but then immediately retreat to the "substantial showing" statutory requirement. Still other courts have been noncommittal, or appear to inconsistently grant COAs across various cases. And, of course, simply because a court applies a particular standard in a high-profile, published case does not mean that it utilizes the same standard internally for the purposes of mundane, day-to-day case management, and especially in those orders not provided to legal publishers.

A. The First School: Permissive Scrutiny, Cognizability, and the "Quick Look"

At one end of the spectrum is a standard focusing on cognizability—literally, "stating a valid claim." Developed in an array of circuits, and deployed in what is undoubtedly a haphazard fashion, this approach to procedural COAs involves only a cursory analysis of the merits of the underlying petition. Namely, if the petition conceivably states a cognizable federal habeas claim, then the appeal should be allowed to proceed in some fashion. The "debate" by the jurists of reason is over the cognizability of the claim, not the underlying merits of the claim. This approach, which takes Justice Kennedy's language from *Slack* at its word, would in theory allow all appeals of procedural denials or dismissals that are both debatable and affect a claim that is not indisputably meritless.

One variant of this permissive approach germinated in the Seventh Circuit shortly after *Slack* was decided. In *Jefferson v. Welborn*,[120] the underlying habeas corpus petition had been dismissed by the district court as untimely filed.[121] Jefferson sought a COA from the Seventh Circuit so that he might have "an opportunity to have the district court consider his claims on the merits."[122] Because the district court had not reached those merits, the COA determination turned on "whether the district court correctly applied the rules governing the limitations period for filing § 2254 petitions that are found in 28 U.S.C. § 2244(d)."[123] In its opinion, the Seventh Circuit did not merely find the timeliness analysis by the district court to be debatable by jurists of reason, it flatly disagreed with the outcome, determining that the petitioner's state court action was "'properly filed' for purposes of § 2244(d)(2)," rendering the eventual federal habeas petition filed "within the permitted time" to do

120. 222 F.3d 286 (7th Cir. 2000).
121. *Id.* at 287.
122. *Id.*
123. *Id.* at 288.

so by statute.[124] But what of the merits of Jefferson's petition, which had not yet been addressed? Referencing *Slack*, the court took a "quick look at the claims Jefferson want[ed] to raise in his petition," and "did not find them so thoroughly lacking that such a step would be appropriate right now," because at least some of the allegations (such as ineffective assistance of counsel) "facially allege[d] the 'denial of a constitutional right.'"[125] Hence, in a single maneuver, the court granted a COA, vacated the judgment of the district court, and remanded for further proceedings.[126]

Jefferson is a fascinating case because it relies directly on *Slack* for the proposition that the merits scrutiny should be minimal. Judge Wood saw no need to parse Justice Kennedy's statement about "stating a valid claim," instead finding it uncontroversial—or at least not deserving of a tortured exegesis—that *Slack* had cautioned the lower courts to avoid an in-depth analysis of the merits of the petition before considering a COA.

Other courts adopted similar standards, at least for a time. In *Lambright v. Stewart*,[127] the Ninth Circuit employed a "quick look" analysis of its own, emphasizing that Justice Kennedy's articulation of "state a valid claim" was meant to echo the plaintiff-favoring application of FED. R. CIV. P. 12(b)(6).[128] The First,[129] Fourth,[130] and Tenth[131] Circuits have also mentioned taking a "quick look" at a petition. The Tenth has clarified that its "quick look" is intended to discern whether the petitioner has "facially alleged" a constitutional

124. *Id.* at 289.

125. *Id.* (citing Slack v. McDaniel, 529 U.S. 473, 484–85 (2000)). The court also suggested that, if Jefferson's claims were all "utterly without merit," it could "affirm the dismissal on that alternative ground." Jefferson v. Welborn, 222 F.3d 286, 289 (7th Cir. 2000). It is unclear whether the court was suggesting that it would "deny a COA if the claims were not cognizable," "deny a COA if the claims were cognizable but facially meritless," or "grant a COA and summarily affirm if the claims were cognizable but facially meritless."

126. *Id.*

127. 220 F.3d 1022 (9th Cir. 2000).

128. *Id.* at 1026–27 & n.5.

129. *See, e.g.*, St. Yves v. Merrill, 78 F. App'x 136, 137 (1st Cir. 2003); Mateo v. United States, 310 F.3d 39, 41 (1st Cir. 2002).

130. *See, e.g.*, Rouse v. Lee, 314 F.3d 698, 702 (4th Cir. 2003), *vacated on other grounds en banc*, 339 F.3d 238, 243 (4th Cir. 2003); Hernandez v. Caldwell, 225 F.3d 435, 437 (4th Cir. 2000).

131. *See, e.g.*, Gibson v. Klinger, 232 F.3d 799, 802–03 (10th Cir. 2000); *see also* Frazier v. Colorado, 405 F. App'x 276, 278 (10th Cir. 2010) (concluding that a "pure matter of state law is simply not cognizable in habeas" and, therefore, cannot support a COA request).

claim,[132] language also briefly adopted by the Fifth Circuit and hinted at in some opinions by the Third.[133] The First Circuit, in *Mateo v. United States*, alternatively referred to its "quick look" goal as seeking to determine whether "the constitutional claim is . . . colorable."[134] While acknowledging that *Slack* did not mandate the "quick look" approach, Judge Boudin thought it presented "the same impulse as *Slack* to protect nascent constitutional claims; and it certainly does not bend the language of § 2253 any more than *Slack* itself."[135]

Despite gaining prominence in some circuits immediately after *Slack* was decided, the permissive standard either fell out of favor among those courts or simply faded into the background soon after it was established. For example, although an early adopter, the Ninth Circuit has referenced its "quick look" standard only a few times since *Lambright*, and has forsaken it entirely in its published output since 2008.[136] Only the Tenth Circuit, which, as stated above, has the intriguing tendency to release all of its COA determinations for archiving by legal publishers (and which, as a possible consequence, tends to write full opinions for use at this screening stage), has recently invoked the permissive approach to COA determinations.[137]

B. The Second School: The Strict Standard and Fealty to the Language of § 2253(c)(2)

Some courts have found that a more searching analysis of the merits is appropriate at the COA stage. In these cases, whether the petitioner has "stated" a cognizable habeas claim is simply one part of the overall inquiry. If a sufficient record exists to allow the appellate court to ascertain that the petitioner's claims are of little overall merit, it may deny a COA on that ground without reaching the district court's procedural ruling.

132. Fleming v. Evans, 481 F.3d 1249, 1259 (10th Cir. 2007) (quoting *Gibson*, 232 F.3d at 803).

133. United States v. Asemani, 77 F. App'x 264, 265 (5th Cir. 2003); Hubley v. Superintendent, SCI Camp Hill, 57 F. App'x 927, 930 (3d Cir. 2003).

134. *Mateo*, 310 F.3d at 40.

135. *Id.* at 41.

136. Sechrest v. Ignacio, 549 F.3d 789, 803–04 (9th Cir. 2008); Lopez v. Schriro, 491 F.3d 1029, 1040 (9th Cir. 2007); Nardi v. Stewart, 354 F.3d 1134, 1139–40 (9th Cir. 2004).

137. *See, e.g.*, McCosar v. Standifird, 488 F. App'x 311, 313 n.1 (10th Cir. 2012). Reflecting the Tenth Circuit's unorthodox approach to COAs, the *McCosar* panel both granted a COA and deferred to the district court's procedural determination in a single opinion. *See id.* at 313–14.

The Fifth Circuit has perhaps most explicitly articulated this "sliding scale" approach. In *Houser v. Dretke*,[138] where a petition had been dismissed by the district court as procedurally barred for failure to exhaust administrative remedies,[139] the court tackled the difficult question of "what *Slack* had in mind" for procedural appeals:[140]

> Assume that petitioner has stated a "debatable" issue concerning the correctness of the district court's procedural denial of habeas relief. Then, if the district court pleadings, the record, and the COA application demonstrate that reasonable jurists could debate whether the petitioner has *made a valid claim* of a constitutional deprivation, a COA will issue. If those same materials make it clear that reasonable jurists could not debate whether the petitioner has made a valid claim of a constitutional deprivation, the COA will be denied. If those materials are unclear or incomplete, then COA should be granted, and the appellate panel, if it decides the procedural issue favorably to the petitioner, may have to remand the case for further proceedings.[141]

Note what the court has done here. For cases in which it is impossible to analyze anything other than the wisdom of the lower court's procedural ruling, a "quick-look-like" approach is endorsed. However, in cases where the record is available, a COA should be denied when the petitioner has not "made" a valid claim of the denial of a constitutional right.[142] The court, in "looking to [the] application for a COA, [the] original petition, the district court's opinion, the record, and the briefs filed in the district court on [the respondent's] behalf," determined that "no reasonable jurist could debate that [the petitioner] fails to state a constitutional deprivation *for which habeas relief is warranted*."[143] In other words, a more searching merits standard is employed when the record allows for it.

In another recent decision, the Fifth Circuit suggested that the *Slack* "state a valid claim" language should be interpreted to require the petitioner to raise "reasonably debatable claims of the denial of

138. 395 F.3d 560 (5th Cir. 2004).

139. *Id.* at 561. *Houser* involved an unusual substantive issue: a challenge to a good-time credit revocation proceeding by a state prisoner. *Id.* at 561–62.

140. *Id.* at 562.

141. *Id.* (emphasis added) (citation omitted).

142. *Id.*; *see also* Singleton v. Cooper, 456 F. App'x 417, 418–19 (5th Cir. 2011) (per curiam) (employing "made" language, as well).

143. *Houser*, 395 F.3d at 562 (emphasis added).

constitutional rights."[144] Again, note that while *Slack* suggested that the "stating" of the claims should be reasonably debatable, the Fifth Circuit's construction shifts the debate from the invocation of the claims to the claims themselves. Other Fifth Circuit cases show this searching standard in action, often while facially invoking the "state a valid claim" requirement of *Slack*.[145]

The Eighth Circuit has endorsed a similar reading of *Slack*. The court observed that no COA should be issued if "there is no merit to the substantive constitutional claims."[146]

Other courts effectively equate "stating a valid claim" of the denial of a constitutional right with making a "substantial showing" of one. For example, in *Bell v. Florida Attorney General*,[147] the Eleventh Circuit observed that the district court "erred in failing to specify whether jurists of reason would find it debatable that Bell's petition states a valid claim of the denial of a constitutional right."[148] It therefore vacated the COA as improvidently granted, and remanded for the district court to consider "what claims, if any, in Bell's petition for habeas corpus make a 'substantial showing of the denial of a constitutional right.'"[149] In an alternative articulation, the court read the *Slack* pronouncement to require a "show[ing] . . . that one or more of the claims [the petitioner] has raised presents *a substantial constitutional issue*."[150]

144. Womack v. Thaler, 591 F.3d 757, 758 (5th Cir. 2009) (per curiam).

145. *See, e.g.*, Morris v. Dretke, 379 F.3d 199, 206 (5th Cir. 2004); Graves v. Cockrell, 351 F.3d 143, 154–55 (5th Cir. 2003). However, it is, as always, difficult to distinguish between the actual use of a stricter standard and descriptive richness.

146. Khaimov v. Crist, 297 F.3d 783, 786 (8th Cir. 2002); *see also* Langley v. Norris, 465 F.3d 861, 863 (8th Cir. 2006) (utilizing the "no merit" language in the context of an ineffectiveness claim to resolve the prejudice prong at the COA stage).

147. 614 F.3d 1230 (11th Cir. 2010) (per curiam).

148. *Id.* at 1232.

149. *Id.* (quoting 28 U.S.C. § 2253(c)(2)).

150. Gordon v. Sec'y, Dep't of Corr., 479 F.3d 1299, 1300 (11th Cir. 2007) (per curiam) (emphasis added). It is worth noting here that early post-*Slack* cases from the Eleventh Circuit could be interpreted as leaning more toward the "quick-look" end of the analytical spectrum. *See, e.g.*, Franklin v. Hightower, 215 F.3d 1196, 1199–1200 (11th Cir. 2000) (per curiam) (quoting the "states a valid claim" language, while also noting that "the first claim's merit is certainly debatable among jurists of reason"); Roberts v. Sutton, 217 F.3d 1337, 1340 n.4 (11th Cir. 2000) ("[A]s long as reasonable jurists would find *the merits of at least one procedurally barred claim to be debatable*, we may move on and weigh the merits of the petitioner's procedural argument to determine if it satisfies the COA standard.") (emphasis added). In retrospect, the circuit's later decision that the merits—and not the actual cognizability—should be "debatable" finds its genesis in these opinions. *See also* Gonzalez v. Sec'y for the Dep't of Corr., 366 F.3d 1253, 1264 (11th Cir. 2004)

The upshot of the stricter standard is the denial of COAs in cases where a procedural question may exist, but the merits of the underlying petition are either undeveloped or appear to lack a chance of success. Even if a loosely cognizable constitutional claim may exist, a court can evaluate the facts presented by the record to deem it presumptively meritless and thus unworthy of a COA.[151] Thus, a petitioner may successfully "assert[ ] a constitutionally cognizable right in his habeas petition," but may still fail if he does not articulate a "reasonably debatable infringement of that right."[152] The sliding-scale approach implies an odd paradox for the petitioner: the more developed his claims are or the more complete the factual record is, the less forgiving the reviewing court will be, even when the district court's disposition did not touch upon the merits. For example, if a bare-bones habeas petition were erroneously denied on timeliness grounds, a "sliding scale" approach might compel a COA grant and a remand to the district court for further development of the factual record. By contrast, the same bare-bones petition which included, as an attachment, a state court decision disposing of the claim might not be so lucky, as an appellate court would no longer be looking at the claim in a factual vacuum.

C. The Pabon *Problem: The Third Circuit's Third Way*

Of all recent COA decisions, however, the Third Circuit's opinion in *Pabon v. Superintendent SCI Mahanoy*[153] is one of the most puzzling. Despite its pre-*Slack* activity in the COA realm, the Third Circuit fell quiet soon after *Slack* was decided.[154] In *Pabon*, the Circuit reentered the field, but in so doing it committed, in my opinion, an interpretive mistake: it tangled the *Slack* and *Miller-El* standards. The court was asked, in part, to tackle the question of

(characterizing COA as "filter[ing] out from the appellate process cases in which the possibility of reversal is too unlikely to justify the cost to the system of a full appellate examination"). These Eleventh Circuit decisions were, in fact, relied on by some of the "quick-look" courts to support those outcomes. *See, e.g.*, Mateo v. United States, 310 F.3d 39, 41 (1st Cir. 2002) (citing *Roberts*, 217 F.3d at 1339–40.).

151. *See, e.g.*, Houser v. Dretke, 395 F.3d 560, 562–63 (5th Cir. 2004).

152. Williams v. Martinez, 586 F.3d 995, 1001 (D.C. Cir. 2009).

153. 654 F.3d 385 (3d Cir. 2011), *cert. denied*, 132 S. Ct. 2430 (2012).

154. Although, it did suggest sympathy to the "quick-look" approach in 2007, *see* Goldblum v. Klem, 510 F.3d 204, 214 (3d Cir. 2007), and it did articulate a "raises a substantial constitutional question" test shortly after *Slack* was decided, *see* United States v. Brooks, 230 F.3d 643, 646 (3d Cir. 2000). *But see* Tomlin v. Britton, 448 F. App'x 224, 228 (3d Cir. 2011) (declining to issue a COA when the petitioner "has made no showing whatsoever, and certainly not a substantial showing, that his constitutional right to the effective assistance of appellate counsel was denied").

the *Slack* "state a valid claim" language, but appeared to instead embrace the more searching analysis of *Miller-El*, a case whose teachings were arguably inapposite. While the opinion suggests an attempt at imposing a kinder standard upon habeas petitioners, it may have accomplished precisely the opposite—an object lesson in the need to keep distinct the two separate *Slack* tests and to separate the language of *Miller-El* from the language of *Slack*.

Pabon, like *Slack*, is a bull's eye case on procedural appeals, with an interesting procedural twist. Pabon's petition raised *Bruton*[155] claims. The district court denied equitable tolling and dismissed the petition as untimely. Pabon appealed and sought a COA. The court of appeals issued a COA, but its order mentioned only the equitable tolling claim and did not "indicate" any constitutional claims.[156] Recognizing that its COA order was problematic, the court stayed oral argument and directed the parties to brief "whether Pabon had made a substantial showing of the denial of a constitutional right as required by 28 U.S.C. § 2253(c)."[157]

In its opinion, then, the court had to solve two different inquiries: first, had Pabon satisfied the merits-based portion of the *Slack* procedural test to bring the COA order into compliance; and, second, should the district court's order be affirmed or reversed? The first question appeared to be an excellent opportunity to finally settle the circuit's approach to *Slack*. Quickly, however, the court veered from *Slack* to *Miller-El*, quoting the latter's merits-based discussion of the showing required by the petitioner to succeed.[158] And while proceeding to make "a threshold inquiry regarding the application of *Bruton* and its progeny to Pabon's trial and conviction,"[159] the court's "cursory" consideration of the applicable facts of the case spanned numerous pages.[160] In the end, it concluded that "[u]nder the *Miller-El* standard, Pabon's alleged *Bruton* violation need only be debatable" and "[f]or the reasons explained

155. Bruton v. United States, 391 U.S. 123 (1968). "*Bruton* and its progeny establish[ ] that in a joint criminal trial before a jury, a defendant's Sixth Amendment right of confrontation is violated by admitting a confession of a non-testifying codefendant that implicates the defendant, regardless of any limiting instruction given to the jury." Johnson v. Tennis, 549 F.3d 296, 298 (3d Cir. 2008).

156. *See* Pabon v. Superintendent SCI Mahanoy, C.A. No. 08-1536 (3d Cir. Nov. 19, 2008).

157. *Pabon*, 654 F.3d at 392. The actual order did mention *Slack*'s "state a valid claim" language, albeit in a parenthetical. *See* Pabon v. Superintendent SCI Mahanoy, C.A. No. 08-1536 (3d Cir. May 12, 2010).

158. *Pabon*, 654 F.3d at 392–93.

159. *Id.* at 393.

160. *See id.* at 387–91, 393–98.

above, we conclude that reasonable jurists could debate whether Pabon has a meritorious claim."[161] Thus bringing the COA order into compliance, the court proceeded to the merits of Pabon's equitable tolling argument, reversing and remanding to the district court for an evidentiary hearing on Pabon's claim.[162]

Yet the first part of the court's exercise was arguably unnecessary—a clear demonstration of the problems of mixing the substantive *Miller-El* and procedural *Slack* standards. The court was not evaluating the merits of Pabon's petition; indeed, it had granted a COA on only the procedural question. Thus, the court's extraordinarily in-depth discussion of both the facts of the case and Pabon's strong merits showing, which would perhaps have been appropriate in a *Miller-El* analysis, was not necessary under the procedural prong of *Slack*. The court had jumbled the appropriate standards, and failed to recognize that, while procedural COA determinations should be guided by the spirit of *Miller-El*, applying that case's merits analysis is dangerous folly.

Odder still, the decision in *Pabon* is written as one generous to the petitioner, stressing the "threshold inquiry" and explicitly disclaiming the Commonwealth's efforts to impose a heightened standard of COA review.[163] Despite this, the court might possibly have made future success by procedural appellants more difficult. Subsequent panels might take *Pabon* to impose *Miller-El*'s merits standard as the equivalent of "stating a valid claim" in procedural appeals, requiring a searching inquiry that might even exceed the stricter standard imposed by other non-quick-look circuits. Of course, the opposite could be true as well; simply because the court arguably overshot the merits analysis in *Pabon* does not mean that the level of inquiry conducted therein is required. *Slack* does, after all, emphasize that a petitioner must show "at least" that he has stated a valid claim of the denial of a constitutional right.[164] Under this reading, *Pabon* can be safely characterized as illustrative rather than restrictive, showing a sufficient, but not necessary, proffer of merit, and at least one post-*Pabon* case has adopted that approach.[165] Taken liter-

161. *Id.* at 398.

162. *Id.* at 404.

163. *Id.* at 392–93 & n.9. I suspect that the extended merits analysis in *Pabon* might have been intended to signal to the district court—to which the matter was being remanded—that the Third Circuit found substantive worth in the underlying habeas corpus petition.

164. Slack v. McDaniel, 529 U.S. 473, 484 (2000).

165. Specifically, in *Gerber v. Varano*, 512 F. App'x 131 (3d. Cir. 2013) (per curiam), the Third Circuit emphasized its commitment to a threshold standard while declining to engage in a lengthy merits analysis à la *Pabon*. *See id.* at 134 &

ally, however, *Pabon*'s casual mixing of procedural and substantive merits standards could portend difficulty for future petitioners.

Indeed, the court's failure to explicitly distinguish between the procedural and the substantive modes of appellate COA review is further demonstrated in the Commonwealth's decision to seek certiorari. In footnote nine of the opinion, the court rejected the Commonwealth's argument that AEDPA deference should be used in evaluating the merits of the petition at the COA stage, observing that this was "precisely what the Supreme Court rejected in *Miller-El*."[166] The Commonwealth charged that, in so holding, the Third Circuit had functionally endorsed de novo review of state court decisions, in clear violation of AEDPA and in contradiction of other circuit decisions that had held that AEDPA deference inheres in the application of *Miller-El*.[167] In my reading, both the court and the Commonwealth had a point, but they were simply arguing at cross purposes. If one interprets footnote nine as applying only to procedural COA considerations, the court is probably correct. As no federal court has yet deferred to the state courts' analysis of the merits of the habeas petition, it is reasonable to suggest that, under *Slack*, the state courts' denial of the petitioner's constitutional claims should not be taken into account in determining whether jurists of reason could debate if the petitioner has stated a valid claim of the denial of a constitutional right.[168] The Commonwealth, by contrast, seized upon the court's overreliance on *Miller-El*, and there *it* had a point. In determining whether "jurists of reason could disagree with the district court's resolution of [the petitioner's] constitutional claims"[169]—the relevant consideration in a merits COA analysis—incorporation of AEDPA deference is proper because it bears on the district court's actual analysis of the petitioner's constitutional claims. In sum, the court's use of *Miller-El* as

n.2. As an unpublished decision, *Gerber* does not bind the circuit, *see* Chehazeh v. Att'y Gen., 666 F.3d 118, 127 n.12 (3d Cir. 2012), and may not reliably signal the court's thinking about *Pabon*. A recent precedential case, *United States v. Thomas*, 713 F.3d 165 (3d Cir. 2013), granted a procedural COA (assuming, without deciding, that one was needed) without discussing the merits of the underlying matter at all. *Id.* at 166 n.1. Indeed, discussion of the merits in *Thomas* would have been complicated by the absence of an actual filed 28 U.S.C. § 2255 motion.

166. *Pabon*, 654 F.3d at 392 n.9.

167. Petition for Writ of Certiorari at 16–21, Kerestes v. Pabon, No. 11-958, 2012 WL 379762, at *16–21 (citing Dockins v. Hines, 374 F.3d 935, 938 (10th Cir. 2004); Medellin v. Dretke, 371 F.3d 270, 275 (5th Cir. 2004)).

168. Of course, this applies only to habeas petitions filed by state prisoners.

169. Miller-El v. Cockrell, 537 U.S. 322, 327 (2003) (citing *Slack*, 529 U.S. at 484).

its primary standard instead of *Slack*, and its inconsistent distinction between the merits-based and procedure-based COA standards, prompted the Commonwealth to argue that the court had advocated something far more novel than it had intended. While it is perilous to read much into a denial of certiorari,[170] I would not be surprised to learn that the Supreme Court untangled this skein in declining to take the case.[171]

Pabon's impact is not yet known, both within the Third Circuit and without. It may yet be marshaled to impose a stricter standard on attempts to obtain COAs on procedural claims; or, on the other hand, it may have very little impact in that realm at all.[172] But regardless of its ultimate effect, *Pabon* speaks to the continued uncertainty over the proper standard to use in evaluating the merits of procedural COA appeals, and the danger of mixing the *Slack* and *Miller-El* approaches to two very distinct questions.

VII.
THE BEST APPROACH IS THE MOST PERMISSIVE

In my view, those circuits that follow the permissive approach—taking only the briefest look at the merits of a petition before granting a COA on procedural grounds—are most accurately following the letter of *Slack* and the spirit of *Miller-El*. It is my contention that an approach that focuses on claim cognizability would be ideal.[173] It should function like this: when presented with a petition that was denied on either erroneous or questionable procedural grounds, a court of appeals should check to see whether the petition facially alleges the deprivation of a constitutional right. If the petition does so, the court should grant a COA, identifying the constitutional ground in its order or other decree to satisfy 28 U.S.C. § 2253(c)(3). Having allowed the appeal to proceed, a court now has several options. Summary action, either in favor of the government or the petitioner, can be employed if the procedural

170. *See* Hughes Tool Co. v. Trans World Airlines, Inc., 409 U.S. 363, 365 n.1 (1973); Pamela S. Karlan, *Some Thoughts on Autonomy and Equality in Relation to Justice Blackmun*, 26 HASTINGS CONST. L.Q. 59, 65 n.32 (1998).

171. I note, too, that the Third Circuit, with the same judge authoring, later acknowledged the ambiguities contained in its earlier statement of certainty in footnote nine. *See* Tomlin v. Britton, 448 F. App'x 224, 227 n.3 (3d Cir. 2011).

172. Of course, *Pabon* also involved a lengthy discussion of equitable tolling, for which the case has been repeatedly cited. *See, e.g.*, Munchinski v. Wilson, 694 F.3d 308, 329 (3d Cir. 2012).

173. A "quick-look" standard, which might serve to weed out certain petitions that advance cognizable but facially meritless claims, would be an acceptable alternative.

grounds upon which the petition was denied below are clearly erroneous *or* if the petition is clearly meritless. As I argue below, this approach would not consume significantly more resources than a "strict" COA standard. Otherwise, the case can, as always, proceed to briefing, appointment of counsel, and so on, which may be appropriate if the procedural ground is ambiguous.

There are several clear reasons, both legal and prudential, why this standard should be adopted by the circuit courts and affirmed (if need be) by the Supreme Court. First, and most importantly, Justice Kennedy's opinion in *Slack* uses language that is indelibly associated with a minimal showing: stating a claim, which traditionally refers to articulating the elements, facts, and actors required to plead a cognizable allegation. True, Kennedy modified this language by inserting "valid," into "stating a claim," and "stating a claim" has itself become a tougher burden in recent years (*Ashcroft v. Iqbal*,[174] anyone?). But interpreting "valid claim" to mean "winning claim" is clearly not right, as language in both *Slack* and *Miller-El* demonstrates.[175] Furthermore, while Justice Kennedy does not anchor his "valid claim" construction on any prior precedent, the Supreme Court *has* discussed "valid claims" in the context of cognizability. In *Estelle v. Gamble*, a seminal case about Eighth Amendment medical-care claims by prisoners, the Court observed that "a complaint that a physician has been negligent in diagnosing or treating a medical condition does not state a valid claim of medical mistreatment."[176] Having reemphasized this principle, the Court proceeded to "consider whether respondent's complaint states a *cognizable* § 1983 claim."[177] Although *Estelle* was decided by a different Court and a different Justice—and on a different (if tangentially related) subject—the case still shows that "valid" can be a

174. 556 U.S. 662 (2009).

175. *See* Miller-El v. Cockrell, 537 U.S. 322, 338 (2003) ("Indeed, a claim can be debatable even though every jurist of reason might agree, after the COA has been granted and the case has received full consideration, that petitioner will not prevail."); Slack v. McDaniel, 529 U.S. 473, 484–85 (2000) (articulating the "jurists of reason" standard, while emphasizing that the COA is a "threshold" test).

176. 429 U.S. 97, 106 (1976).

177. *Id.*; *see also* Newport v. Fact Concerts, Inc., 453 U.S. 247, 253 n.7 (1981) ("Both the District Court and the Court of Appeals determined that respondent had stated valid claims for relief under federal and state law . . . ."); Urie v. Thompson, 337 U.S. 163, 196 (1949) (referring to a "valid claim" as one cognizable under the relevant statute). I do not mean to suggest that this is the only reasonable way of interpreting "valid claim," but rather that doing so is not outlandish and has support in earlier Supreme Court case law.

part of a cognizable/noncognizable continuum rather than a successful/unsuccessful one.

None of the alternative approaches is universally sustainable across the wide variety of procedural postures that can give rise to an appeal. A stricter approach, like the "sliding scale" one favored by the Fifth Circuit, fails when the record has not been developed to a sufficient extent to allow for the evaluation of the merits of the petition at all. In the aforementioned Fifth Circuit test, a petition without much factual development is evaluated at a "quick look" level, whereas one with more factual development is evaluated at a stricter level. In effect, then, the Fifth Circuit is using two different tests, each of which can lead to a COA. Because a cognizability/quick-look approach is still used in these more demanding circuits when circumstances warrant, those circuits are, in effect, acknowledging that a minimal showing is all that is actually required by *Slack*, and that they are interposing a greater requirement upon the appellant.

Second, and following from the above, the fact that multiple tests are percolating throughout the circuits hides a starker reality: because most COA orders are not available to or indexed by legal publishers, courts can, in fact, do whatever they please on a case-by-case basis. After all, COA orders, like most summary dispositions, need not contain justification or reasoning. While the Introduction to this Article decried the possible petitioner-disfavoring ramifications of this state of affairs, unclear and inconsistent standards can also work against the government. Take, for example, a Fifth Circuit case, *Bailey v. Cain*.[178] As the appellate court's opinion reveals, the district court raised the issue of timeliness sua sponte, and a COA was granted by the court of appeals "on the issues of whether the district court should have raised the limitations issue sua sponte and, if so, whether the dismissal was proper."[179] The order that granted a COA, however, makes no mention of the Fifth Circuit's balancing test; indeed, it fails to reference the "merits" of Bailey's petition at all, in contravention of both the balancing test and 28 U.S.C. § 2253(c)(2).[180] Circuit appeals always contain a certain amount of variability, and an appellant's fortunes can be determined by the particular panel he draws. Yet, as shown by *Bailey*, the differences between a precedential opinion articulating a COA standard and an interstitial order (that is, for all intents and purpose, unsearchable) can be striking.

178. 609 F.3d 763 (5th Cir. 2010).

179. *Id.* at 765.

180. *See* Bailey v. Cain, No. 08-31222 (5th Cir. Aug. 13, 2009).

Third, any test that relies on a searching merits analysis will fail on those occasions where the record is not developed to a point where real evaluation of a petitioner's claims is possible.[181] In these circumstances, the kind of in-depth analysis that is seen in cases like *Pabon* would be difficult, although the appellate court could sit as an ersatz district court and conduct a detailed factual inquiry. Furthermore, in extreme cases, a "category 1" procedural dismissal—one unrelated to either habeas or AEDPA, such as failure to follow court rules, pay the filing fee, and so on—can present the appellate court with the absence of a final, filed habeas petition.[182]

Fourth, at least in the context of actions brought by prisoners acting pro se, courts are obligated to construe liberally the filings of the unrepresented.[183] This requirement is constrained and expanded by competing impulses. On the one hand, a court does not assume the role of an advocate;[184] on the other, courts may sometimes decide to grant relief based on an entirely different legal theory than the one identified by the pro se petitioner.[185] A risk inherent in denying procedural COAs based on marginal showings by a petitioner is that, simply, there may be more in the record that could plausibly suggest an alternative outcome if the petition were

181. *See, e.g.*, United States v. Thomas, 713 F.3d 165 (3d Cir. 2013).

182. *See, e.g.*, United States v. Saro, 252 F.3d 449, 453 (D.C. Cir. 2001) (addressing a situation in which "the underlying § 2255 motion was never filed because the district court denied leave to file," but declining to reach the central question—the standard of review).

It may be the case that the most extreme situations under category 1 need not always require a COA. The subject could likely support an article by itself, but, in brief: In *Harbison v. Bell*, 556 U.S. 180 (2009), the Supreme Court distinguished between appeals of "final" orders, which require a COA, and appeals of orders that are nonfinal—in that case, an order "den[ying] a motion to enlarge the authority of appointed counsel"—which do not require a COA. *Id.* at 183. At least one court has tentatively applied *Harbison* to situations in which a petition is denied in such a way that prevents it from being "filed" in the first place—in other words, a nonfinal predisposition order that is so defective as to be immediately appealable. *See, e.g.*, Pero v. Duffy, No. 11-4532 (3d Cir. Mar. 7, 2012); Jones v. Pa. Bd. of Prob. & Parole, No. 11-3461 (3d Cir. Dec. 5, 2011).

183. *See, e.g.*, Meador v. Branson, 688 F.3d 433, 436 (8th Cir. 2012); Figuereo-Sanchez v. United States, 678 F.3d 1203, 1207 (11th Cir. 2012); Koons v. United States, 639 F.3d 348, 353 n.2 (7th Cir. 2011).

184. *See* Mala v. Crown Bay Marina, Inc., 704 F.3d 239, 244 (3d Cir. 2013); United States v. Viera, 674 F.3d 1214, 1216 n.1 (10th Cir. 2012) (citing Yang v. Archuleta, 525 F.3d 925, 927 n.1 (10th Cir. 2008)); *see also* Robert Bacharach & Lyn Entzeroth, *Judicial Advocacy in Pro Se Litigation: A Return to Neutrality*, 42 IND. L. REV. 19, 42 (2009).

185. *See, e.g.*, *Mala*, 704 F.3d at 245; *Meador*, 688 F.3d at 436. *But see* Castro v. United States, 540 U.S. 375, 385 (2003) (Scalia, J., concurring) (warning against moving from liberal construction to recharacterization).

to be liberally construed. Whether an appellate court chooses to vacate or affirm, the act of having to justify its decision at the merits stage could lead to a fairer examination of the pro se petition, especially in contexts where the district court's interpretation was unnecessarily narrow.

Fifth, a lenient standard maximizes the value of adopting a single standard for this kind of COA determination. As discussed above, the various circuit courts have promulgated markedly different rules with regard to briefing, presentation of the issues, and release of their opinions and orders to legal publishers. Some of these approaches—most notably, the Tenth Circuit's—can seem, at times, like a combination COA/merits determination. Using a single standard for procedural COA applications would ensure that, regardless of the reviewing court's ultimate output, no COA denials—whether in lengthier opinion form or shorter order form—would be "stealth" analyses of the merits of a petition. This benefit would be in addition to the intrinsic value contained in a uniform implementation of the rule, such as easier cross-circuit applicability and streamlined procedures for petitioners.[186] In other words, even if a court issued a nondescript order declining to issue a COA because the petitioner had failed to meet the *Slack* standard, a reviewing court or an outside observer would be assured that: (a) the district court's procedural determination was not debatable among jurists of reason; or (b) the petitioner had failed to allege a cognizable denial of a constitutional right. All parties to the system would be better informed, issues surrounding successive petitions would be clarified, and counsel for the petitioner would be better able to tailor her arguments if attacking a COA denial at a rehearing or certiorari stage.

The above leads smoothly into a discussion of several prudential rationales for adopting a more lenient standard across the circuits, starting with the one that is most central to my thesis: as AEDPA aimed, in part, to give prisoners a single attempt to collaterally challenge their convictions and sentences in federal court, those petitioners who successfully navigate the labyrinth that AEDPA erects should be afforded a decision on the merits of their petitions.[187] By "decision on the merits," I mean to suggest a *real*

186. *Cf.* Daniel S. Tomson, Note, *Rule 58's Dirty Little Secret: The Problematic Lack of Uniform Enforcement of Federal Rule of Civil Procedure 58 Within the Federal Court System*, 36 VAL. U. L. REV. 767, 816–17 (2002) (discussing the importance of uniformity of federal rules in connection with FED. R. CIV. P. 58).

187. *See* Uhrig, *supra* note 10, at 1222–23 (discussing the complex field that petitioners, often pro se, are expected to navigate).

decision on the merits, not one shrouded in the procedural uncertainty of a COA denial that suggests a failure to satisfy an ambiguous gatekeeping standard. Those prisoners who lose on the merits before the district court have been afforded one full look at the merits of their petition; thus, their arguments are subject to a higher standard of scrutiny before being granted a COA. Those petitioners who lose on a procedural ground, by contrast, have not been afforded an equal courtesy. Whether the appellate court acts via a summary affirmance on alternative grounds or remands to the district court for further consideration, a habeas petitioner whose filing is denied on procedural grounds that are erroneous or arguable has, under this model, a better chance of receiving a decision on the merits at some stage of the litigation. And while not true in all contexts, courts do, on occasion, indulge a "strong presumption in favor of deciding cases on the merits."[188]

I see many benefits and few clear pitfalls in taking the above approach. First and foremost, removing a stealth merits analysis from the COA stage renders the appellate court's decision clearer to other courts and to the parties, especially in those circuits whose COA dispositions are traditionally neither detailed nor available to legal publishers. When a petition for certiorari is filed to the Supreme Court from a denial of a COA,[189] the Court does not possess insight into what motivated the lower court's decision to deny a COA; all it has is the order. In procedural cases, shifting matters to the merits stage would give the high court a greater ability to determine what exactly motivated the appellate court's disposition.

Further, the petitioner himself, especially in marginal cases, would understand the grounds upon which his petition was denied, which might have differed markedly from the grounds relied upon by the district court. A petitioner whose collateral attack on his conviction was denied on spurious procedural grounds, and whose COA application was nevertheless denied based on his "failure to state a valid claim" of the denial of a constitutional right, could reasonably take from these decisions a feeling that his petition was denied on grounds that were grossly unfair. This, in turn, could lead to exactly the sort of serial filing of petitions (both in good faith and bad) frowned upon by AEDPA, in an attempt to reair griev-

188. Garcia-Perez v. Hosp. Metropolitano, 597 F.3d 6, 7 (1st Cir. 2010) (quoting Malot v. Dorado Beach Cottage Assocs., 478 F.3d 40, 43 (1st Cir. 2007)) (discussing dismissals under FED. R. CIV. P. 41(b)).

189. *See generally* Hohn v. United States, 524 U.S. 236 (1998).

ances that were not clearly addressed in earlier attempts.[190] And a court reviewing a subsequent application under the prefiling authorization provisions of 28 U.S.C. § 2244(b) might not have easy access to the reasoning that undergirded the previous decisions.

Take, for example, this simple scenario: a prisoner files a habeas petition of marginal worth that is denied as untimely by the district court based on a miscalculation of the periods of statutory tolling. Without this miscalculation, the petition is timely. In a circuit applying a stricter standard, an appeals court could deny a COA by simply stating that the petitioner failed to "state a valid claim" or make a "substantial showing of the denial of a constitutional right," if the petition had little apparent merit after a cursory analysis. Indeed, boilerplate language citing *Slack* might be sufficient to deny permission to appeal. But the petitioner, who may or may not have a reasonable merits argument, has a *strong* argument that the district court's procedural decision was wrong. He may think, understandably, that the circuit court simply missed the district court's mathematical error, or, more darkly, might wonder whether the same forces that "railroaded [him] . . . to . . . conviction" were keeping him from his putative savior, the "federal habeas judge."[191] He probably knows that, if he wishes to file a second habeas petition, he confronts a daunting task in vaulting AEDPA's restrictions on successive petitions.[192] That bar contains numerous

190. Recall that habeas relief exists in obvious tension with the finality of a conviction and sentence; even before AEDPA, courts were required to balance the power of habeas corpus with its potential for abuse. *See, e.g.*, McCleskey v. Zant, 499 U.S. 467, 490–91 (1991). AEDPA places tremendous restrictions on serial filings, requiring a prisoner to meet the demanding requirements of 28 U.S.C. § 2244(b) (for state prisoners) or 28 U.S.C. § 2255(h) (for federal prisoners) to proceed. But even the least meritorious application to pursue a second or successive petition requires judicial resources to process and deny, a feat that becomes more daunting when the actual disposition of the original petition is less than clear.

191. *See* William N. Eskridge, Jr., *Some Effects of Identity-Based Social Movements on Constitutional Law in the Twentieth Century*, 100 MICH. L. REV. 2062, 2225 (2002).

192. State petitioners cannot pursue second or successive petitions unless they rely on new rules of constitutional law that have been made retroactively applicable or satisfy a "newly discovered facts" standard that requires them to show that "but for constitutional error, no reasonable factfinder would have found the applicant guilty of the underlying offense." 28 U.S.C. § 2244(b) (2006). Federal prisoners must satisfy a similar standard. *See* 28 U.S.C. § 2255(h) (2006); *see also* Uhrig, *supra* note 10, at 1250. Moreover, both groups must receive prefiling authorization from the relevant court of appeals, after which (assuming the application is granted) the district court must undertake its own analysis of the § 2244(b) factors before proceeding to the merits. *See, e.g.*, Case v. Hatch, 731 F.3d 1015, 1026–27 (10th Cir. 2013) (describing the screening process); FHCPP, *supra* note 3, § 28.3; *see also* Burris v. Parke, 116 F.3d 256, 257 (7th Cir. 1997) (referring to the

exceptions, though, some of which are open to creative (if frequently futile) interpretation by litigants and attorneys of both the jailhouse and licensed variety.[193] Federal prisoners might cast a keen eye upon 28 U.S.C. § 2255(e), for example, which allows for those prisoners who have used up their "one shot" at regular § 2255 relief to pursue a 28 U.S.C. § 2241 habeas corpus petition to "bring a second or successive attack on [their] conviction[s] or sentence[s] under 28 U.S.C. § 2241, without reference to § 2255(h)'s restrictions."[194] Or a prisoner might delve into history, seeking solace in one of the residual writs that, in the modern age, are of limited general applicability.[195] This sort of serial filing of "disguised" habeas petitions (which are usually dismissed as unauthorized filings or as facially meritless[196]) is hardly unique to prisoners who have been the subject of unfair procedural denials—far from it.[197] Yet it is certainly possible that a relaxed approach to granting procedural COAs would lead to a more satisfying resolution on the merits for a number of petitioners, and would indeed simplify

AEDPA successive requirements as "more restrictive" than the pre-AEDPA "abuse of the writ" standard).

193. *See, e.g.*, D.G. Maxted, Panetti v. Quarterman*: Raising the Bar Against Executing the Incompetent*, 4 DUKE J. CONST. L. & PUB. POL'Y SIDEBAR 99, 104–05 (2009), *available at* http://scholarship.law.duke.edu/djclpp_sidebar/28/.

194. Prost v. Anderson, 636 F.3d 578, 584 (10th Cir. 2011), *cert. denied*, 132 S. Ct. 1001 (2012); *see also* Sarah French Russell, *Reluctance to Resentence: Courts, Congress, and Collateral Review*, 91 N.C. L. REV. 79, 116–17 (2012) (discussing the scope of the "savings clause"). Legitimate petitions under § 2241 in this context can only be sought when § 2255 relief is inadequate or ineffective, an extraordinarily narrow category. *See* FHCPP, *supra* note 3, § 41.2(b) n.27. That the exception is narrowly drawn does not prevent petitioners from trying. *See, e.g.*, Adderly v. Zickefoose, 459 F. App'x 73, 75 (3d Cir. 2012) (per curiam).

195. *See, e.g.*, Massey v. United States, 581 F.3d 172, 174 (3d Cir. 2009) (per curiam) (discussing writ of audita querela), *cert. denied*, 130 S. Ct. 2426 (2010); *see also* FHCPP, *supra* note 3, § 41.2(b) (discussing residual remedies validly available to federal prisoners).

196. *See, e.g.*, United States v. Ledesma-Cuesta, 443 F. App'x 685, 686 (3d Cir. 2011) (per curiam).

197. While I have no concrete evidence of this, I suspect that petitioners who do not receive adjudications on the merits are probably more likely to attempt to file serial petitions to relitigate claims that were never squarely addressed by the lower courts. During my time as a staff attorney, the majority of serial filers I encountered (at least, the ones who were not clearly vexatious or suffering from mental illness) fell into this category, raising claims that had been denied on procedural grounds either pre- or post-AEDPA. Undoubtedly, many of these procedural denials were correct, but some likely were arguable; an eventual decision on the merits might have stemmed the flow of serial petitions, or at the very least would have aided the court's analysis of the petitioner's claims when he complained of alleged injustices from trials occurring thirty or more years in the past.

things tremendously for those tribunals asked to rule upon those serial petitions. They would be able to proceed with the certainty and security that the petitioner had been afforded his day in court.

Taking this approach would also allow circuit courts to be more proactive in pointing out procedural errors made by the district courts while allowing for the more rapid development of the relevant jurisprudence. The procedural landscape of habeas after AEDPA is nothing if not wild and wooly, and district courts—especially those with crushing caseloads—can make the occasional, understandable mistake, in addition to touching on issues that, by dint of their complexity, might simply be "arguable" by jurists of reason. Even in this late hour, the basic components of AEDPA's procedural toolkit are still being tweaked. To cite just one example, the Supreme Court recently ruled that a sentence-reduction motion under Rhode Island law is "an application for 'collateral review' that triggers AEDPA's tolling provision."[198] Certainly, this new rule probably affected some cases that were winding their way through the system, and will likely affect many more to come. Yet in applying the rule outside of Rhode Island, district and circuit courts will likely be confronted by more marginal than meritorious petitions. With a strict COA gatekeeping standard, those marginal petitions, even if denied on incorrect procedural grounds under the new rule, may yet be rejected at the COA stage. But with a more lenient standard, an appellate court might earlier be able to reach the relevant procedural question, even if its opinion was unpublished or otherwise lacked binding precedential value. This in turn would lead to a swifter growth of cases discussing the new rule.

The central argument against a lenient COA standard is clear: such a standard will require courts to expend more time, money, and energy on matters of uncertain merit. However, some of these concerns, while well intentioned, are not borne out by the reality on the ground. As a preliminary matter, the connection between granting COAs and additional resources is largely one of the courts' own creation. If COAs are to only be granted in circumstances of likely merit, such that they imply the presumption of appointment of counsel, extended briefing schedules, and so on, then of course COAs will be associated with delay and other administrative or financial unpleasantness. Severing this connection, and recharacterizing COAs as modest gatekeepers instead of cruel ones, will do much to correct this misconception. Those same judiciary employ-

198. Wall v. Kholi, 131 S. Ct. 1278, 1287 (2011).

ees,[199] who now bear the responsibility of screening COA applications, will undoubtedly be able to spot those COA candidates that, despite being allowed to proceed to the merits stage, should not receive the full accoutrement of services that are granted petitions of potential merit.

Nor will shifting more appeals to the merits stage "frustrate[ ] the 'E' in AEDPA."[200] Courts will not have to invest more intellectual firepower in disposing of these cases. Behind each order denying a COA because a petitioner has failed to meet the merits prong of the *Slack* procedural standard must lie an analysis of the merits, one that could be easily transformed into an opinion. Moreover, courts already have mechanisms in place to expedite appeals of little to no merit, or those with clear errors that deserve quick remand to the district court. Summary affirmance, for example, can allow a court to dispose of an appeal sua sponte without requiring full briefing if no substantial question is presented.[201] Summary action can operate against the petitioner or in his favor. For example, if the appellate court wishes to summarily vacate and remand to the district court to reach the merits of a petition in the first instance, it can certainly do so.[202]

A case from the Third Circuit demonstrates a simple procedural approach to the summary action question in the habeas context. In *Buxton v. Pennsylvania*, after the district court had (potentially erroneously) dismissed a petition as unexhausted without reaching whether its claims may have been defaulted, the court of appeals issued an order to show cause requesting that the parties "address whether the claims raised in appellant's petition for a writ of habeas corpus are either procedurally defaulted . . . or barred under the AEDPA statute of limitations."[203] After the response period had ex-

199. Staff attorneys, for example. *See, e.g.*, Roxie Bacon, *Retain the Ninth Circuit: An Efficient and Excellent Bench*, ARIZ. ATT'Y, Sept. 2005, at 35, 44, *available at* http://www.myazbar.org/AZAttorney/PDF_Articles/0905Con3.pdf; Marin K. Levy, *The Mechanics of Federal Appeals: Uniformity and Case Management in the Circuit Courts*, 61 DUKE L.J. 315, 335–39 (2011) (addressing case management among different circuits); David R. Stras & Shaun M. Pettigrew, *The Rising Caseload in the Fourth Circuit: A Statistical and Institutional Analysis*, 61 S.C. L. REV. 421, 443 (2010).

200. Cutler, *supra* note 10, at 357.

201. *See, e.g.*, Murray v. Bledsoe, 650 F.3d 246, 247–48 (3d Cir. 2011) (per curiam) (citing 3D CIR. R. 27.4; 3D CIR. I.O.P. 10.6 (2010)); McKenna v. Powell, 631 F.3d 581, 582 (1st Cir. 2011) (per curiam) (citing 1ST CIR. R. 27.0(c)).

202. *See, e.g.*, Wyatt v. Thaler, 395 F. App'x 113, 114–15 (5th Cir. 2010) (per curiam) (vacating and remanding the district court's dismissal of defendant's habeas petition because he had been given insufficient procedural fairness).

203. No. 10-1203 (3d Cir. Mar. 31, 2010).

pired, the court granted a COA, vacated, and remanded in a single step.[204]

The filtering mechanism of the COA requirement does not complicate the process of taking summary action in those circumstances when an appellate court wishes to summarily affirm adverse to the interests of the petitioner—thus potentially reaching the merits in the first instance on appeal—rather than summarily vacating due to a problematic procedural outcome below. Ordinarily, when granting a COA on a procedural question, the merits of the case are not separately "certified" for appeal, although (as discussed further *infra*) a proper COA order should "indicate" at least one issue that passes the *Slack* test. Yet as explained by the Federal Rules of Appellate Procedure (if, curiously, not by the statutory text), "[a] certificate of appealability is not required when a state or its representative or the United States or its representative appeals."[205] Mindful of the rule that Congress generally does not mean to depart from standard federal procedure unless it states so explicitly,[206] and embracing the axiom that an appellate court can "affirm a lower court's ruling on any grounds adequately supported by the record, even grounds not relied upon by the district court,"[207] several courts have held that the scope of the COA does not limit the grounds that the state may argue to defend the judgment. These courts, therefore, have preserved the discretion of the appellate court to affirm on any basis supported by the record.[208] Thus, appellate courts may easily deny a petition as meritless, notwithstanding that the district court had declined to reach those issues and the COA order declined to separately certify them for appeal.

In sum, there are several good reasons for adopting, in this context, the more lenient "quick-look" or "facially cognizable" COA standards for procedural habeas appeals. These positive legal and prudential considerations meet with few downsides. Allowing more procedural COAs to be granted, and therefore disposing of more petitions at the merits stage, will have little effect on a court's use of

204. *See generally* Buxton v. Pennsylvania, 398 F. App'x 704, 706 n.2, 708 (3d Cir. 2010) (per curiam).

205. FED. R. APP. P. 22(b)(3).

206. *See* Rhodes v. Robinson, 621 F.3d 1002, 1007 (9th Cir. 2010) (quoting Jones v. Bock, 549 U.S. 199, 216 (2007)).

207. Elwell v. Byers, 699 F.3d 1208, 1213 (10th Cir. 2012); *see also* Smith v. Phillips, 455 U.S. 209, 215 n.6 (1982) ("Respondent may, of course, defend the judgment below on any ground which the law and the record permit, provided the asserted ground would not expand the relief which has been granted.").

208. *See, e.g.*, Woodward v. Williams, 263 F.3d 1135, 1139 n.2 (10th Cir. 2001); Garcia v. Lewis, 188 F.3d 71, 75 n.2 (2d Cir. 1999).

resources. The manpower that has gone into deciding that a petition is marginal or meritless will already have been spent, no attorney need be appointed to represent a pro se petitioner, and a court's ability to take summary action in whatever way it pleases is unaffected by AEDPA. Further, courts will be hewing closer to the language of *Slack* and the spirit of *Miller-El*, while resolving a greater number of petitions on the merits and "out in the open." If made readily available, those opinions would hasten the further development and refinement of the procedural habeas landscape post-AEDPA.

VIII.
WHAT DOES A "COMPLIANT" COA ORDER LOOK LIKE IN 2013?

One further question merits brief discussion: what does a compliant procedural COA order look like after fifteen years of AEDPA jurisprudence? Recently, in *Gonzalez v. Thaler*,[209] the Supreme Court held that a failure to comply with the exact statutory language of AEDPA and "indicate" substantive constitutional issues does not deprive an appellate court of jurisdiction.[210] Nevertheless, following the language of § 2253(c) is still "mandatory,"[211] although defects can be easily remedied when "a party timely raises the COA's failure to indicate a constitutional issue."[212]

In order to satisfy the Court's *Gonzalez* ruling, then, all procedural COA orders should endeavor to identify *a* claim that passes the low-level scrutiny advocated for above. For example, a COA order might grant a certificate of appealability on a procedural default claim, while identifying an ineffective assistance of counsel argument as "stating a valid claim of the denial of a constitutional right." Citation to 28 U.S.C. § 2253(c)(2) is unnecessary, especially if the authoring panel is of the opinion that the relevant language of *Slack* was intended to supplant the literal statutory terms. The court need not take from its "indication" that it has separately certified the merits of the claim for appeal.

CONCLUSION: THE MODEST PROPOSAL

In 2007, District Judge James Robertson decried the "procedural obstacles that confront prisoners seeking review on the merits of

209. 132 S. Ct. 641 (2012).
210. *Id.* at 649.
211. *Id.* at 651.
212. *Id.*

their [habeas corpus] petitions."[213] According to Judge Robertson, "virtually every habeas petition and § 2255 application on [his] docket has to be dismissed" for one of a number of sundry procedural reasons.[214] While conceding that "[m]ost post-conviction claims do lack merit," he nevertheless "suspect[ed] that [judges] expend a lot more energy crafting careful opinions explaining why we cannot reach the merits than we would if we simply ruled on the issues that petitioners ask us to decide."[215] Judge Reinhardt of the Ninth Circuit has also assailed the "impenetrable procedural rules designed to make habeas relief unavailable to all but the most fortunate."[216] Sympathy for the plight of the convicted and the pro se is hardly universal, of course, and given the seventeen years since AEDPA's enactment, prisoners should arguably be on notice regarding the statute's procedural pitfalls.[217] Yet the statistics are striking. Even before AEDPA's passage, a substantial percentage of federal habeas petitions were dismissed on procedural grounds.[218] A recent study suggests that a majority of noncapital petitions filed by state petitioners continues to be disposed of on procedural grounds without a discussion of the merits.[219] And, of course, "virtually all habeas

213. James Roberston, *2007 James McCormick Mitchell Lecture:* Quo Vadis, *Habeas Corpus?*, 55 BUFF. L. REV. 1063, 1083 (2008).

214. *Id.* at 1084.

215. *Id.*

216. Leavitt v. Arave, 682 F.3d 1138, 1141 (9th Cir. 2012) (Reinhardt, J., concurring).

217. *See, e.g.*, Cutler, *supra* note 10, at 357. Indeed, the habeas petition forms themselves contain warnings about potential procedural issues. *See, e.g.*, PETITION FOR RELIEF FROM A CONVICTION OR SENTENCE BY A PERSON IN STATE CUSTODY, U.S. COURTS 6 (June 2013), *available at* http://www.uscourts.gov/uscourts/FormsAndFees/Forms/AO241.pdf ("CAUTION: To proceed in the federal court, you must ordinarily first exhaust (use up) your available state-court remedies . . . ."). Of course, these warnings are of little use to those who do not glance at the form before errors have been committed in state court proceedings, or who do so after the time to file has lapsed.

218. *See* PAUL H. ROBINSON, AN EMPIRICAL STUDY OF FEDERAL HABEAS CORPUS REVIEW OF STATE COURT JUDGMENTS 4(b), 13 (1979) (noting that 55% of petitions challenging state court convictions and sentences were dismissed because of procedural defects; at the time, 60% of those dismissals were due to failure to exhaust, and 15.1% failed to state a cognizable claim). Fifteen years later, that number was substantially unchanged. *See* Jonah Wexler, Note, *Fair Presentation and Exhaustion: The Search for Identical Standards*, 31 CARDOZO L. REV. 581, 593 n.74 (2009) (citing ROGER A. HANSON & HENRY W.K. DALEY, BUREAU OF JUSTICE STATISTICS, FEDERAL HABEAS CORPUS REVIEW: CHALLENGING STATE COURT CRIMINAL CONVICTIONS 17–18 (1995), *available at* http://www.bjs.gov/content/pub/pdf/FHCRCSCC.PDF).

219. *See* NANCY J. KING ET AL., FINAL TECHNICAL REPORT: HABEAS LITIGATION IN U.S. DISTRICT COURTS: AN EMPIRICAL STUDY OF HABEAS CORPUS CASES FILED BY STATE PRISONERS UNDER THE ANTITERRORISM AND EFFECTIVE DEATH PENALTY ACT OF

petitions are brought pro se by prisoners"[220] who are poorly equipped to navigate arcane procedural requirements and for whom warnings at stage ten are of little help in preventing procedural mistakes, errors, and oversights at stage one.

In enacting AEDPA, Congress aimed to streamline a process long thought to be unnecessarily protracted—especially in the capital context—while emphasizing the values of finality and comity. Agree or not with Congress's decision, the concerns it identified were real, and the historical circumstances that led to AEDPA's swift passage were certainly galvanizing. Yet the procedural additions that AEDPA grafted on to an already complex field of jurisprudence may not have helped the goals of expediency and regularity, and, undoubtedly, the general poor quality of the law's drafting meant that courts would have to expend more effort, not less, on navigating the hidden shoals of the new order. At the end of the first stage of this navigation, the courts are instructed to apply an additional, opaque standard—the COA requirement—to determine whether an appeal can proceed. As this Article has hopefully shown, the combination of these new procedural requirements and a strict procedural COA standard (or one, at the very least, applied inconsistently across circuits) serves to frustrate the goals of Congress, the courts, habeas practitioners, and litigants.

Short of eliminating the COA requirement entirely, which would contradict over a hundred years of gatekeeping habeas jurisprudence, the best solution to AEDPA's procedural quagmire is to ensure fairness to pro se litigants, while simultaneously honoring the impulses of comity and finality. Adopting a lenient procedural COA standard accomplishes this goal, while imposing few additional demands on the resources of the federal judiciary. COA jurisprudence may not be as exciting as the big issues in habeas corpus jurisprudence, but it remains a daily component of the dimmer parts of the criminal justice system. My aim is to rescue some portion from the shadows.

1996 45 (2007), *available at* http://www.ncjrs.gov/pdffiles1/nij/grants/219559.pdf; *see also* Erwin Chemerinsky, *Losing Faith: The Supreme Court and the Abandonment of the Adjudicatory Process*, 60 HASTINGS L.J. 1129, 1133 (2009).

220. Chemerinsky, *supra* note 218, at 1134.

THE LAW OF INTENDED CONSEQUENCES: DID THE AMERICANS WITH DISABILITIES ACT AMENDMENTS ACT MAKE IT EASIER FOR CANCER SURVIVORS TO PROVE DISABILITY STATUS?

*BARBARA HOFFMAN**

> Parent of Student (seeking to fire his daughter's coach): "Nobody is talking about the elephant in the room. The fact is you have cancer. How are you going to coach when you are in and out of hospitals? I mean, who knows what else is gonna go wrong with you this year. . . ."
>
> Cathy Jamison (teacher and coach who has melanoma): "If you try to fire the lady with cancer, you better hire a damn good lawyer."[1]

ABSTRACT

When Congress passed the Americans with Disabilities Act of 1990 ("ADA"), cancer survivors had reason to believe that the Act prohibited cancer-based employment discrimination. Federal courts, however, failed to apply the ADA to cancer-based discrimi-

* Legal Research and Writing Faculty, Rutgers Law School—Newark; Founding Chair, National Coalition for Cancer Survivorship. I am very grateful to: (i) Andy Gimigliano for his generous, detailed, and insightful suggestions; one of the most rewarding aspects of teaching is learning from a gifted student; (ii) Elise Smith for her highly professional research and practical comments; and (iii) Jennifer Falkenstern for her thorough research and dedication to cancer advocacy.

Additional information about legal services for cancer survivors is available from the National Cancer Legal Services Network, a coalition of legal services providers and cancer advocacy organizations that provides resources to improve access to free legal services programs for cancer survivors:

National Cancer Legal Services Network
New York Legal Assistance Group
7 Hanover Square, 18th Floor
New York, NY 10004
www.nclsn.org

1. *The Big C: The Little C* (Showtime television broadcast Aug. 1, 2011). Cathy Jamison is the protagonist of the Showtime series *The Big C*, a fictional television series that chronicles the personal and work life of a woman who has melanoma. The author is grateful to Darlene Hunt, creator and producer of *The Big C*, for dramatizing the employment concerns of cancer survivors.

nation consistently and denied standing to many plaintiffs. Although some courts recognized that cancer was a disability covered by the ADA, other courts precluded relief for cancer survivors by holding that they did not have a disability as defined by the ADA. In response, Congress attempted to expand employees' access to courts with the Americans with Disabilities Act Amendments Act ("ADAAA").

The first part of this Article reviews cases filed under Title I of the ADA by employees who claimed they were discriminated against, in part, because of their cancer history. The discussion focuses on cases in which courts considered whether a cancer survivor was a "qualified individual with a disability" under Title I at the summary judgment stage. The second half of the Article compares these results to plaintiffs' outcomes after the passage of the ADAAA. It concludes that the ADAAA has fulfilled its legislative goal in that it has significantly improved the ability of cancer survivors to prove disability status under the ADA. The Article concludes by providing suggestions for plaintiffs' attorneys to preserve Title I claims.

Introduction ..... 845 R
I. Cancer Survivors and Work ..... 848 R
II. The Americans with Disabilities Act of 1990 ..... 851 R
A. Title I of the ADA ..... 851 R
B. References to Cancer as a Disability in the ADA Regulations and Legislative History ..... 854 R
1. Compliance Manual to the Original EEOC Regulations ..... 854 R
2. Legislative History ..... 856 R
III. How Cancer Survivors Fared in Employment Discrimination Cases Under the ADA ..... 856 R
A. Judgments for Employers ..... 856 R
1. Plaintiffs Did Not Have a Disability Because They Were Not Substantially Limited in the Major Life Activity of Working ..... 857 R
2. Plaintiffs Did Not Have a Disability Because They Could Not Establish a Record of a Disability ..... 863 R
3. Plaintiffs Did Not Have a Disability Because Their Employers Did Not Regard Their Cancer as Disabling ..... 865 R
B. Judgment for Employees ..... 867 R
1. Major Life Activity Other Than Work ..... 867 R

2. Record of Having an Impairment That Substantially Limits a Major Life Activity ..... 869 R
3. Regarded as Having an Impairment That Substantially Limits a Major Life Activity ..... 871 R
IV. How Cancer Survivors Fared in Employment Discrimination Cases Under the ADA Amendments Act ..... 873 R
A. Congress Attempts to Plug Holes in the ADA with the Americans with Disabilities Act Amendments Act ..... 873 R
B. The EEOC Regulations to Title I of the ADAAA . 877 R
1. Definition of an Individual with a Disability .. 879 R
2. Impairments in Remission ..... 879 R
3. Record of a Substantially Limiting Impairment..... 880 R
4. Regarded as Having a Substantially Limiting Impairment..... 881 R
C. Cancer Survivors' Employment Discrimination Cases Under the ADAAA ..... 881 R
V. Lessons for Plaintiffs' Attorneys..... 887 R
Conclusion ..... 890 R

INTRODUCTION

Disabilities rights advocates cheered[2] on July 26, 1990, when President George H.W. Bush signed the Americans with Disabilities Act ("ADA") into law.[3] Cancer advocates in particular heralded the law as an end to a history of employment discrimination[4] against

2. *See e.g.*, 135 Cong. Rec. 19,807 (1989) (statement of Sen. Edward Kennedy) ("The Americans with Disabilities Act . . . has the potential to become one of the great civil rights laws of our generation."); Steven A. Holmes, *Rights Bill for Disabled Is Sent to Bush*, N.Y. TIMES, July 14, 1990, *available at* http://www.nytimes.com/1990/07/14/us/rights-bill-for-disabled-is-sent-to-bush.html ("More than 100 supporters of the measure, some in wheelchairs, gathered in a reception room off the Senate chamber and cheered wildly when a security guard announced the final vote. 'No longer will people with disabilities be second-class citizens,' said Pat Wright, executive director of the Disability Rights Education and Defense Fund, a lobbying and advocacy group.").

3. 42 U.S.C. §§ 12101–12213 (2006).

4. In hearings before the House of Representatives, cancer advocates testified that federal legislation was needed to address employment discrimination against cancer survivors who could work, yet faced disparate treatment. *See Employment Discrimination Against Cancer Victims and the Handicapped: Hearings on H.R. 370 and H.R. 1294 Before the Subcomm. on Empl't Opportunities of the H. Comm. on Educ. and Labor*, 99th Cong. 15 (1985) (statement of Robert J. McKenna, M.D., President, American Cancer Society) (reporting that individual misconceptions and social

cancer survivors.[5] With the passage of the ADA, Congress appeared to create a clear mandate against, and adequate remedies for, employment decisions based on a person's cancer history instead of on his or her individual qualifications.[6] Or so we thought.

From the very beginning, federal courts seemed confused about how to determine when a cancer survivor was a "person with a disability" as defined by the ADA.[7] Some courts recognized that cancer was a disability covered by the ADA and thus focused on whether the plaintiff proved that he or she was qualified for the job in question.[8] Other courts, however, forced plaintiffs into a Catch-22 in dismissing survivors' claims prior to trial.[9] These decisions held that survivors who could work did not have an "impairment that substantially limits . . .[a] major life activity" precisely because they could work, while survivors whose cancer limited their ability to work were not "qualified" to perform the essential functions of

attitudes contributed to employment discrimination against cancer survivors). Dr. McKenna identified three classifications of work-related discrimination: (1) dismissal, demotion, and reduction or elimination of work-related benefits; (2) problems arising from coworkers' attitudes; and (3) problems related to the cancer survivors' attitudes about how they should be perceived by coworkers resulting in alienation and avoidance by others. *Id.* at 19.

5. The term "cancer survivor" is now widely recognized to refer to "anyone with a diagnosis of cancer, whether newly diagnosed or in remission or with recurrence or terminal cancer." Ellen Stovall & Elizabeth Johns Clark, *Survivors as Advocates*, *in* A CANCER SURVIVOR'S ALMANAC: CHARTING YOUR JOURNEY 274, 276 (Barbara Hoffman ed., 2014); *see also* FROM CANCER PATIENT TO CANCER SURVIVOR: LOST IN TRANSITION 23–24 (Maria Hewitt et. al. eds., 2006) [hereinafter LOST IN TRANSITION].

6. The purpose of the Americans with Disabilities Act of 1990 was:

> (1) to provide a clear and comprehensive national mandate for the elimination of discrimination against individuals with disabilities;
> (2) to provide clear, strong, consistent, enforceable standards addressing discrimination against individuals with disabilities;
> (3) to ensure that the Federal Government plays a central role in enforcing the standards established in this chapter on behalf of individuals with disabilities; and
> (4) to invoke the sweep of congressional authority, including the power to enforce the fourteenth amendment and to regulate commerce, in order to address the major areas of discrimination faced day-to-day by people with disabilities.

42 U.S.C. § 12101(b) (2006).

7. *See* discussion *infra* Part III.

8. *See* Barbara Hoffman, *Between a Disability and a Hard Place: The Cancer Survivors' Catch-22 of Proving Disability Status Under the Americans with Disabilities Act*, 59 MD. L. REV. 352, 394–407 (2000) [hereinafter Hoffman, *The Cancer Survivors' Catch-22*].

9. *See id.* at 353.

their jobs.[10] The United States Supreme Court further diluted the ADA by rejecting the Equal Employment Opportunities Commission ("EEOC") regulations that determined which employees stated a claim under the ADA.[11] As a result, the very plaintiffs the ADA was designed to protect found their claims cut short by motions to dismiss and motions for summary judgment.[12] Cancer survivors and other Title I claimants faced long odds to secure a trial on the merits.[13]

The ADA, a historical civil rights law, was broken. The National Council on Disability voiced the concern of the disability rights community that Congress needed to restore employees' access to

10. *Id.*

11. The Court rejected the EEOC regulations that explicitly stated that the "determination of whether an individual is substantially limited in a major life activity must be made on a case-by-case basis, without regard to mitigating measures such as medicines, or assistive or prosthetic devices." Sutton v. United Air Lines, Inc., 527 U.S. 471, 480, 482 (1999) (quoting 29 C.F.R. § 1630, app. §1630.2(j) (1998)). For a further discussion of *Sutton* and *Murphy v. United Parcel Service, Inc.*, 527 U.S. 516 (1999), see Hoffman, *The Cancer Survivors' Catch-22*, *supra* note 8, at 420–26.

12. *See* Ruth Colker, *The Americans with Disabilities Act: A Windfall for Defendants*, 34 HARV. C.R.-C.L. L. REV. 99, 108, 125–26 (1999) ("My investigation of the cases in which courts of appeals have affirmed summary judgment decisions in favor of defendant-employers suggests that courts may be too quick to take cases from juries as well as too willing to render judgments in favor of defendants in ADA cases."); Louis S. Rulli & Jason A. Leckerman, *Unfinished Business: The Fading Promise of ADA Enforcement in the Federal Courts Under Title I and Its Impact on the Poor*, 8 J. GENDER RACE & JUST. 595, 616 (2005) (arguing that federal court hostility towards plaintiffs' Title claims are driving lawyers away from filing cases); Melanie D. Winegar, Note, *Big Talk, Broken Promises: How Title I of the Americans with Disabilities Act Failed Disabled Workers*, 34 HOFSTRA L. REV. 1267, 1317 (2006) (concluding that the promise of the ADA to provide remedies for individuals with disabilities "has largely proven untrue").

13. *See* Hoffman, *The Cancer Survivors' Catch-22*, *supra* note 8, at 376 n.121; Jane Byeff Korn, *Cancer and the ADA: Rethinking Disability*, 74 S. CAL. L. REV. 399, 415–16 (2001) (reporting that Title I cases are often dismissed when a question is raised regarding the claimant's inclusion in the protected class of "disability"). *But see* Sharona Hoffman, *Setting the Matter: Does Title I of the ADA Work?*, 59 ALA. L. REV. 305, 343 (2008) (arguing that "the Title I environment is less bleak than suggested by previously published studies. It is undeniable that plaintiffs rarely win in cases that are resolved through judicial opinion and that there has been no apparent increase in employment rates for those with disabilities since the ADA's enactment. However, ADA plaintiffs do not fare poorly with respect to EEOC merit resolutions, and evidence suggests that they also obtain meaningful relief through case settlements and requests for workplace accommodation that are granted by employers. These successes may explain the continued employee-initiated activity under Title I of the ADA." (citations omitted)).

court.[14] Congress finally responded in 2008 with the Americans with Disabilities Act Amendments Act ("ADAAA").[15] Did Congress and the EEOC finally craft a statute and regulations that federal courts could apply uniformly? Unlike the ADA, did the ADAAA fulfill its intended consequences for cancer survivors?

This Article reviews cases filed under Title I of the ADA by employees who claimed they were discriminated against, in part, because of their cancer history. The discussion focuses on cases in which courts considered whether a cancer survivor was a "qualified individual with a disability"[16] under Title I. The Article then compares plaintiffs' outcomes before and after the passage of the ADAAA, and concludes that the ADAAA has significantly improved cancer survivors' ability to obtain favorable judgments under the statute. The Article concludes by summarizing the lessons learned from this analysis.

I.
CANCER SURVIVORS AND WORK

Of the more than 13.7 million cancer survivors in the United States, roughly one-half are of working age.[17] The majority of these individuals are willing and able to work.[18] Due to significant improvements in cancer care, most working-aged adults can work dur-

14. NATIONAL COUNCIL ON DISABILITY, RIGHTING THE AMERICANS WITH DISABILITIES ACT (2004), *available at* http://www.ncd.gov/publications/2004/Dec12004 (criticizing restrictive judicial interpretation of the definition of disability under the ADA); NATIONAL COUNCIL ON DISABILITY, THE AMERICANS WITH DISABILITIES ACT POLICY BRIEF SERIES: RIGHTING THE ADA, NO. 6, DEFINING "DISABILITY" IN A CIVIL RIGHTS CONTEXT: THE COURT'S FOCUS ON EXTENT OF LIMITATIONS AS OPPOSED TO FAIR TREATMENT AND EQUAL OPPORTUNITY (2003), *available at* http://www.ncd.gov/publications/2003/Feb242003.

15. ADA Amendments Act of 2008, Pub. L. No. 110-325, 122 Stat. 3553 (codified at 42 U.S.C. §§ 12101–12213 (Supp. V 2011)).

16. The statute only protects "qualified individuals." *See* 42 U.S.C. § 12112(a) (2006).

17. On January 1, 2012, approximately 3,195,640 men and 4,170,340 women in the United States between the ages of twenty and sixty-nine had been diagnosed with cancer. *Cancer Treatment and Survivorship Statistics 2012*, 62 CA: A CANCER J. FOR CLINICIANS 220, 222–23 (2012). The population of cancer survivors grows significantly each year. For example, approximately 1.6 million Americans were expected to be diagnosed with cancer in 2013. *Cancer Statistics 2013*, 63 CA: A CANCER J. FOR CLINICIANS 11, 16 (2013).

18. *See* Barbara Hoffman, *Cancer and Work: Protections Under the Americans with Disabilities Act*, 24 ONCOLOGY: NURSE EDITION, Apr. 2010, at 15 [hereinafter *Cancer and Work*].

ing and after cancer treatment.[19] For example, one survey of ten studies that assessed return-to-work rates of 1,904 survivors from 1986 to 1999 found that a mean of 62% returned to work.[20]

Nonetheless, cancer can be a devastating disease. Some survivors experience significant physical or mental limitations that affect their ability to work.[21]

Whether a cancer survivor continues to work during treatment or returns to work after treatment—and if so, whether that survivor's diagnosis or treatment will result in work limitations—depends on medical and socioeconomic factors.

Table 1: Factors that Affect a Cancer Survivor's Ability to Work[22]

Medical Factors	Socioeconomic Factors
Age	Financial Status
Type of Cancer	Education
Stage of Cancer	Access to Health Insurance
Side Effects[23]	Access to Transportation
Late Effects[24]	Access to Quality Cancer Care
Other Chronic Health Conditions	Essential Job Functions

19. *Id.* at 15–16 (discussing studies of cancer survivors' work experiences); *see also* Michael Feuerstein et al., *Work in Cancer Survivors: A Model for Practice and Research*, 4 JOURNAL OF CANCER SURVIVORSHIP 415, 416 (2010) (discussing the results of a meta-analysis of cancer and employment in both the United States and Europe that noted an increased risk of unemployment and cancer survivors).

20. *See* Evelien Spelten et al., *Factors Reported to Influence the Return to Work of Cancer Survivors: A Literature Review*, 11 PSYCHO-ONCOLOGY 124, 131 (2002).

21. *See* Cathy J. Bradley & Heather L. Bednarek, *Employment Patterns of Long-Term Cancer Survivors*, 11 PSYCHO-ONCOLOGY 188, 194–96 (2002); Pamela Farley Short et al., *Employment Pathways in a Large Cohort of Adult Cancer Survivors*, 103 CANCER 1292, 1293 (2005); Betty R. Ferrell et al., *"Bone Tired": The Experience of Fatigue and Its Impact on Quality of Life*, 23 ONCOLOGY NURSING FORUM 1539 (1996).

22. "For example, survivors in physically demanding jobs have higher disability rates than those in more sedentary jobs; survivors with advanced education have higher return to work rates than those with less education. Medical treatment decisions that consider quality of life and the shift towards providing cancer treatment in outpatient settings have contributed to the increasing number of survivors who can work during their treatment." Hoffman, *Cancer and Work*, *supra* note 18, at 16 (citations omitted).

23. A short-term side effect is a "problem that is caused by treatment of a disease but usually goes away after treatment. Short-term side effects of cancer treatment include nausea, vomiting, diarrhea, hair loss, fatigue, and mouth sores." *NCI Dictionary of Cancer Terms: Short-Term Side Effect*, NATIONAL CANCER INSTITUTE, http://www.cancer.gov/dictionary?CdrID=730423 (last visited Feb. 14, 2014).

24. A late effect is a "health problem that occurs months or years after a disease is diagnosed or after treatment has ended. Late effects may be caused by

Prior to the passage of state and federal employment rights laws, employment discrimination against cancer survivors was common.[25] Such discrimination imposed devastating physical, emotional, and financial consequences.[26] After the passage of the ADA and comparable state laws, survivors reported decreasing incidences of work problems attributable to their cancer.[27] Most employers accommodate survivors and their caregivers.[28] In a 2006 survey, "three out of five survivors reported receiving co-worker support, such as help with work or random acts of kindness."[29] Survivors and their caregivers also experienced "low incidences of negative reactions from their employers and coworkers."[30]

Over the past generation, enforcement of laws like the ADA, improvements in cancer treatment, and a sea change in perceptions about living with and beyond cancer have greatly enhanced the employment opportunities of cancer survivors. Although incidences of cancer-based employment discrimination have decreased, some survivors still face disparate treatment and seek legal redress. Between 1997 and 2011, 2.3% to 3.9% of claims brought under Title I of the ADA alleged "cancer" as a disability.[31] The most common problems survivors experience at work are receiving less work and responsibility after their diagnoses, "being fired or laid off," being "denied a raise or promotion," and being "denied health insurance benefits."[32] Court confusion as to the proper interpretation of the ADA has prevented these survivors from obtaining the relief that the statute was intended to provide.

---

cancer or cancer treatment. They may include physical . . . problems and second cancers." *NCI Dictionary of Cancer Terms: Late Effect*, NATIONAL CANCER INSTITUTE, http://www.cancer.gov/dictionary?CdrID=390292 (last visited Feb. 14, 2014).

25. *See* Hoffman, *The Cancer Survivors' Catch-22*, *supra* note 8, at 356–60.

26. Brad Zebrack, *Cancer and Job Loss*, 24 ONCOLOGY: NURSE EDITION, Apr. 2010, at 19.

27. *See* Hoffman, *Cancer and Work*, *supra* note 18, at 16; Barbara Hoffman, *Cancer Survivors at Work: A Generation of Progress,* 55 CA: A CANCER J. FOR CLINICIANS, 271, 273 (2005) (discussing studies of cancer survivors' work experiences).

28. Hoffman, *Cancer and Work*, *supra* note 18, at 16 (citing Fleishman-Hillard Research, Breakaway from Cancer Survey (2006) (unpublished survey results) (on file with author)).

29. *Id.*

30. *Id.*

31. EQUAL EMP. OPPORTUNITY COMM'N, *ADA Changes Data by Impairment/Base Receipts, FY 1997–FY 2013*, *available at* http://www.eeoc.gov/eeoc/statistics/enforcement/ada-receipts.cfm (last visited Feb. 15, 2014) [hereinafter EEOC, *ADA Receipts*].

32. Hoffman, *Cancer and Work*, *supra* note 18, at 16 (citing Fleishman-Hillard Research, Breakaway from Cancer Survey (2006) (unpublished survey results) (on file with author)).

II.
THE AMERICANS WITH DISABILITIES ACT OF 1990

A. Title I of the ADA

The ADA was the first federal law to prohibit discrimination by large private employers against individuals with disabilities.[33] Title I of the ADA, which prohibits employment discrimination, provided that "[n]o covered entity shall discriminate against a qualified individual with a disability because of the disability of such individual in regard to job application procedures, the hiring, advancement, or discharge of employees, employee compensation, job training, and other terms, conditions, and privileges of employment."[34] It defined a "qualified individual with a disability" as a person "who, with or without reasonable accommodation, can perform the essential functions of the employment position that such individual holds or desires."[35] The ADA provided three separate definitions of a disability: "(A) a physical or mental impairment that substantially limits one or more of the major life activities of such individual; (B) a record of such an impairment; or (C) being regarded as having such an impairment."[36]

Key terms in the ADA definition of "disability" were defined by EEOC regulations as follows:

> **"Substantially limits"** means "(i) [u]nable to perform a major life activity that the average person in the general population can perform" or "(ii) [s]ignificantly restricted as to the condition, manner or duration under which an individual can perform a particular major life activity as compared to the condition, manner, or duration under which the average person in the general population can perform that same major activity."[37]
>
> **"Physical . . . impairment"** means "any physiological disorder, or condition, cosmetic disfigurement, or anatomical loss affecting one or more of the following body systems: neurological,

33. Prior to the ADA, the only federal law that prohibited employment discrimination based on disability was the Federal Rehabilitation Act of 1973. The Federal Rehabilitation Act prohibited employment discrimination only by programs that received federal financial assistance or by public employers. 29 U.S.C. § 794(a). The ADA covers employers that have "15 or more employees for each working day in each of 20 or more calendar weeks in the current or preceding calendar year, and any agent of such person." 42 U.S.C. § 12111(5)(A) (2006).

34. § 12112(a).

35. § 12111(8).

36. § 12102(2).

37. 29 C.F.R. § 1630.2(j)(1) (1992).

musculoskeletal, special sense organs, respiratory (including speech organs), cardiovascular, reproductive, digestive, genitourinary, hemic and lymphatic, skin, and endocrine."[38]

"Major Life Activities" means "functions such as caring for oneself, performing manual tasks, walking, seeing, hearing, speaking, breathing, learning, and working."[39]

To pursue an employment discrimination claim in federal court, a plaintiff must first file a complaint with the EEOC within 180 days of the alleged discrimination.[40] Most Title I claims are resolved at the administrative level. The EEOC finds that approximately 60% of all claims have no reasonable cause and the charging party withdraws approximately 6% of all claims.[41] Only 20–25% of claims result in a favorable outcome to the charging party.[42]

Cancer survivors experience outcomes similar to those of other claimants. Approximately one in four claims alleging cancer as an impairment are resolved favorably to the charging party.[43] Like all disability-based claims, most cancer-based employment discrimination claims are resolved prior to formal litigation.[44]

Table 2: Cancer-Based Employment Discrimination Claims Filed with the EEOC Between 1997 and 2011

	FY 1997	FY 1998	FY 1999	FY 2000	FY 2001	FY 2002	FY 2003	FY 2004	FY 2005	FY 2006	FY 2007	FY 2008	FY 2009	FY 2010	FY 2011
1	448	460	396	434	454	455	442	427	407	493	578	707	799	978	951
2	2.5	2.6	2.3	2.7	2.8	2.9	2.9	2.8	2.7	3.2	3.3	3.6	3.7	3.9	3.7
3	81	102	103	162	124	121	103	115	103	127	135	173	161	232	257
4	18.1	25.1	26.0	37.3	27.3	26.6	23.3	26.9	25.3	25.8	23.4	24.5	20.2	23.7	27.0
5	16.3	19.1	23.3	30.5	30.9	25.8	23.9	23.1	22.8	22.7	21.9	20.9	19.8	20.8	22.7

38. § 1630.2(h).

39. § 1630.2(i).

40. *See* 42 U.S.C. § 2000e–5(e)(1) (2006).

41. *See* EQUAL EMP. OPPORTUNITY COMM'N, *Americans with Disabilities Act of 1990 (ADA) Charges (Includes Concurrent Charges with Title VII, ADEA, and EPA) FY 1997-FY 2012*, http://www.eeoc.gov/eeoc/statistics/enforcement/ada-charges.cfm (last visited Feb. 14, 2014) [hereinafter EEOC, *ADA Charges*].

42. *See id.*

43. *See infra* Table 2.

44. As a result, most Title I claims do not generate a reported decision because they are resolved at the administrative level, through settlements, or in unpublished judicial decisions. *See* Sharona Hoffman, *supra* note 13, at 312–13.

Key to Table 2

1	Total Number of Charges Filed with the EEOC Under Title I of the ADA Claiming Cancer as an Impairment[45]
2	Cancer Charges as a Percentage of Total Number of Charges Filed[46]
3	Total Number of Cancer Charges with Outcomes Favorable to Charging Parties and/or Charges with Meritorious Allegations (these include negotiated settlements, withdrawals with benefits, successful conciliations, and unsuccessful conciliations)[47]
4	Percentage of Cancer Charges with Outcomes Favorable to Charging Parties and/or Charges with Meritorious Allegations
5	Percentage of All Impairment Charges with Outcomes Favorable to Charging Parites and/or with Meritorious Allegations[48]

Although the EEOC has the authority to file a lawsuit,[49] most ADA lawsuits are initiated by the charging party. For example, the EEOC filed only eighty merit-based lawsuits out of 25,742 Title I claims in Fiscal Year 2011.[50] Like many civil rights claimants, Title I plaintiffs must first exhaust administrative remedies to avoid flooding federal courts with claims that could be addressed in a more

45. EEOC, *ADA Receipts*, *supra* note 31.

46. *Id.*

47. EQUAL EMP. OPPORTUNITY COMM'N, *ADA Charge Data by Impairments/Bases - Merit Factor Resolutions FY 1997-FY 2012*, http://www.eeoc.gov/eeoc/statistics/enforcement/ada-merit.cfm (last visited Feb. 14, 2014) [hereinafter EEOC, *ADA Merit*].

48. *See* EQUAL EMP. OPPORTUNITY COMM'N, *supra* note 41. To date, the percentage of EEOC claims resolved favorably for the charging party has increased, though not significantly, since the passage of the ADAAA. *See supra*, Table 2. The reason cancer survivors may not yet see significant improvement in outcomes at the administrative level may be due to many factors, including: (1) one purpose of the EEOC is to screen out nonmeritorious claims, see Laura M. Hyer, *Is Cooperation with the EEOC an Implied Requirement for Exhaustion of Administrative Remedies?*, 98 IOWA L. REV. 1351, 1356 (2013); (2) some claimants and defendants are not represented by attorneys in EEOC proceedings, see Kathryn Moss et al., *Unfunded Mandate: An Empirical Study of the Implementation of the Americans with Disabilities Act by the Equal Employment Opportunity Commission*, 50 U. KAN. L. REV. 1, 105–06 (2001); and (3) some cases, especially those with attorney representation, are withdrawn voluntarily to proceed in court, *id.* at 105.

49. 42 U.S.C. § 2000e–5(e)(1) (2006).

50. *See* Barry A. Hartstein, *EEOC Receives a Record Number of Private Sector Discrimination Charges and Secures Highest Amount in Damages in FY 2011*, WASHINGTON D.C. EMP. L. UPDATE (Nov. 18, 2011), http://www.dcemploymentlawupdate.com/2011/11/articles/eeoc-1/eeoc-receives-a-record-number-of-private-sector-discrimination-charges-and-secures-highest-amount-in-damages-in-fy-2011/ (last visited Feb. 14, 2014); EEOC, *ADA Receipts*, *supra* note 41.

efficient setting.[51] Typically, fewer than one in ten complainants resolve their claims in federal court.[52]

B. References to Cancer as a Disability in the ADA Regulations and Legislative History

Nowhere in the ADA itself does the word "cancer" appear. The original EEOC regulations to Title I and the legislative history of the ADA, however, suggest that Congress intended the statute to prohibit cancer-based employment discrimination.

1. Compliance Manual to the Original EEOC Regulations

The ADA authorized the EEOC to issue regulations explaining how Title I should apply to specific circumstances.[53] The EEOC Compliance Manual ("Manual"), first issued in January 1992 to supplement the federal regulations, specifically illustrated how the EEOC intended the ADA to apply to cancer-based discrimination through eight references:[54]

1. The ADA protects not only individuals with a visible disability, but also those who have "hidden disabilities, such as . . . cancer."[55]

51. *See* Sharona Hoffman, *supra* note 13, at 314.

52. In a typical month in 2011, approximately 150 Title I cases were filed in federal court compared with 2145 complaints filed with the EEOC. *Compare Increase in Employment Discrimination Lawsuits under the Americans with Disabilities Act*, TRAC REPORTS, http://trac.syr.edu/tracreports/civil/282/ (last visited Feb. 14, 2014), *with* EEOC, *ADA Charges*, *supra* note 41.

53. *See* 42 U.S.C. § 12116 (2006) (stating that "the [EEOC] shall issue regulations in an accessible format to carry out this subchapter"); *see also* Bragdon v. Abbott, 524 U.S. 624, 647 (1998) (noting in a parenthetical that 42 U.S.C. § 12116 authorizes the "EEOC to issue regulations implementing Title I").

54. The EEOC issued the Technical Assistance Manual "to help employers . . . and persons with disabilities learn about their obligations and rights under the employment provisions of the Americans with Disabilities Act," providing "guidance on the practical application of legal requirements established in the statute and EEOC regulations." EQUAL EMP. OPPORTUNITY COMM'N, A TECHNICAL ASSISTANCE MANUAL ON THE EMPLOYMENT PROVISIONS (TITLE I) OF THE AMERICANS WITH DISABILITIES ACT, at i (1992), *reprinted in* EQUAL EMP. OPPORTUNITY COMM'N, ADA COMPLIANCE GUIDE app. IV (2013), *available at* 2005 WL 4899269.

55. 2 EQUAL EMP. OPPORTUNITY COMM'N, NO. 130, A TECHNICAL ASSISTANCE MANUAL ON THE EMPLOYMENT PROVISIONS (TITLE 1) OF THE AMERICANS WITH DISABILITIES ACT § 902 (1995) [hereinafter 1995 TECHNICAL ASSISTANCE MANUAL], *cited in The Cancer Survivors' Catch-22*, *supra* note 8, at 368 n.83.

2. "Most forms of heart disease and cancer" are types of medical conditions that may substantially limit a major life activity.[56]
3. The ADA "protects former cancer patients from discrimination based on their prior medical history" as individuals with a "record of a substantially limiting condition."[57]
4. An individual with a genetic marker for cancer is covered by the ADA as they are "regarded as having a substantially limiting impairment."[58]
5. Employers may not use "medical inquiries or medical examinations before making a conditional job offer" to screen out individuals with a "hidden disability such as . . . cancer."[59]
6. How an individual develops an impairment, such as whether an individual "develops lung cancer as a result of smoking," is irrelevant to ADA standing.[60]
7. In an illustration of flex-time as a reasonable accommodation, the Manual states that cancer survivors are entitled to reasonable modifications in their work schedules to accommodate the side effects of cancer treatment.[61]
8. In an example of the impact of coworkers' attitudes toward individuals with disabilities, the Manual instructs that an employer may not discriminate against an employee on the

56. *Id.* at § 902.4, *cited in The Cancer Survivors' Catch-22*, *supra* note 8, at 368 n.84.

57. *See* 29 C.F.R. § 1630.2(k) (1992) (stating that "a record of such an impairment [meant that the individual] has a history of, or has been misclassified as having, a mental or physical impairment that substantially limits one or more major life activities"). The Manual states that this language refers to "people with a history of cancer . . . or other debilitating illness whose illnesses are either cured, controlled or in remission." EQUAL EMP. OPPORTUNITY COMM'N, EEOC-M-1A, EEOC COMPLIANCE MANUAL: A TECHNICAL ASSISTANCE MANUAL ON THE EMPLOYMENT PROVISIONS (TITLE I) OF THE AMERICANS WITH DISABILITIES ACT, pt. 2, § 2.2(b), at II-8 (1992), *available at* http://ia700504.us.archive.org/21/items/technicalassista00unse/technicalassista00unse_bw.pdf [hereinafter 1992 TECHNICAL ASSISTANCE MANUAL].

58. *See, e.g.*, 1995 TECHNICAL ASSISTANCE MANUAL, *supra* note 55, at § 902.8 (giving as an example that a person with a genetic marker for colon cancer will be covered by the ADA).

59. 1992 TECHNICAL ASSISTANCE MANUAL pt. 6, §6.3, at VI-3 to VI-4.

60. 1995 TECHNICAL ASSISTANCE MANUAL, *supra* note 55, at § 902.2.

61. *See* EEOC COMPLIANCE MANUAL (1999), *reprinted from* EQUAL EMP. OPPORTUNITY COMM'N, NO. 915.002, ENFORCEMENT GUIDANCE: REASONABLE ACCOMMODATION AND UNDUE HARDSHIP UNDER THE AMERICANS WITH DISABILITIES ACT (Mar. 1, 1999), *available at* 1999 WL 33305876, at *26.

basis that other employees react negatively to the employee because of the employee's cancer history.[62]

2. Legislative History

The House and Senate subcommittees considered the testimony of cancer advocates during hearings on the ADA.[63] The final reports of the subcommittees illustrated that cancer-based employment discrimination should fall within the scope of the ADA.[64] The reports recognized that although the ADA could not include an exhaustive, comprehensive list of every disorder that could be covered, the term "disability" includes "such conditions, diseases and infections as: . . . cancer."[65] The reports also confirmed that "persons with histories of . . . cancer" had a record of an impairment.[66]

Despite favorable language in the ADA, EEOC regulations, and legislative history, federal courts often dismissed cancer survivors' employment discrimination claims between 1992 and 2008.

III. HOW CANCER SURVIVORS FARED IN EMPLOYMENT DISCRIMINATION CASES UNDER THE ADA

A. Judgments for Employers

Like many ADA plaintiffs,[67] cancer survivors often found themselves on the losing end of motions for summary judgment. As noted above, plaintiffs could prove disability status under the ADA in three ways: by demonstrating (1) a physical or mental impairment that substantially limits one or more of their major life activi-

62. *See id.*

63. For a detailed description of this testimony, see *The Cancer Survivors' Catch-22*, *supra* note 8, at 371–73.

64. *See* H.R. REP. NO. 101-485, pt. 3, at 53, *reprinted in* 1990 U.S.C.C.A.N. 303, 334–35; S. REP. NO. 101-116, at 23, *reprinted in* 1 LEGISLATIVE HISTORY OF PUBLIC LAW 101–336, THE AMERICANS WITH DISABILITIES ACT, at 120 (1990).

65. H.R. REP. NO. 101-485, pt. 2, at 51, *reprinted in* 1990 U.S.C.C.A.N. 303, 333; S. REP. NO. 101-116, at 22, *reprinted in* 1 LEGISLATIVE HISTORY OF PUBLIC LAW 101–336, THE AMERICANS WITH DISABILITIES ACT, at 120 (1990).

66. H.R. REP. NO. 101-485, pt. 2, at 52–53, *reprinted in* 1990 U.S.C.C.A.N. 303, 334–35; S. REP. NO. 101-116, at 23, *reprinted in* 1 LEGISLATIVE HISTORY OF PUBLIC LAW 101–336, THE AMERICANS WITH DISABILITIES ACT, at 121 (1990).

67. *See* Colker, *The Americans with Disabilities Act*, *supra* note 12, at 109. Colker also notes that in the first few years of the ADA, "defendants prevail[ed] in more than ninety-three percent of reported ADA employment discrimination cases decided on the merits at the trial court level. Of those cases that [were] appealed, defendants prevail[ed] in eight-four percent of reported cases." *Id.* at 100.

ties; (2) a record of such an impairment; or (3) being regarded as having such an impairment. At the summary judgment stage, however, federal courts were quick to find that cancer rendered plaintiffs either too sick to be qualified for their jobs or not sufficiently impaired to be disabled.[68] As discussed below, courts routinely held that survivors whose cancers were treated successfully, despite the survivors' records of hospitalization and grueling treatments, were not substantially limited in major life activities because their cancer did not have a "permanent or long-term impact."[69] Similarly, courts often granted summary judgment to employers under the second and third prongs. As a result, survivors were unable to defeat motions for summary judgment.

1. Plaintiffs Did Not Have a Disability Because They Were Not Substantially Limited in the Major Life Activity of Working

Few appellate decisions set back survivors' rights more than the Fifth Circuit's narrow reading of the ADA in *Ellison v. Software Spectrum, Inc.*[70] Phyllis Ellison was diagnosed with breast cancer in August 1993, nineteen months after becoming a full-time buyer for Software Spectrum.[71] To accommodate her treatment and recovery, Ellison worked a modified schedule from September 1993 through February 1994.[72] The following month, Software Spectrum fired Ellison and three other employees as part of an alleged downsizing.[73]

The Fifth Circuit affirmed the lower court's grant of summary judgment to Software Spectrum on the ground that Ellison's breast cancer was not a disability.[74] Noting that Ellison never missed a day of work during her cancer treatment, the Fifth Circuit affirmed the lower court's finding that Ellison was not substantially limited in her ability to work.[75] Relying on Ellison's admission that, although her treatment left her feeling sick and fatigued, she was able to perform her job with the reasonable accommodation of a modified work schedule, the Fifth Circuit concluded that she was not suffi-

68. *See The Cancer Survivors' Catch-22*, *supra* note 8, at 375–94 (discussing cancer-based employment discrimination cases pre-*Sutton*).

69. 29 C.F.R. 1630, app. 1630.2(j)(4) (2012); *see infra* Part III.A.

70. 85 F.3d 187 (5th Cir. 1996).

71. *Id.* at 189.

72. *Id.*

73. *Id.* Ellison was subsequently rehired to a different position several weeks later. The Fifth Circuit relied in part upon Ellison's paucity of harm to conclude that she failed to raise a genuine issue of material fact. *Id.* at 193.

74. *Id.* at 193.

75. *Id.* at 191.

ciently limited in her ability to work to establish standing under the ADA.[76]

Thus, *Ellison* sent a halting message to cancer survivors bringing Title I claims: employees who find a way to work during and after cancer treatment are not disabled within the meaning of the ADA. This decision effectively imposed a Catch-22 on plaintiffs who alleged that their cancer substantially limited the major life activity of working. Under this reasoning, had Ellison chosen to take medical leave instead of working a modified schedule during her treatment, her cancer would have substantially limited her major life activity of working.[77] This logic ironically punishes employees for seeking and taking reasonable accommodations to help them perform their essential job functions, a fundamental goal of the ADA.

For nearly fifteen years, the tidal effects of *Ellison* capsized survivors' claims. Courts in almost every circuit adopted the Fifth Circuit's analysis to conclude that cancer survivors who were healthy enough to work were not sufficiently impaired to be considered disabled under the ADA.[78] The unfortunate result was that plaintiffs who failed to allege that their cancer substantially limited a major life activity besides working rarely made it past the summary judgment stage.

a. First Circuit

Ellen Whitney worked as an executive assistant in an accounting firm when she was diagnosed with ovarian cancer.[79] She claimed that dementia caused by chemotherapy for ovarian cancer substantially limited her ability to work and learn.[80] "Chemobrain" is a common side effect of some chemotherapies that manifests as cognitive dysfunction, such as problems with memory and concen-

76. *Id.*

77. *See The Cancer Survivors' Catch-22*, *supra* note 8, at 380.

78. *See, e.g.*, Cook v. Robert G. Waters, Inc., 980 F. Supp. 1463, 1469 (M.D. Fla. 1997) (granting summary judgment to the employer of a brain tumor patient who the court held was not substantially limited in her ability to work because "except for during doctor appointments and during Plaintiff's headaches, Plaintiff state[d] that she could fulfill the essential requirements of the job . . . until she was terminated"); Madjlessi v. Macy's W., Inc., 993 F. Supp. 736, 741 (N.D. Cal. 1997) (finding that a breast cancer survivor who relied on an accommodation of flex-time to perform her job was not protected by the ADA because her "breast cancer did not substantially limit her ability to work"); Hirsch v. National Mall & Serv., Inc., 989 F. Supp. 977, 981–82 (N.D. Ill. 1997) (noting that Hirsch continued to work, although he indicated that it was difficult to do so).

79. Whitney v. Greenberg, Rosenblatt, Kull, & Bitsoli, P.C., 258 F.3d 30, 31 (1st Cir. 2001).

80. *Id.* at 31–32.

tration.[81] A psychologist's testimony that Whitney's condition was "mild yet significant"[82] was precisely the type of factual allegation that begs a trial, as "mild" suggests her impairment was not substantially limiting, yet "significant" suggests the opposite. Nevertheless, the trial court granted summary judgment for the accounting firm, holding that Whitney did not allege "sufficient evidence to demonstrate that her cognitive disability was severe or lengthy enough to substantially limit her ability to work or to learn."[83] The First Circuit affirmed summary judgment for her employer, finding that Whitney was not substantially limited in her ability to work or learn because her "cognitive impairment was mild, reversible, and short lived."[84]

b. Second Circuit

The Second Circuit affirmed summary judgment for an employer who failed to accommodate a cancer survivor.[85] Steven Thomsen claimed that he had a disability because he had to take three medical leaves for surgery for bowel cancer.[86] Even though Thomsen could not work for six weeks after each surgery, the court held that Thomsen "produced no evidence that his impairments—whatever they might have been during his medical leaves of absence—were significantly restrictive, persistent or permanent."[87] The court relied on Thomsen's testimony that he was not limited in any major life activity, including working, and that at the time he was fired, "he was cancer-free."[88]

c. Sixth Circuit

Rick Stokes took a three-month leave of absence from his job as a county emergency medical technician to receive treatment for kidney cancer.[89] He alleged that he was subsequently fired because

81. *See* LOST IN TRANSITION, *supra* note 5, at 101.

82. 258 F.3d at 31.

83. *Id.* at 32.

84. *Id.* at 34.

85. Thomsen v. Stantec, Inc., 785 F. Supp. 2d 20, 23 (W.D.N.Y. 2011), *aff'd*, 483 F. App'x 620 (2d Cir. 2012), *cert. denied*, 133 S. Ct. 931 (2013).

86. *Id.*

87. *Id.* ("'[S]ubstantially limits' means '[s]ignificantly restricted as to the condition, manner or duration under which an individual can perform a particular major life activity as compared to the condition, manner, or duration under which the average person in the general population can perform that same major life activity.'") (quoting 29 C.F.R. § 1630.2(j)(1)(ii)).

88. *Id.*

89. Stokes v. Hamilton Cty., 113 Fed. App'x 680, 682 (6th Cir. 2004).

the county feared that it would have to pay for his medical care if his cancer recurred.[90] The Sixth Circuit rejected this argument, holding that Stokes had failed to prove that he had an actual disability because his kidney cancer did not impose a long-term burden on his ability to work.[91] Like Thomsen, Stokes was essentially punished for returning to work after successful cancer treatments.

d. Seventh Circuit

Paul Hirsch was fired after working for several years while receiving treatment for non-Hodgkin's lymphoma.[92] The District Court for the Northern District of Illinois granted his employer summary judgment.[93] The court held that Hirsch's cancer did not substantially limit his ability to work because he "continue[d] to work,"[94] even though his illness forced him to work part-time from home and to be occasionally absent.[95] The court failed to reconcile how Hirsch could not be substantially limited in his ability to work, yet was so ill that he "asked to work at home part-time"[96] and "was forced to be occasionally absent from work."[97] He ultimately died from his lymphoma seventeen months after he was fired.[98]

e. Eighth Circuit

Susan Treiber was a breast cancer survivor who was denied reappointment as a school teacher.[99] Although the court recognized that Treiber's breast cancer was an impairment, it held that her

90. *Id.*

91. *Id.* at 684. The Sixth Circuit also rejected his claim that the county had discriminated against him based on a perceived disability. *Id.* However, much of the court's reasoning was impermissibly based on a case that correctly held that "perceived disability" claims do not apply to employers who discriminate to reduce worker compensation payments. *See* Mahon v. Crowell, 295 F.3d 585, 593 (6th Cir. 2002). The purpose of the ADA is to prohibit disability-based employment discrimination, not discrimination against employees who file workers' compensation claims to recover for job-related injuries.

92. Hirsch v. National Mall & Serv., Inc., 989 F. Supp. 977, 981–82 (N.D. Ill. 1997).

93. *Id.* at 983.

94. *Id.* at 981 (noting that Hirsch continued to work, although he indicated that it was difficult to do so).

95. *Id.*

96. *Id.*

97. *Id.*

98. *Id.* at 978–79.

99. Treiber v. Lindbergh Sch. Dist., 199 F. Supp. 2d 949, 953–54 (E.D. Mo. 2002).

cancer did not substantially limit any major life activities, including her ability to work:

> Rather than evidence that her cancer and its treatment caused substantial limitations, Plaintiff has introduced evidence to the contrary. She returned to work after approved leave for surgery, scheduled her chemotherapy treatments so that her recovery time would primarily fall on the weekends, and underwent radiation therapy in the summer. There is no testimony that the cancer or treatment adversely affected her ability to perform her teaching duties.[100]

Like Susan Treiber, Joan Rickert was caught in the Catch-22. She successfully scheduled her breast cancer treatment during the summer when the college was not in session to avoid missing work, and the "volleyball team changed the practice schedule to accommodate Rickert's medical appointments."[101] Rickert then sued Midland Lutheran College for failing to promote her from a part-time to a full-time volleyball coach because she had breast cancer.[102] The court granted Midland summary judgment, finding that Rickert was not substantially limited in her ability to work because she never missed a day of work.[103]

f. Ninth Circuit

Virginia Madjlessi unknowingly undermined her subsequent Title I claim by working during grueling breast cancer treatment.[104] In granting summary judgment to her employer, the court relied on *Ellison* and on Madjlessi's evidence that she was qualified for her job to support its conclusion that she was not substantially limited in her ability to work.[105] It recognized that she was not only qualified for her job, but "that she worked even harder to move herself up the career path. Her coworkers praise her ability to work through adversity."[106] In essence, the court punished Madjlessi for working "while suffering the side effects of vomiting, weakness and nausea," because by doing so, she proved she was not substantially

100. *Id.* at 961.

101. Rickert v. Midland Lutheran Coll., No. 8:07CV334, 2009 WL 2840528, at *3 (D. Neb. Sept. 2, 2009).

102. *Id.* at *1.

103. *Id.* at *14.

104. Madjlessi v. Macy's W., Inc., 993 F. Supp. 736, 741 (N.D. Cal. 1997) (granting summary judgment to employer). Madjlessi took only four days off from work every month for cancer treatment. *Id.*

105. *Id.*

106. *Id.*

limited.[107] In granting summary judgment to her employer, the court concluded that Madjlessi failed to prove that she had a disability.[108]

g. Eleventh Circuit

Donna Dalton experienced long-term lingering arm pain and swelling caused by lymphedema resulting from breast cancer treatment.[109] She produced medical evidence that although she needed some time off for pain treatment, when she was at work she could perform all of her responsibilities.[110] Seeking an accommodation, she asked to work part-time or to transfer to a less stressful position, but her employer denied her requests.[111] Her employer then fired her, claiming that she failed to meet her job's performance standards.[112] Instead of allowing this factual dispute to be resolved by a jury, the court granted summary judgment to her employer.[113] Although the court recognized that Dalton's cancer-related pain, fatigue, and depression were impairments,[114] it determined that she was able to work, and thus not protected by the ADA.[115]

107. *Id.* (citing Ellison v. Software Spectrum, Inc., 85 F.3d 187, 191 (5th Cir. 1996) ("Obviously, her ability to work was affected [by breast cancer]; but as reflected in the above quoted statutes and regulations, far more is required to trigger coverage under § 12102(2)(A).")).

108. 933 F. Supp. at 740–41.

109. Dalton v. Geico Annuity & Ins. Co., Civil Action No. 7:08–CV–130 (HL), 2010 WL 2640097, at *7 (M.D. Ga. June 28, 2010).

110. *Id.* at *2.

111. *Id.*

112. *Id.* at *4.

113. *Id.* at *9.

114. *Id.* at *6.

115. *Id.* at *7 (finding that, absent taking short medical leaves, Dalton was cleared by her physicians to work).

Many other Eleventh Circuit cases granted summary judgment to employers under the ADA. *See also* Schwertfager v. City of Boynton Beach, 42 F. Supp. 2d 1347, 1360 (S.D. Fla. 1999) (granting summary judgment to employer because plaintiff's breast cancer treatment did not render her "incapacitated"—a standard unsupported by the statutory language—and despite the court's concession that plaintiff's breast cancer limited her ability to perform, other major life activities including her ability to care for herself, dress, and cook); Cook v. Robert G. Waters, Inc., 980 F. Supp. 1463, 1469 (M.D. Fla. 1997) (holding that plaintiff was not substantially limited in her ability to work despite her severe headaches and diminished ability to concentrate caused by plaintiff's brain tumor); Gordon v. E.L. Hamm & Assocs., Inc. 100 F.3d 907, 912 (11th Cir. 1996) (rejecting plaintiff's claim that his lymphoma was a disability because "except for a couple of days of medical testing and a [ten-day] leave of absence . . . Gordon was fully capable of working").

By concluding that a survivor who is healthy enough to perform essential job functions does not have standing under the ADA, the courts precluded a determination of whether the employer discriminated against the employee because of his or her cancer. These decisions illustrate how courts erected pretrial barriers to employees who claimed that their cancer substantially limited their ability to work. Survivors who demonstrated that they were "qualified individuals"[116] were trapped in a Catch-22 of being unable to prove that they were substantially limited in the major life activity of working. Thus, these decisions prevented survivors from having the opportunity to argue the merits of their claims.

2. Plaintiffs Did Not Have a Disability Because They Could Not Establish a Record of a Disability

Some cancer survivors claimed that they had standing under the ADA's "record" of a disability prong because they were hospitalized for surgery and/or treatment.[117] The EEOC Compliance Manual to the ADA stated that "people with a history of cancer . . . or other debilitating illness, whose illnesses are either cured, controlled or in remission" have a "[r]ecord of a substantially limiting condition."[118] The EEOC specifically noted that "this provision protects former cancer patients from discrimination based on their prior medical history."[119] Moreover, the Supreme Court recognized in *School Board of Nassau County v. Arline* that hospitalization is evidence of a record of a disability.[120]

116. 42 U.S.C. § 12111(8) (2006).

117. *See* § 12102(2)(B); *see infra* text accompanying notes 122–28.

118. 1992 TECHNICAL ASSISTANCE MANUAL, pt. 2, §2.2(b), at II-8.

119. 29 C.F.R. § 1630, app. § 1630.2(k) (1992).

120. Because the "ADA's definition of disability is drawn almost verbatim from the definition of 'handicapped individual' included in the Rehabilitation Act of 1973," courts have relied on Rehabilitation Act cases to interpret the ADA. Bragdon v. Abbott, 524 U.S. 624, 631–32 (1998).

In interpreting identical language in the Rehabilitation Act, the Supreme Court held that a school teacher with tuberculosis had a record of an impairment because she was hospitalized. Sch. Bd. of Nassau Cty. v. Arline, 480 U.S. 273, 281 (1987) ("This impairment was serious enough to require hospitalization, *a fact more than sufficient to establish that one or more of her major life activities were substantially limited by her impairment*. Thus, Arline's hospitalization for tuberculosis in 1957 suffices to establish that she has a 'record of . . . impairment' within the meaning of 29 U.S.C. § 706(7)(B)(ii), and is therefore a handicapped individual.") (emphasis added). *But see* EEOC v. Gallagher, 181 F.3d 645, 655 (5th Cir. 1999) (holding that because the ADA requires an individualized inquiry, "it is not enough for an ADA plaintiff to simply show that he has a record of a cancer diagnosis" because he was hospitalized; he must prove that he was substantially limited in a major life activity).

Despite the *Arline* decision, most courts rejected the argument that hospitalization alone proved a record of a disability. Courts were also reluctant to find a record of a disability if the duration of the cancer was short or temporary, thus creating another barrier to recovery for ADA plaintiffs.[121]

Silvia Day returned to work following surgery and chemotherapy for breast cancer.[122] Day was hospitalized for *four months*, yet the Fifth Circuit concluded that her impairment was "temporary."[123] The Fifth Circuit affirmed summary judgment for her employer, finding that, although Day's "cancer may have been severe, its duration was short and its long-term impact minimal."[124] Thus she did not have a "record of an impairment that substantially limited a major life activity."[125]

Several courts applied the same reasoning to deny survivors standing under the "record of an impairment" prong. The District Court for the Eastern District of Missouri held that the fact that Susan Treiber was hospitalized for breast cancer surgery "in and of itself" did not prove she had a record of a disability.[126] The District Court for the Northern District of California rejected Madjlessi's claim that her extensive cancer treatment proved she had a history of a disability.[127] And the Second Circuit explicitly rejected the argument that hospitalization alone is sufficient to prove a record of a substantially limiting impairment because the "ADA requires an in-

121. *See, e.g.*, Schwertfager v. City of Boynton Beach, 42 F. Supp. 2d 1347, 1359–60 & n.10 (S.D. Fla. 1999) ("Merely having a record of hospitalization is insufficient to establish that Schwertfager had a record of a substantially limiting impairment."). *But see* Mark v. Burke Rehabilitation Hosp., No. 94 Civ. 3596 RLC, 1997 WL 189124, at *4 (S.D.N.Y. Apr. 16, 1997) (denying summary judgment to employer because employee's hospitalization for cancer surgery created a record of a substantially limiting impairment).

122. Day v. Earthgrains, Co., 211 F.3d 124 (5th Cir. 2000) (unpublished table decision), *available at* 2000 WL 309415, at *1.

123. *Id.*

124. *Id.*

125. *Id*; *see also* Ellison v. Software Spectrum, Inc., 85 F.3d 187, 192 (5th Cir. 1996) (affirming summary judgment for the employer because, in part, "nothing in Ellison's personnel file has ever indicated that she was substantially limited by a physical or mental impairment either in her ability to perform her job or in any other respect").

126. Treiber v. Lindbergh Sch. Dist., 199 F. Supp. 2d 949, 961–62 (E.D. Mo. 2002).

127. Madjlessi v. Macy's W., Inc., 993 F. Supp. 736, 742 (N.D. Cal. 1977) ("[T]he mere fact that Madjlessi had cancer and was utterly incapacitated for brief periods of time after chemotherapy does not mean she was 'substantially limited' for purposes of the ADA.").

dividualized inquiry beyond the mere existence of a hospital stay."[128]

Other survivors unsuccessfully argued that having cancer in remission creates a history of a disability.[129] For example, Rene Olmeda claimed that the New York State Department of Civil Service failed to hire him as a parole officer because he had been diagnosed with and successfully treated for leukemia seven years earlier.[130] The trial court dismissed his complaint, holding that he did not have standing under the ADA solely because Olmeda "testified that his leukemia has been in remission since 1987 and asserted that he is not limited in any way."[131]

3. Plaintiffs Did Not Have a Disability Because Their Employers Did Not Regard Their Cancer as a Disability

Cancer survivors who allege that their employers regarded their cancer as a disability[132] also face significant pretrial hurdles because this alternate prong requires evidence of an employer's beliefs.[133] This third prong is intended "to combat the effects of 'archaic attitudes,' erroneous perceptions, and myths that work to the disadvantage of persons with or regarded as having disabilities."[134] Under the original ADA, many courts did little to further the purpose of this prong—"to combat the effects of 'archaic attitudes,' erroneous perceptions, and myths"[135]—because they required uncontroverted evidence that an employer regarded the plaintiff's cancer as a disability. Plaintiffs could not rely on evidence that their employers knew they had cancer or that their employers made general statements linking their cancer with their inability to

128. Colwell v. Suffolk Cty. Police Dep't, 158 F.3d 635, 646 (2d Cir. 1998) (citing Burch v. Coca-Cola, Inc., 119 F.3d 305, 317 (5th Cir. 1997)).

129. *See, e.g.*, Thomsen v. Stantec, Inc., 785 F. Supp. 2d 20, 23 (W.D.N.Y. 2011) (despite acknowledging that plaintiff was cancer-free, court declined to address whether he proved he had a record of a disability), *aff'd*, 483 F. App'x 620 (2d Cir. 2012), *cert. denied*, 133 S. Ct. 931 (2013).

130. Olmeda v. N.Y. State Dep't of Civil Serv., No. 96 Civ. 7557(HB), 1998 WL 17729, at *1–2 (S.D.N.Y. Jan. 16, 1998).

131. *Id.* at *2.

132. *See* 42 U.S.C. § 12102(2)(C) (2006); *infra* notes 138–49 and accompanying text.

133. *See* 29 C.F.R. § 1630.2(l) (2012) (defining an individual as "'regarded as having such an impairment' if the individual is subjected to a prohibited action because of a . . . perceived . . . impairment"); *cf.* Olds v. United Parcel Serv., 127 Fed. App'x 779, 782 (6th Cir. 2005) (holding that plaintiff raised a genuine issue of material fact as to whether his employer regarded his cancer as disabling).

134. Wooten v. Farmland Foods, 58 F.3d 382, 385 (8th Cir. 1995).

135. *Id.*

work.[136] Even direct admissions by employers that they doubted their employees could perform their jobs because they had cancer proved insufficient.[137]

For example, James McElroy claimed that he was fired because his employer mistakenly assumed his prostate and colon cancers were disabilities that rendered him unfit for his job.[138] The Sixth Circuit held that to prevail under the "regarded as" prong, McElroy had to prove "not only that his employer thought he was disabled, but also that his employer thought his disability would prevent him from performing a broad class of jobs."[139] Although McElroy presented evidence that his supervisor referred to his past "serious health problems" as the reason that he needed to "rekindle his skills,"[140] the Sixth Circuit held that this evidence fell short of McElroy's burden of proof.[141] In affirming summary judgment for his employer, the court stated, "taking the evidence together and in the light most favorable to plaintiff, there is, at best, a weak inference that Wyatt may have believed McElroy's past health problems had affected his performance as an account manager."[142]

The Fifth Circuit took an even tougher stance, holding in *Paulsen v. Beyond, Inc.* that a plaintiff must not only prove that the employer discriminated against the employee based upon a perceived disability, but also that the employer perceived the disability to have substantially limited a major life activity. Lynn Paulsen had been successfully treated for cancer ten years before she became a sales

136. *See, e.g.*, Stokes v. Hamilton Cty., 113 Fed. App'x at 680, 864 (6th Cir. 2004) (affirming summary judgment, in part, because plaintiff failed to allege how his employer "'entertain[ed] misperceptions' about Stokes's past battle with cancer which related to his ability to perform his job."); Trieber v. Lindbergh Sch. Dist., 199 F. Supp. 2d 949, 962 (E.D. Mo. 2002) (granting summary judgment to employer, in part, because plaintiff failed to prove that her employer perceived her breast cancer as a disability).

137. *See, e.g.*, Pikoris v. Mount Sinai Med. Ctr., No. 96 CIV. 1403(JFK), 2000 WL 702987, at *13 (S.D.N.Y. May 30, 2000) (finding that the plaintiff, who had been diagnosed with and treated for breast cancer, had not established that she was perceived as having a disabling impairment based on her supervisor's two comments expressing concern that the plaintiff could not perform her job as a medical resident because she could not handle the stress of cancer).

138. McElroy v. Philips Med. Sys., 127 Fed. App'x 161, 168–69 (6th Cir. 2005) (summarizing plaintiff's claims that his employer terminated him to retaliate against his EEOC claims alleging he was regarded as having a disability).

139. *Id.* at 168.

140. *Id.* at 169.

141. *Id.*

142. *Id.*

manager for a software company.[143] She subsequently took off several weeks to have surgery to remove scar tissue caused by her earlier cancer treatment.[144] When she informed the software company that she was ready to return to work, she was fired.[145] Paulsen alleged that when she sought to return to work, one supervisor inquired, "You used to have cancer, didn't you? Aren't you afraid it's going to come back?," and another supervisor questioned whether her current medical problems were related to her prior cancer.[146] Paulsen alleged that she was protected by the ADA because her employer regarded her cancer history as a disability.[147] The Fifth Circuit affirmed judgment for the software company, ruling that Paulsen failed to present evidence that her employer "treated her as if the supposed cancer substantially limited a major life activity."[148] The Fifth Circuit discounted Paulsen's evidence of how her cancer history in fact substantially limited major life activities, including reproduction, as well as Paulsen's evidence of her supervisors' concerns about her cancer history.[149]

B. Judgments for Employees

Unlike the majority of cancer-survivor plaintiffs, some cancer survivors prevailed in pretrial motions by proving that they had a disability, a record of a disability, or were regarded as having a disability. These survivors avoided the Catch-22 by pleading that their cancer substantially limited a major life activity other than work or that they had standing under the "record of" or "regarded as" prongs.

1. Major Life Activity Other Than Work

Survivors who argued that the major life activity affected by their cancer was a quality of life issue fared better than survivors who argued that their cancer substantially limited only their ability to work. For example, Nancy DeMarah claimed that her supervisor harassed her because she was receiving treatment for breast can-

143. Paulsen v. Beyond, Inc., 91 F.3d 140 (5th Cir.1996) (unpublished table decision) (No. 95-40107), *available at* 1996 WL 40029, at *1.

144. *Id.*

145. *Id.* at *2.

146. *Id.* at *1–2.

147. *Id.* at *4.

148. *Id.* at *5.

149. *Id.* at *5–6. Paulsen alleged that she had many reproductive organs removed and that she was subsequently unable to have children. *Id.*

cer.[150] She alleged that her cancer affected her major life activities of walking and caring for herself.[151] The trial court denied summary judgment to her employer and found that DeMarah was disabled because, after she had a mastectomy, reconstructive surgery, and chemotherapy, she could walk only short distances and could not care for herself or her youngest son.[152]

Judith Keller, who was fired from her job as a school superintendent, proved that her breast cancer treatment substantially limited her ability to have sexual relations.[153] Recognizing that sexual intercourse and reproduction are major life activities,[154] the court denied her employer's motion for summary judgment.[155] It relied on a short, but comprehensive list of cancer survivors who had proven that their cancer was a disability under the ADA.[156] Similarly, Claudia Berk successfully alleged that her breast cancer substantially limited her major life activity of reproduction because her

150. DeMarah v. Texaco Grp., Inc., 88 F. Supp. 2d 1150, 1151 (D. Colo. 2000).

151. *Id.* at 1155.

152. *Id.*

153. Keller v. Board of Educ., 182 F. Supp. 2d 1148, 1154 (D.N.M. 2001).

154. *Id.* at 1155.

155. *Id.* at 1164.

156. *Id.* at 1155. The trial court cited to nine cases where the court recognized that the plaintiff's allegation that his or her cancer was a disability was sufficient to survive summary judgment:

> Furthermore, many courts have found that cancer is a disability, or at a minimum that such a determination is a question of fact, not law. *Demarah v. Texaco,* 88 F. Supp. 2d 1150 (D. Colo. 2000) (finding genuine issues of material fact as to whether former employee whose breast cancer was in remission was substantially limited in the life activities of working, walking, and caring for herself for a sufficient duration so as to be considered disabled, precluded summary judgment on employee's ADA claim); *Wilson v. Md.–Nat'l Capital Park & Planning Comm'n,* 178 F.3d 1289 (4th Cir.1999) (accepting district court assumption that bladder cancer is a disability); *Berk v. Bates Adver. USA, Inc.,* 25 F.Supp.2d 265 (S.D.N.Y.1998) (holding breast cancer is a disability); *Olbrot v. Denny's, Inc.,* 1998 WL 525174 (N.D.Ill. Aug.19, 1998) (holding issue of whether breast cancer is a disability is a question of fact); *Bizelli v. Amchem,* 981 F.Supp. 1254 (E.D.Mo.1997) (holding testicular cancer is a disability); *Mark v. Burke Rehab. Hosp.,* 1997 WL 189124 (S.D.N.Y. Apr.17, 1997) (holding lymphoma is a disability); *Wojciechowski v. Emergency Technical Servs. Corp.,* 1997 WL 164004 (N.D.Ill. Mar.27, 1997) (accepting breast cancer in deceased plaintiff as a disability); *Milton v. Bob Maddox Chrysler Plymouth, Inc.,* 868 F.Supp. 320 (S.D.Ga.1994) (finding summary judgment inappropriate in case involving bronchial cancer); *Braverman v. Penobscot Shoe Co.,* 859 F.Supp. 596 (D.Me.1994) (finding summary judgment inappropriate in case alleging prostate cancer as a disability).

Id.

physicians testified "that her particular type of breast cancer would put her life at risk if she became pregnant."[157]

2. Record of Having an Impairment That Substantially Limits a Major Life Activity

At least one appellate court recognized that cancer treatment can create a record of having an impairment that substantially limits a major life activity by correctly applying the EEOC regulations to the evidence.[158] Joan Eshelman was diagnosed with breast cancer after working for the same company for seventeen years.[159] She took six months of medical leave for treatment, including chemotherapy, and then returned to work part-time.[160] Like many survivors, Eshelman suffered from "chemo brain" in the form of short-term memory loss.[161] Initially, her employer agreed that Eshelman could perform the essential functions of her job.[162] In light of her memory loss, however, she sought reasonable accommodations, including the ability to telecommute at times.[163] After she requested accommodations, Eshelman was laid off in a staff reduction.[164] The court allowed Eshelman to proceed to trial where a jury awarded Eshelman $170,000 in back pay and $30,000 in compensatory damages.[165]

The Third Circuit affirmed, finding that Eshelman's cancer treatment established a record of impairment upon which her employer relied in deciding to terminate her employment.[166] It rejected the argument that Eshelman's work limitations were too temporary to be covered by the ADA based on three factors:[167] First, the court held that "Eshelman was, unquestionably, substantially impaired in the major life activity of working during her six-month absence for cancer treatment."[168] Second, her employer knew about her cancer, chemotherapy, and surgery.[169] Third, Eshelman's cancer affected her ability to work beyond "a relatively

157. Berk v. Bates Adver. USA, Inc., 25 F. Supp. 2d 265, 268 (S.D.N.Y. 1998).
158. Eshelman v. Agere Sys., Inc., 554 F.3d 426, 438–39 (3d Cir. 2009).
159. *Id.* at 430.
160. *Id.*
161. *Id.* at 430–31; *see also supra* note 21.
162. 554 F.3d 426 at 431.
163. *Id.*
164. *Id.*
165. *Id.* at 432.
166. *Id.* at 436.
167. *Id.* at 437.
168. *Id.*
169. *Id.*

short-term absence from work"[170] because she received cancer-related care before and after her medical leave, experienced cognitive problems when she returned to work, and her supervisors knew about her medical care and its side effects.[171] "Paired with Eshelman's six-month absence and [her employer's] knowledge of her condition, this cognitive dysfunction permitted the jury to conclude that Eshelman had demonstrated a record of impairment that substantially limited her ability to think and work" and had shown that her employer impermissibly relied upon her cancer experience as a factor in deciding to lay her off.[172]

Another appellate case recognized that cancer in remission can establish a record of a substantially limiting impairment.[173] The Fifth Circuit, which typically embraces the cancer survivor's Catch-22, reversed summary judgment for an employer in a case that egregiously ignored the purpose of the ADA.[174] After working his way up from salesman to president of a company over twenty years, Michael Boyle was diagnosed with leukemia.[175] He was hospitalized for thirty days to receive chemotherapy.[176] His doctor then declared his cancer to be in remission and gave him permission to return to work without limitations.[177] When Boyle returned to work the next week, his employer demoted him and reduced his salary.[178] The trial court granted Boyle's employer summary judgment, concluding that he was not disabled, did not have a record of a disability, and was not perceived as having a disability.[179] The court held that Boyle's month-long hospitalization for chemotherapy did not raise a question of material fact as to whether he had a record of a disability.[180] The Fifth Circuit reversed and remanded, instructing the trial court "to determine whether the record of Boyle's impairment includes a substantial effect on a major life activity."[181] Quoting the EEOC's interpretative regulations that the

170. *Id.*

171. *Id.* at 438.

172. *Id.*

173. EEOC v. Gallagher, 181 F.3d 645, 656–57 (5th Cir. 1999).

174. *Id.* at 656–57 (reversing summary judgment for employer because whether an employee whose blood cancer was in remission had a record of a substantial limitations on any major life activities was a question for the jury).

175. EEOC v. Gallagher, 959 F. Supp. 405, 406–407 (S.D. Tex. 1997), *aff'd in part, vacated in part*, 181 F.3d 645 (5th Cir. 1999).

176. *Id.* at 407.

177. *Id.*

178. *Id.*

179. *Id.* at 409.

180. *Id.*

181. EEOC v. Gallagher, 181 F.3d 645, 656 (5th Cir. 1999).

ADA "protects former cancer patients from discrimination on the basis of their prior medical history," the Fifth Circuit concluded that although a cancer history does not per se prove discrimination based on a record of a disability, the trial court erroneously failed to consider whether Boyle's cancer history established a record of a disability.[182]

Finally, at least one trial court case correctly held that hospitalization constitutes evidence of a record of a disability. Although most courts declined to hold that hospitalization alone could prove a record of a disability, the District Court for the Eastern District of Missouri found that John Bizelli's hospitalization for cancer surgery raised sufficient evidence of a record of a disability to preclude summary judgment.[183] Bizelli took six months of medical leave for testicular cancer treatment.[184] After his physician cleared him to work with lifting restrictions, his employer refused to allow him to return to work unless he successfully completed a physical examination to which no other employee was subject.[185] The trial court held that Bizelli's cancer history established a record of an impairment,[186] and that the ADA "[p]lainly . . . intended to ensure that former cancer patients are not discriminated against on the basis of their prior medical history."[187] The trial court appropriately permitted a jury to consider whether Bizelli was discriminated against because of his cancer.[188]

3. Regarded as Having an Impairment That Substantially Limits a Major Life Activity

Only those survivors who provided very strong evidence of an employer's misconceptions about how their cancer affected major life activities survived pretrial motions on whether their employer

182. *Id.* at 655–56.

183. Bizelli v. Amchem, 981 F. Supp. 1254, 1257 (E.D. Mo. 1997) (recognizing that Bizelli's surgery and treatment for testicular cancer established a record of a disability).

184. *Id.* at 1255–56.

185. *Id.* at 1256.

186. *Id.* at 1257.

187. *Id.*

188. *Id.* The jury found that Bizelli's employer breached its duty to accommodate Bizelli's return to work with a temporary lifting restriction and that he was fired because of his testicular cancer history. Bizelli v. Parker Amchem, 17 F. Supp. 2d 949, 951 (E.D. Mo. 1998). It awarded him lost wages and compensatory damages. *Id.*

regarded their cancer as substantially limiting a major life activity.[189]

In *Eshelman v. Agere Systems, Inc.*, the Third Circuit affirmed the jury's determination that Eshelman's employer discriminated against her because it regarded her breast cancer treatment as substantially limiting a major life activity.[190] The Third Circuit agreed with Eshelman's argument that the jury had sufficient evidence upon which to conclude "that although she had excelled at her job and was a highly valued employee," her employer "nonetheless erroneously viewed her memory impairment caused by her chemotherapy treatment as substantially limiting two major life activities; namely, Eshelman's ability to think and her ability to work."[191]

Similarly, a trial court in the Western District of New York denied summary judgment for an employer who fired a cancer survivor because she raised a material dispute concerning whether her employer regarded her cancer treatment as substantially limiting her ability to do her job.[192] Donna Shandrew worked as a phlebotomist for Quest Diagnostics for twenty-three years; she was fired two months after completing chemotherapy.[193] The trial court held that Shandrew alleged sufficient evidence for a jury to consider "that Quest mistakenly believed her cancer and/or treatment substantially limited her in the major life activity of working"[194] by preventing her from working in "a broad class of jobs."[195] The court found that Shandrew raised a genuine issue of material fact that her supervisor "treated her differently when she returned from her medical leave, making twenty or more comments to her in regards to her chemotherapy treatment making her slow and forgetful, and suggesting she should stay home and rest and/or take pain medica-

189. *See* Eshelman v. Agere Sys., Inc., 554 F.3d 426, 436 (3d Cir. 2009) (affirming jury verdict for cancer survivor on the grounds that her employer regarded her memory problems caused by chemotherapy as substantially limiting her ability to think).

190. *Id.* (finding that plaintiff presented sufficient evidence for a reasonable fact finder to conclude that the employer regarded her as substantially limited in a major life activity). The court also held that Eshelman's employer discriminated against her because of her record of an impairment. *See id.* at 439.

191. *Id.* at 434, 436 (finding that the jury had reasonable evidence to conclude that Eshelman's employer perceived that her cancer-related memory problems created substantial limitations in her ability to work).

192. Shandrew v. Quest Diagnostics Inc., 819 F. Supp. 2d 181, 188 (W.D.N.Y. 2011).

193. *Id.* at 183–184.

194. *Id.* at 187.

195. *Id.*

tion."[196] Considered in the light most favorable to Shandrew, these allegations indicated that her supervisor may have believed that she was unable to adequately perform her job duties because of her cancer treatment.[197]

Similarly, at least one appellate court held that an employer can regard cancer in remission as a disability.[198] Mark Olds worked as a UPS driver for sixteen years when he was diagnosed with multiple myeloma.[199] UPS denied his request to return to work with weight-lifting restrictions after his cancer treatment.[200] In granting summary judgment to UPS, the trial court held that UPS did not regard Olds's cancer history as an impairment because it denied his accommodation requests, even though his restrictions resulted from his cancer treatment.[201] The Sixth Circuit correctly reversed, finding that Olds presented evidence as to how UPS perceived his cancer history to affect his job qualifications.[202] Thus, a "reasonable jury could infer that UPS believed that Olds'[s] condition was significantly more disabling than it actually was, and for that reason UPS did not want to reinstate him."[203]

IV.
HOW CANCER SURVIVORS FARED IN EMPLOYMENT DISCRIMINATION CASES UNDER THE ADA AMENDMENTS ACT

A. Congress Attempts to Plug Holes in the ADA with the Americans with Disabilities Act Amendments Act

The often dismal results Title I complainants experienced in federal court prompted Congress to reject and repair the Supreme Court's narrow construction that emasculated the ADA.[204] In hear-

196. *Id.*

197. *Id.*

198. Olds v. United Parcel Serv., 127 Fed. App'x 779, 782 (6th Cir. 2005) (holding that the plaintiff raised a genuine issue of material fact as to whether his employer regarded his post-cancer condition as disabling).

199. *Id.* at 780–81.

200. *Id.*

201. *Id.* at 781.

202. *Id.* at 783.

203. *Id.* Olds presented evidence that a UPS "lawyer said that Olds should not be reinstated because he had cancer" and that a UPS employee admitted that " 'we didn't want to come right out and say that because of Mark's condition, his cancerous condition, that he couldn't do the job.' " *Id.*

204. For evidence that the Supreme Court narrowly interpreted the ADA, see Sutton v. United Air Lines, Inc., 527 U.S. 471, 482–84 (1999) (rejecting EEOC regulations that instructed courts to determine whether an individual had a disa-

ings on the ADAAA, the House Committee on Education and Labor declared that it intended "to lessen the standard of establishing whether an individual has a disability for purposes of coverage under the ADA, and to refocus the question on whether discrimination on the basis of disability occurred."[205] The Committee illustrated why congressional action was needed to ease the burden of proving standing under the ADA with the story of a cancer survivor whose case was dismissed on summary judgment:

> The Committee expects that the bill will affect cases such as . . . Pimental v. Dartmouth-Hitchcock Clinic, 236 F. Supp. 2d 177 (D.N.H. 2002) in which the court concluded that the plaintiff's stage three breast cancer did not substantially limit her ability to care for herself, sleep, or concentrate.[206] The Committee expects that the plaintiffs in each of these cases could establish a material restriction on major bodily functions that would qualify them for protection under the ADA."[207]

The ADAAA, which took effect on January 1, 2009, retained the three alternate definitions for the term "disability,"[208] but lowered the threshold for determining whether a person is "disabled" under the first and second alternatives. Congress explicitly stated that one purpose of the ADAAA was:

---

bility without regard to mitigating measures); Albertsons, Inc. v. Kirkinburg, 527 U.S. 555, 556 (1999) (holding that mitigating measures "must be taken into account in judging whether an individual possesses a disability").

205. H. REP. NO. 110-730, pt. 1, at 7 (2008).

206. The court trapped Pimental in a Catch-22 because she did not claim that her cancer substantially limited her ability to work. It held that Pimental's "own assertions that the cancer did not substantially impair her ability to perform various tasks associated with her employment tend[ed] to undermine her claim that it did substantially affect her ability to, for example, care for herself on a long-term basis. *See, e.g.*, Memorandum of Law in Support of Plaintiff's Objection to Defendant's Motion for Summary Judgement, Pimenta v. Dartmouth-Hitchcock Clinic, 236 F. Supp. 2d 177 (2002) (No. C.01-292-M) (stating that, upon her return from medical leave, plaintiff "had no problems performing her duties as a nurse"). Thus, she has failed to demonstrate that her illness substantially affected her ability to care for herself, sleep, or to concentrate on a permanent or long-term basis." 236 F. Supp. 2d at 183–84.

207. H. REP. NO. 110-730, pt. 1, at 12 (footnote 206 added).

208. "The term 'disability' means, with respect to an individual—(A) a physical or mental impairment that substantially limits one or more of the major life activities of such individual; (B) a record of such an impairment; or (C) being regarded as having such an impairment." ADA Amendments Act of 2008, Pub. L. No. 110-325, sec. 4(a), § 3(1), 122 Stat. 3553 (codified at 42 U.S.C. § 12102(1) (Supp. V 2011)). The ADAAA further added a statutory definition of "being regarded as having such an impairment." *Id.* at sec. 4(a), § 3(1)(C).

> [T]o convey congressional intent that the standard created by the Supreme Court in the case of Toyota Motor Manufacturing, Kentucky, Inc. v. Williams, 534 U.S. 184 (2002) for "substantially limits," and applied by lower courts in numerous decisions, has created an inappropriately high level of limitation necessary to obtain coverage under the ADA, to convey that it is the intent of Congress that the primary object of attention in cases brought under the ADA should be whether entities covered under the ADA have complied with their obligations, and to convey that the question of whether an individual's impairment is a disability under the ADA should not demand extensive analysis.[209]

The ADAAA made four key changes to the plaintiff's burden of proof. First, the ADAAA expands the list of "major life activities" covered. In an end-run around the Supreme Court's unwillingness to defer to the EEOC regulations,[210] Congress added the definition of "major life activities" to the statute itself.[211] The ADAAA broadens the definition of "major life activity" to include "the operation of a major bodily function, including but not limited to, functions of the immune system, normal cell growth, digestive, bowel, bladder, neurological, brain, respiratory, circulatory, endocrine, and reproductive functions."[212] Almost all cancers substantially limit one or more of those major bodily functions.[213] Additionally, the ADAAA supplements the nonexhaustive list of "major life activities" by adding "eating, sleeping, walking, standing, lifting, bending, speaking, breathing, learning, reading, concentrating, thinking, communicating and working."[214] This far more expansive definition of "major life activities" should allow practically all cancer survivors to identify a major life activity affected by their cancer.

Second, the ADAAA explicitly renounced the Supreme Court's decision that a court must consider whether an individual used mitigating measures, such as taking medication, in determining

209. *Id.* at sec. 2(b)(5).

210. *See* Sutton v. United Air Lines, Inc., 527 U.S. 471, 482–84 (1999).

211. ADA Amendments Act, sec. 4(a), § 3(2) (codified at 42 U.S.C. § 12102(2) (Supp. V 2011)).

212. *Id.* at sec. 4(a), § 3(2)(B) (codified at 42 U.S.C. § 12102(2)(B) (Supp. V 2011)).

213. *See* 29 C.F.R. § 1630, app. § 1630.2(i) (2011) ("Because impairments, by definition, affect the functioning of body systems, they will generally affect major bodily functions. For example, cancer affects an individual's normal cell growth . . . .").

214. *Id.* at sec. 4(a), § 3(2)(A) (codified at 42. U.S.C. § 12102(2)(A) (Supp. V 2011)).

whether he or she had a disability:[215] "The determination of whether an impairment substantially limits a major life activity shall be made without regard to the ameliorative effects of mitigating measures such as . . . medication . . . or prosthetics."[216] Under this language, a court may no longer consider how cancer treatment mitigated the effects of cancer on an individual. For example, a sarcoma survivor whose leg was amputated, but who can walk with the use of a prosthetic leg, has a disability because her leg was amputated. A court may no longer consider how well the survivor could walk with the prosthesis. Similarly, antiemetics may help a survivor whose chemotherapy-induced nausea substantially limits his ability to eat. A court may no longer consider how well the survivor could eat while taking antinausea medication.

Third, prior to the ADAAA, a person whose impairment was episodic or in remission would be unlikely to prove that it substantially limited a major life activity.[217] Thus, a survivor whose cancer was in remission or whose cancer only occasionally affected a major life activity may not have been covered under the ADA.[218] The ADAAA addresses this obstacle by specifying that an "impairment that is episodic or in remission is a disability if it would substantially limit a major life activity when active."[219] This language will benefit the large numbers of cancer survivors whose cancer is chronic, but often managed.[220] Many survivors live for years or decades with

215. The ADAAA explicitly "reject[s] the requirement enunciated by the Supreme Court in Sutton v. United Air Lines, Inc., 527 U.S. 471 (1999) and its companion cases that whether an impairment substantially limits a major life activity is to be determined with reference to the ameliorative effects of mitigating measures." ADA Amendments Act, sec. 2(b)(2).

216. *Id.* at sec. 4(a), § 3(4)(E)(i) (codified at 42 U.S.C. § 12102(4)(E)(i) (Supp. V 2011)).

217. The Supreme Court held that, under the original ADA, "[to] be substantially limited . . . an individual must have an impairment that prevents or severely restricts the individual from doing activities that are of central importance to most people's daily lives. The impairment's impact must also be permanent or long term." Toyota Motor Mfg., Kentucky, Inc. v. Williams, 534 U.S. 184, 198 (2002) (citing 29 C.F.R. §§ 1630.2(j)(2)(ii)–(iii)(2001)).

218. *See, e.g.*, Olmeda v. N.Y. State Dep't of Civil Serv., No. 96 Civ. 7557(HB), 1998 WL 17729, at *2 (S.D.N.Y. Jan. 16, 1998).

219. ADA Amendments Act, sec. 4(a), § 3(4)(D) (codified at 42 U.S.C. § 12102(4)(D) (Supp. V 2011)). *See also* discussion *infra* text accompanying notes 254–62 (discussing cases applying this provision).

220. For example, some cancers, like certain lymphomas, are not curable, yet many individuals live for years or decades with these cancers as a chronic condition. *See, e.g.*, Bradley & Bednarck, *supra* note 21 at 188 ("[C]ertain types of cancer are no longer perceived as terminal illnesses, but instead, chronic diseases that require regular monitoring . . . treatment . . . and life style modification.").

their cancer, and at times are not substantially limited by their diagnoses. Now even those survivors whose cancer is successfully treated in fewer than six months are covered by the ADAAA.[221]

Fourth, under the original ADA, an employee could be covered if his or her employer "regarded" him or her as having a disability that affected a major life activity.[222] The ADAAA no longer requires that the employer actually believe that the employee is substantially limited in a major life activity; a burden of proof few plaintiffs could meet. Now the employee need only prove that "he or she has been subjected to an action prohibited under [the ADA] because of an actual or perceived physical or mental impairment whether or not the impairment limits or is perceived to limit a major life activity."[223] Thus, cancer survivors who prove that they are treated differently because of their cancer, regardless of whether their cancer substantially limits any major life activity, may be protected under the ADAAA.

B. The EEOC Regulations to Title I of the ADAAA

As exemplified by the Sutton trilogy,[224] federal courts often ignored the EEOC regulations and interpretive guidance when determining whether a plaintiff has a disability.[225] Yet, as a practical matter, Congress generally outlines major policy goals in statutes and defers to the expertise of administrative agencies like the EEOC to provide detailed definitions and illustrations for judicial guidance.[226] Accordingly, the EEOC issued extensive regulations to further define the ADAAA.[227] The ADAAA regulations take direct

221. The EEOC regulations to the ADAAA instruct that "[t]he effects of an impairment lasting or expected to last fewer than six months can be substantially limiting." 29 C.F.R. § 1630.2(j)(1)(ix) (2012). The ADAAA overturns the Supreme Court's holding in *Toyota Motor Manufacturing Kentucky, Inc.* that "the impairment's impact must also be permanent or long term." *Compare id.*, *with* Toyota Motor Mfg. Ky., Inc. v. Williams, 534 U.S. 184, 198 (2002).

222. 42 U.S.C. § 12102(2)(C) (2006).

223. ADA Amendments Act, sec. 4(a), § 3(3)(A) (codified at 42 U.S.C. § 12102(3)(A) (Supp. V 2011)).

224. The *Sutton* trilogy is composed of three cases. *Cf.* Sutton v. United Air Lines, Inc., 527 U.S. 471 (1999); Murphy v. United Parcel Service, Inc., 527 U.S. 516 (1999); Albertson's, Inc. v. Kirkingburg, 527 U.S. 555 (1999).

225. See Evan Sauer, *The ADA Amendments Act of 2008: The Mitigating Measures Issues, No Longer A Catch-22*, 36 OHIO N.U. L. REV. 215, 233–34 (2010).

226. *See id.* at 234.

227. The EEOC regulations to Title I of the ADAAA were issued on March 25, 2011, and took effect on May 24, 2011. 29 C.F.R. § 1630.1(a)(c)(4) (2011).

Instead of redefining "substantially limits" in the ADAAA, Congress added "rules of construction" to the statute to instruct courts that they should interpret

aim at the federal court decisions that restricted a plaintiff's ability to prove that he or she has a disability as defined by Title I:[228]

> Broad coverage. The primary purpose of the ADAAA is to make it easier for people with disabilities to obtain protection under the ADA. Consistent with the Amendments Act's purpose of reinstating a broad scope of protection under the ADA, the definition of "disability" in this part shall be construed broadly in favor of expansive coverage to the maximum extent permitted by the terms of the ADA. The primary object of attention in cases brought under the ADA should be whether covered entities have complied with their obligations and whether discrimination has occurred, not whether the individual meets the definition of disability. The question of whether an individual meets the definition of disability under this part should not demand extensive analysis.[229]

The EEOC also expressly recognized that cancer survivors are individuals with a disability. In a clear message to federal courts that Congress intended cancer survivors to have standing under the ADA, the regulations direct that "it should easily be concluded that the following types of impairments will, at a minimum, substantially limit the major life activities indicated: . . . cancer substantially limits normal cell growth."[230]

---

the term "disability" broadly. ADA Amendments Act, sec. 4(a), § 3(4) (codified at 42 U.S.C. § 12102(4) (Supp. V 2011)). Congress then gave the EEOC the authority to interpret and illustrate these rules to provide further guidance to the courts and litigants. *Id.* at sec. 6(a)(2), § 506 (codified at 42 U.S.C. § 12205a (Supp. V 2011)); *see also* E. Pierce Blue, *Arguing Disability Under the ADA Amendments Act: Where Do We Stand?*, FED. LAW, Dec. 2012, at 38, 39–41 (describing the rules of construction adopted by the EEOC that "provide a fuller picture of the relationship between the changes made in the law, the repudiation of *Sutton* and *Toyota*, and the broad interpretation of 'substantially limits'"). Although the regulations list conditions that presumptively prove disability, the ADAAA and its regulations retain a court's obligation to assess whether a particular plaintiff has a disability on an individualized basis. 29 C.F.R. § 1630.2(j)(1)(iv) (2012) ("The determination of whether an impairment substantially limits a major life activity requires an individualized assessment.").

228. For a discussion of the Supreme Court's weakening of employee rights under the ADA, see *The Cancer Survivors' Catch-22*, *supra* note 8, at 413–32.

229. 29 C.F.R. § 1630.1(c)(4) (2012).

230. 29 C.F.R. § 1630.2(j)(3)(iii) (2012). "The determination of whether an impairment substantially limits a major life activity requires an individualized assessment. However, in making this assessment, the term 'substantially limits' shall be interpreted and applied to require a degree of functional limitation that is lower than the standard for 'substantially limits' applied prior to the ADAAA." § 1630.2(j)(1)(iv).

The Appendix to the EEOC regulations is replete with examples of how Congress intended to cover cancer survivors under Title I of the revised statute.

1. Definition of an Individual with a Disability

The Appendix recognizes that most plaintiffs should be able to identify the link between their impairment and the affected bodily functions "because impairments, by definition, affect . . . bodily functions."[231] It illustrates this relationship by acknowledging that "cancer affects an individual's normal cell growth."[232] In referencing the legislative history of the ADAAA, the Appendix refers to the House Education and Labor Committee's intention that the inclusion of major bodily functions would "affect cases such as . . . *Pimental v. Dartmouth–Hitchcock Clinic*, in which the court concluded that the plaintiff's stage three breast cancer did not substantially limit her ability to care for herself, sleep, or concentrate."[233] The Appendix also explains that a plaintiff need not prove that multiple bodily functions are affected by a disability, illustrated by the following example: "An individual whose normal cell growth is substantially limited due to lung cancer need not also show that she is substantially limited in breathing or respiratory function."[234]

2. Impairments in Remission

One of the most beneficial descriptions in the Appendix explains how cancer survivors in remission would be protected by the ADAAA: "An impairment that is episodic or in remission is a disability if it would substantially limit a major life activity in its active state."[235] The Appendix again refers to Congress's disapproval of cases like *Pimental v. Dartmouth–Hitchcock Clinic*,[236]

> where the courts have discounted the impact of an impairment [such as cancer] that may be in remission as too short-lived to be substantially limiting. It is thus expected that individuals with impairments that are episodic or in remission (e.g., epilepsy, multiple sclerosis, cancer) will be able to establish coverage if, when active, the impairment or the manner in which it

231. 29 C.F.R. § 1630, app. § 1630.2(i) (2011).
232. *Id.*
233. *Id.*
234. § 1630, app. § 1630.2(j)(1)(viii).
235. § 1630, app. § 1630.2(j)(1)(vii).
236. 236 F. Supp. 2d 177, 182–83 (D.N.H. 2002).

> manifests (e.g., seizures) substantially limits a major life activity.[237]

The Appendix also relies on the legislative history of the ADA to support its interpretation of the ADAAA regarding impairments in remission. In quoting the testimony of Representative Steny Hoyer, an original cosponsor of the ADA, the Appendix notes that the ADA drafters "could not have fathomed that people with . . . cancer . . . and other disabilities would have their ADA claims denied because they would be considered too functional to meet the definition of disability."[238] For example, individuals with cancers that have poor long-term prognoses such as ovarian cancer, metastatic breast cancer, and non-Hodgkin's lymphoma may be substantially limited in their ability to work or concentrate while suffering the side effects of chemotherapy. But many survivors who have these types of cancer experience long periods of remission when they are not substantially limited in any majorly life activity, even though their cancers are likely to return. Thus, survivors who experience a remission, regardless of the length or permanency of that remission, have standing under the ADAAA.

3. Record of a Substantially Limiting Impairment

The Appendix describes that the purpose of the second prong of the definition of "disability"—an individual with a record of an impairment that substantially limits or limited a major life activity—is to prohibit discrimination based on a history of a disability. In recognizing that some cancer survivors face discrimination long after their treatment is completed, the Appendix states that "the 'record of' provision would protect an individual who was treated for cancer ten years ago but who is now deemed by a doctor to be free of cancer from discrimination based on that prior medical history."[239] The Appendix also explains that some cancer survivors have standing under both the "disability" prong and the "record of" prong:

> This is a consequence of the rule of construction in the ADAAA and the regulations providing that an individual with an impairment that is episodic or in remission can be protected under the first prong if the impairment would be substantially limiting when active. See 42 U.S.C. 12102(4)(D);

237. 29 C.F.R. § 1630, app. § 1630.2(j)(1)(vii) (quoting H. REP. NO. 110-730, pt. 2, at 19–20 (2008)).

238. § 1630, app. § 1630.2(j)(3).

239. § 1630, app. § 1630.2(k).

> § 1630.2(j)(1)(vii). Thus, an individual who has cancer that is currently in remission is an individual with a disability under the "actual disability" prong because he has an impairment that would substantially limit normal cell growth when active. He is also covered by the "record of" prong based on his history of having had an impairment that substantially limited normal cell growth.[240]

4. Regarded as Having a Substantially Limiting Impairment

Finally, the Appendix provides a cancer example to demonstrate that the amended "regarded as" prong should be applied in a practical, straightforward manner: "[I]f an employer terminates an employee because he has cancer, the employer has regarded the employee as an individual with a disability."[241] As noted above, courts should not consider whether the impairment substantially limits a major life activity under the "regarded as" prong, unlike under the first prong.

In sum, the EEOC reforms promulgated under the ADAAA ensure that cancer survivors have standing under the statute. Employers' defenses to a Title I claim must now properly focus not on whether the plaintiff has a disability, but on whether he or she is qualified to perform the essential functions of the job and on whether the employer offered a reasonable accommodation.[242]

C. Cancer Survivors' Employment Discrimination Cases Under the ADAAA

Because Congress did not intend to apply the ADAAA retroactively,[243] cases involving causes of action that accrued prior to January 1, 2009, applied the original ADA. Once the ADAAA took effect, plaintiffs and their advocates readily tested how the EEOC would apply the statute's reduced burden of proving disability status. Unsurprisingly, Title I complaints increased by 31.8%, from 19,453 in 2008 to 25,742 in 2011.[244] Similarly, the number of cancer-based

240. *Id.*

241. § 1630, app. § 1630.2(l) (2011).

242. The ADA protects only "qualified individuals" with a disability. *See* 42 U.S.C. § 12112(a) (2006) ("No covered entity shall discriminate against a *qualified* individual . . . .") (emphasis added).

243. *See, e.g.*, Melone v. Paul Evert's RV Country, Inc., 455 Fed. App'x 738, 739 n.1 (9th Cir. 2011); EEOC v. Agro Distrib., LLC, 555 F.3d 462, 469 n.8 (5th Cir. 2009).

244. EEOC, *ADA Receipts*, *supra* note 41.

claims increased by 34.5%, from 707 in 2008 to 951 in 2011.[245] The number of cancer-based employment discrimination claims that have been resolved favorably for the plaintiff has nearly doubled since the passage of the ADAAA.[246]

A significant number of cancer survivors who sued under the ADAAA, in contrast to those who sued under the ADA, have successfully survived pretrial motions so that courts can address the merits of their claims for employment discrimination. Most telling, however, is that no reported case interpreting the ADAAA has imposed a Catch-22 on a plaintiff cancer survivor by holding that the plaintiff is too healthy to be disabled, yet too ill to work.[247]

As Congress intended,[248] judicial review of cancer survivors as Title I plaintiffs has focused more on whether the plaintiff was qualified for his or her job yet faced discrimination because of a disability, than whether the plaintiff had standing under the ADAAA. For example, the District Court for the Northern District of Georgia relied on the ADAAA regulations that the determination of whether a plaintiff has a disability should "be construed broadly in favor of expansive coverage" in finding that a breast cancer survivor had a disability because her cancer substantially limited her normal cell growth.[249] Another trial court in the Southern District of New York also relied on the ADAAA regulations to hold that "[c]ancer will 'virtually always' be a qualifying disability."[250]

---

245. EEOC, *ADA Charges*, *supra* note 31.

246. The total number of cancer charges with outcomes favorable to the charging parties and/or charges with meritorious allegations increased from 135 in 2007 to 257 in 2011. *See supra* Table 2.

247. This Article surveyed cases reported through January 1, 2013.

248. H. REP. NO. 110-730, pt. 1, at 7.

249. Coker v. Enhanced Senior Living, Inc., 897 F. Supp. 2d 1366, 1374 (N.D. Ga. 2012) (quoting 29 C.F.R. § 1630.2(j)(1)(i) (2012)) (quoting from regulations stating that "cancer substantially limits normal cell growth").

250. Katz v. Adecco USA, Inc., 845 F. Supp. 2d 539, 548 (S.D.N.Y. 2012) (denying summary judgment to an employer and employment agency that had impermissibly asked a job applicant about her medical history in violation of the ADA, and then failed to offer her a job upon discovering that she was a cancer survivor). The ADAAA regulations provide:

> [T]he individualized assessment of some types of impairments will, in virtually all cases, result in a determination of coverage under paragraphs (g)(1)(i) (the "actual disability" prong) or (g)(1)(ii) (the "record of" prong) of this section. Given their inherent nature, these types of impairments will, as a factual matter, virtually always be found to impose a substantial limitation on a major life activity. Therefore, with respect to these types of impairments, the necessary individualized assessment should be particularly simple and straightforward.

29 C.F.R. § 1630.2(j)(3)(ii).

Remarkably, in some ADAAA cases, the parties *agreed* that the plaintiff had a disability,[251] a rare concession in cases under the ADA.[252] Other ADAAA cases did not directly analyze whether the plaintiff had a disability, but instead addressed issues on the assumption that the plaintiff had a disability.[253] Those survivors whose ADAAA claims were dismissed in pretrial motions appropriately lost because they failed to prove that they were qualified for their jobs or were treated differently because of their cancer, not because they failed to prove that they had a disability.[254]

251. Diehl v. Bank of Am., N.A., 470 Fed. App'x 771, 775 (11th Cir. 2012) (employer did not dispute that plaintiff's breast cancer was a disability); Garity v. APWU Nat'l AFL-CIO, 2:11–CV–01109–PMP–CWH, 2012 WL 2273429, at *6, *8 (D. Nev. June 18, 2012) (denying the employer's motion to dismiss because the parties agreed that the cancer survivor had a disability and because she alleged sufficient facts to show her disability was related to adverse treatment by her employer); Pauling v. Gates, No. 1:10cv1196 (LMB/JFA), 2011 WL 1790137, at *4 n.4 (E.D. Va. May 6, 2011) (parties conceded that the plaintiff, who had been treated for cancer for five years, had a disability).

252. In cases prior to the ADA, defendants often conceded that a plaintiff had an impairment, although they seldom conceded that a plaintiff had a disability.

253. *See* Shelton v. Bridgestone Metalpha, U.S.A., Inc., 3–11–0001, 2012 WL 1609670, at *5 (M.D. Tenn. May 8, 2012) (denying summary judgment to the employer of a thyroid cancer survivor who claimed her employer did not reasonably accommodate her); Estate of Reed v. Ponder Enterprises, Inc., 1:11cv554–CSC, 2012 WL 1031487, at *7 (M.D. Ala. Mar. 27, 2012) (denying the employer's motion to dismiss in finding that "the complaint reasonably support[s] the inference that Crystal Reed was terminated not for poor job performance, but because of her disability and/or in retaliation for her request for accommodations"); Jacobsen v. Dillon Companies, Inc., 10–CV–01944–LTB–BNB, 2012 WL 638122, at *1, *11 (D. Colo. Feb. 28, 2012) (denying summary judgment to the employer of a breast cancer survivor who was fired while undergoing treatment); Moore v. Md. Dep't of Pub. Safety & Corr. Servs., Civil No. CCB–11–553, 2011 WL 4101139, at *4 (D. Md. Sept. 12, 2011) (denying summary judgment to the employer on the issue of whether an employee who took an eight-month leave of absence for breast cancer treatment was entitled to medical leave as a reasonable accommodation, especially given that she had been cleared to return to work just one day after she lost her job).

254. *See, e.g.*, Valdez v. McGill, 462 F. App'x 814, 818–19 (10th Cir. 2012) (implying that the employee's colon cancer was a disability in granting summary judgment to the employer because the employee failed to "show a reasonable accommodation would allow him to perform the essential functions of his job"); Angell v. Fairmount Fire Prot. Dist., No. 11–cv–03025–CMA–CBS, 2012 WL 5389777, at *4–5 (D. Colo. Nov. 5, 2012) (concluding that the plaintiff cancer survivor had a disability but nevertheless granting summary judgment to the defendant because the plaintiff did not prove that he was discriminated against because of his disability); Haley v. Cohen & Steers Capital Mgmt., Inc., 871 F. Supp. 2d 944, 955–56 (N.D. Cal. 2012) (granting summary judgment to the employer because the employee failed to prove a link between her lymphoma and the alleged adverse employment action).

The ADAAA has been particularly helpful to survivors who can prove they were able to work, in part, because their cancer treatment was successful. Although plaintiffs alleging discrimination based on cancer in remission seldom survived summary judgment under the ADA, several recent decisions have recognized that cancer in remission is a disability under the ADAAA.[255]

For example, the District Court for the Eastern District of Pennsylvania held that a plaintiff whose cancer is in remission has a disability. Richard Unangst was laid off following his successful treatment for lymphoma.[256] Although the parties agreed that Unangst had a disability, the court explained why this conclusion was mandated by the ADAAA:

> The ADA was clearly intended by Congress to protect cancer patients from disability discrimination. *See* H. Rep. No. 101–485, pt. 3, at 29 (1990). Cancer is a "paradigmatic example of such an impairment." *Adams v. Rice,* 531 F.3d 936, 952 (D.C.Cir. 2008). Plaintiff has further demonstrated that his chemotherapy treatment substantially limited his ability to perform major life activities, due largely to the fatigue and nausea he experienced as a result of the treatment. Plaintiff's cancer

255. The District Court for the Eastern District of Pennsylvania also held that Michael Chalfont stated a claim that he had a disability based on a history of leukemia. Chalfont v. U.S. Electrodes, No. 10–2929, 2010 WL 5341846, at *9 (E.D. Pa. Dec. 28, 2010). Chalfont took leave from his job to receive chemotherapy from October 2007 until May 2008. *Id.* at *2. He was fired eight months after he returned to work when his cancer was in remission. *Id.* The court held that Chalfont had a disability because he took a seven-month leave of absence to receive chemotherapy for leukemia, which, though in remission when he was fired, "at times cause[d] him to be fatigued and subject to easy bleeding and bruising" and substantially limited "the major life activity of normal cell growth and circulatory function." *Id.* at *9.

Other district courts have similarly held that cancer in remission is a disability under the ADAAA. The District Court for the Northern District of Indiana held that Stephen Hoffman's renal cell carcinoma, which was in remission when he was fired, was a disability under the ADAAA. Hoffman v. Carefirst of Fort Wayne, Inc., 737 F. Supp. 2d 976, 985 (N.D. Ind. 2012). Hoffman claimed that his employer would not accommodate his doctor's advice that he work no more than forty hours per week. *Id.* at 982. The court rejected his employer's argument that Hoffman was not disabled because his cancer was in remission and he could work forty hours per week. *Id.* at 984–85. Citing the EEOC regulations, the court held that "under the ADAAA, because Hoffman had cancer in remission (and that cancer would have substantially limited a major life activity when it was active), Hoffman does not need to show that he was substantially limited in a major life activity at the actual time of the alleged adverse employment action." *Id.* at 985.

256. Unangst v. Dual Temp Co., Inc., No. 10–6811, 2012 WL 931130, at *2 (E.D. Pa. Mar. 19, 2012).

> diagnosis in late 2008 qualified him for the protections of the ADA at that time. Plaintiff was cancer-free as of February 2009, and cleared to return to work without restrictions.[257]

The trial court then correctly held that Unangst's cancer, though in remission when he was laid off, "would substantially limit a major life activity when active."[258] Thus, the court held that Unangst had a disability as defined by the ADA.[259]

Even within the Fifth Circuit, the District Court for the Eastern District of Texas recognized that a plaintiff whose cancer was in remission could be classified as disabled under the ADAAA.[260] One month after he returned to work as a sales manager from leave to receive treatment for kidney cancer, Michael Norton was fired.[261] The court recognized that under the ADAAA, "renal cancer is capable of qualifying as a disability under the ADA" because "when active," it "'substantially limits' the 'major life activity' of 'normal cell growth.'"[262] Relying on the EEOC regulations and interpretive guidance, the court found that it was of "no consequence" that plaintiff's cancer was in remission at the time of his alleged discrimination.[263]

Simply pleading that the plaintiff is a person with a disability because she had cancer, however, is not sufficient to state a claim under the ADAAA. The plaintiff still must allege sufficient facts to show how her cancer substantially limited a major life activity. Several cancer survivors have failed to earn their day in court because of insufficient pleadings.

LaTanya Brandon alleged that the school where she taught failed to accommodate her fatigue resulting from cancer treat-

257. *Id.* at *4 (citations omitted). In *Adams v. Rice*, the plaintiff alleged that the federal government (U.S. Department of State) denied her a position that required "world-wide availability" because she had been treated for breast cancer, thus the Rehabilitation Act, not the ADA, governed her claim. 531 F.3d 936, 939. The court held that Adams did not have a disability because she was "cancer-free" when she was rejected from the Foreign Service. *Id.* at 944. But it denied summary judgment to the State Department on its argument that Adams did not have a record of a disability. *Id.* at 954. The court found that Adams raised sufficient material facts to show that her breast cancer created a record of an impairment because it substantially limited her major life activity of engaging in sexual relations. *Id.* at 949.

258. *Id.*

259. *Id.*

260. Norton v. Assisted Living Concepts, 786 F. Supp. 2d 1173, 1185 (E.D. Tex. 2011).

261. *Id.* at 1178.

262. *Id.* at 1185.

263. *Id.* at 1185–86 (denying employer's motion for summary judgment).

ment.[264] Even though the court acknowledged that the ADAAA lightens the burden of proof on the plaintiff, it held that "it remains the case that 'not every impairment will constitute a disability'" and the plaintiff must still prove a substantial limitation.[265] In dismissing Brandon's complaint, the court found that Brandon's vague pleadings were insufficient to prove disability:

> Brandon has not alleged any facts other than that, upon returning to work, she "would experience fatigue" and "was not to engage in lifting objects." The latter allegation seems to suggest an impairment of a "major life activity"—lifting—identified by the ADA as amended. However, without any additional factual detail, it is virtually impossible to determine whether Brandon's impairment limited her ability to lift objects "as compared to most people in the general population." . . . But absent any details as to how or why Brandon was limited in lifting—let alone any description of what kind of cancer she had or what kind of treatment she received or for how long—the Court simply cannot determine whether eighteen months after returning from medical leave Brandon was any more impaired in lifting her room supplies than any other teacher or any other person. That is not to say that the Court doubts that Brandon's cancer proved extremely difficult for her. But to bring a legal claim against the Academy under the ADA, Brandon must provide some minimal details about what her condition was, what treatment she received for it, how long the treatment lasted, and how it affected her. She has not done so.[266]

In a similar case, the court granted summary judgment to Virginia Larson's employer because Larson failed to offer any "evidence beyond her bare assertion" that her employer, who fired her four months after she concluded treatment for breast cancer, regarded her breast cancer as a disability.[267] Lawson did not dispute her employer's claim "that he did not know that she had a disability and he did not regard her as having a disability."[268]

Finally, Michael O'Connell failed to state a viable claim that his cancer was a disability because he did not claim that his cancer sub-

264. Brandon v. O'Mara, 10 Civ. 5174 (RJH), 2011 WL 4478492, at *1 (S.D.N.Y. Sept. 28, 2011).

265. *Id.* at *7 (quoting 29 C.F.R. § 1630.2(j)(1)(ii)(2011)).

266. *Id.* (footnote omitted).

267. Larson v. Del. Highlands AL Servs. Provider, LLC, No. 10–2295–KHV, 2012 WL 1415521, at *5 (D. Kan. Apr. 24, 2012).

268. *Id.*

stantially limited any major life activities.[269] O'Connell told his employer that he had undergone surgery to remove his testicle and a malignant tumor from his neck.[270] O'Connell complained to an owner of the business that one of his supervisors repeatedly made disparaging comments about his surgery.[271] In his amended complaint, O'Connell alleged that the owner failed for two years to reprimand his supervisor, that the disparaging remarks caused stress that interfered with his ability to work, and that he was ultimately fired for complaining about his supervisor.[272] In granting the employer's motion to dismiss, the court ruled that O'Connell's amended complaint did "not plausibly allege that O'Connell's cancer limited any major life activities."[273]

V.
LESSONS FOR PLAINTIFFS' ATTORNEYS

The introduction to this Article references Cathy Jamison, the protagonist of the television series *The Big C*, cautioning her student's parent that if he insisted that she be fired because of her melanoma, he "better hire a damn good lawyer."[274] If Jamison were fired and decided to file a Title I complaint, she, too, would need a damn good lawyer to ensure that her claim would survive summary judgment.[275] So what lessons have we learned from cancer survivors' experiences as Title I claimants? How should plaintiffs' attorneys craft a complaint to maximize the chance that their clients meet their burden of proving they have a disability, are regarded as having a disability, or have a record of a disability as defined by the ADAAA?[276]

First, plead details.[277] For purposes of stating a claim, a court must accept the plaintiff's factual narrative as true.[278] Specify when

269. O'Connell v. Cont'l Elec. Constr. Co., No. 11 C 2291, 2011 WL 4916464, at *3 (N.D. Ill. Oct. 17, 2011) (dismissing the claim that the plaintiff was fired because of a disability).

270. *Id*. at *1.

271. *Id*.

272. *Id*. at *1–2.

273. *Id*. at *3.

274. *See The Big C: The Little C*, *supra* note 1.

275. *Id.*

276. These suggestions are relevant to plaintiffs with most impairments, not exclusively to those with cancer.

277. See *supra* text accompanying notes 264–73 (summarizing cases in which the plaintiff failed to allege sufficient facts to state an ADAAA claim).

278. A court must accept well-pled facts as true, though it can reject a plaintiff's legal conclusions. Ashcroft v. Iqbal, 556 U.S. 662, 677 (2009) ("[T]he tenet

the plaintiff was diagnosed and with what type of cancer, what type of medical care the plaintiff received, how the plaintiff's cancer and its treatment affected the plaintiff's daily life, and every action the employer took that could possibly be related to the plaintiff's cancer history. In pleadings and discovery, flesh out the plaintiff's narrative by including expert testimony regarding the plaintiff's medical care—including prognosis, late effects, and side effects—as well as the plaintiff's posttreatment physical and mental abilities.

Thus, Cathy Jamison's complaint should detail:

- when she was diagnosed with cancer;
- the type of cancer and stage of diagnosis;
- the types of treatment she received;
- her physicians' recommendations for her care;
- the side effects and late effects of her treatment; and
- every employment event that could possibly be related to her cancer.

Second, plead that the plaintiff's cancer substantially limited as many major life activities as are supported by the facts.[279] Almost all cancers substantially limit "the operation of a major bodily function, including . . . functions of the immune system [and] normal cell growth."[280] Moreover, during and after cancer treatment, many survivors are substantially limited in "eating, sleeping, standing, lifting, bending, reading, concentrating, thinking, [or] communicating."[281] Never focus a complaint solely on how cancer substantially limits the plaintiff's ability to work. Although some survivors may prove that their cancer substantially limits their ability to work, working is the most problematic of the major life activities and should be pled only as an alternative argument.

Thus, Cathy Jamison should allege that her melanoma substantially limited not only her ability to work as a swim coach, but also that it substantially limited:

---

that a court must accept as true all of the allegations contained in a complaint is inapplicable to legal conclusions.").

279. Where Title I plaintiffs fail to allege which major life activities their cancer has effected, courts typically analyze whether the plaintiff was substantially limited in the major life activity of work. *See, e.g.*, Burnett v. LFW, Inc., 472 F.3d 471, 483–84 (7th Cir. 2006).

280. 29 C.F.R. § 1630.2(i)(1)(ii); *see also* § 1630.2 (j)(3)(iii).

281. 42 U.S.C. § 12102(2)(A) (Supp. V 2011); *see also* CANCER CARE FOR THE WHOLE PATIENT: MEETING PSYCHOSOCIAL HEALTH NEEDS 26–29 (Nancy Adler & Ann Page eds., 2008) (summarizing "cancer-induced physical stressors); Bradley & Bednarek, *supra* note 21, at 196 (noting some cancer survivors had some degree of disability with multiple workplace limitations as a result of cancer and its treatment).

- the functions of her immune system;
- her normal cell growth; and
- her ability to sleep, concentrate, and think.

Third, describe how a survivor is substantially limited "without regard to the ameliorative effects of mitigating measures such as . . . medication . . . or prosthetics."[282] Mitigating measures such as chemotherapy and other medications, radiation therapy, rehabilitation therapy, and prosthetics can improve the quality of a cancer survivor's life. Their effects are irrelevant, however, in evaluating whether a survivor is disabled. Thus, Cathy Jamison should allege that although medication helped alleviate the nausea and fatigue caused by her treatment, without the mitigating effects of that medication, she would be substantially limited in at least one major life activity.

Fourth, allege that the plaintiff is covered by each potentially relevant, alternate definition of a person with a disability.[283] For example, cancer survivors whose cancer was in remission at the time of the alleged discriminatory job action should claim both that they have a disability and that their cancer history establishes a record of impairment.[284] Explain how the plaintiff's cancer "would substantially limit a major life activity when active."[285] Thus, even if Cathy Jamison's melanoma were in remission when she was fired as a swim coach, she should allege that her melanoma would have substantially limited a major life activity if it were active.

Fifth, cancer survivors who were cancer-free at the time of the adverse employment action, regardless of when they were initially diagnosed, should plead that their employer regarded them as having a disability.[286] Simply put, explain how the employer adversely treated a cancer survivor "because" of his or her cancer.[287] Thus, if

282. § 12102(4)(E).

283. *See supra* note 208.

284. *See* 29 C.F.R. § 1630, app. § 1630.2(k) (2011). *See, e.g.*, Farrish v. Carolina Commercial Heat Treating, Inc., 225 F. Supp. 2d 632, 636 (M.D.N.C. 2002) (holding that the plaintiff failed to show that the cancer in remission substantially limited a major life activity—therefore not proving a disability—while not considering sua sponte whether the plaintiff had a record of a disability); Alderdice v. Amer. Health Holding, Inc., 118 F. Supp. 2d 856, 863 (S.D. Ohio 2000) (granting summary judgment to the employer where the plaintiff, whose cancer was in remission, did not allege that she had a record of a disability).

285. 42 U.S.C. § 12102(4)(D) (Supp. V 2011).

286. *See generally* 29 C.F.R. app. § 1630.2(l) (explaining that through the ADAAA, Congress broadened the application of the "regarded as" prong so a plaintiff need prove only that an employer treated an employee adversely because of an impairment).

287. § 1630, app. § 1630.2(l).

the producers of *The Big C* had cured Cathy Jamison of her melanoma, she could allege that her high school fired her because it capitulated to pressure from parents to replace "the lady with cancer" as their children's swim coach, despite her coaching skills and successes.

CONCLUSION

Today's Congress has been called everything from gridlocked to broken to dysfunctional for failing to see beyond intransigent partisanship to pass legislation.[288] Congress is often called upon to amend ineffective or judicially emasculated legislation. Yet it seldom answers that call. But when judicial interpretation of the ADA ignored the statute's purpose to provide remedies to disability-based employment discrimination and created significant barriers for cancer survivors to prove disability status under Title I, plaintiffs' advocates had no choice but to turn to Congress to amend the statute and to the EEOC to revise Title I regulations. So is the ADAAA a law of intended consequences? Did Congress and the EEOC fix a broken ADA? Remarkably, the early cases suggest that they did.[289]

288. *See, e.g.*, Ezra Klein, *Fourteen Reasons Why This Is the Worst Congress Ever*, WASHINGTON POST, July 13, 2012, 8:00 AM, http://www.washingtonpost.com/blogs/wonkblog/wp/2012/07/13/13-reasons-why-this-is-the-worst-congress-ever/.

Congressional inaction has been ripe fodder for comedians. For example, Jimmy Fallon joked: "Congress was broadcast live on Facebook for the first time in history. Now you can waste time and not get work done by watching Congress waste time and not get work done." Kevin G. Barkes, *Political Jokes of the Week*, KGB REPORT (Jan. 07, 2011, 8:51 AM), http://www.kgbreport.com/archives/political-jokes-of-the-week/index.shtml.

289. This Article was drafted less than four years after the passage of the ADAAA and less than two years after issuance of the new EEOC regulations. Commentators are just beginning to assess the effectiveness of the ADAAA for all individuals who face disability-based discrimination. The initial assessments seem to agree with the conclusions drawn in this Article. One commentator suggests that plaintiffs with "impairments that were previously denied coverage have survived the summary judgment phase of litigation [under the ADAAA] and anecdotal evidence indicates that it is now more common for parties to agree that a person is covered under the ADA and move to other issues." E. Pierce Blue, *supra* note 228, at 38. Similarly, another commentator optimistically predicts that mediators will easily conclude whether the plaintiff has a disability and shift the focus of mediation to which "job functions are essential, whether an accommodation is indeed reasonable, the depth to which the parties engaged in the interactive process, and to what extent the defense of direct threat is applicable." Mark Travis, *A Change in Focus—Mediation of Claims Under the ADA Amendments Act*, DISP. RESOL. MAG., Spring 2012, at 33, 36.

FACEBOOK USED TAKEDOWN AND IT WAS SUPER EFFECTIVE! FINDING A FRAMEWORK FOR PROTECTING USER RIGHTS OF EXPRESSION ON SOCIAL NETWORKING SITES*

KEVIN PARK†

Introduction ..... 892 R
I. Facebook, I Choose You: A Description of Social Networking Sites ..... 903 R
A. Social Networking Sites Have Become Important Sites for Communication in Contemporary Society Due to Rapid Growth ..... 904 R
B. Social Networking Sites Act as Censors and Filter Both Content and Users ..... 911 R

* This title draws inspiration from a popular meme called "It's Super Effective!" based on the widely popular Pokémon videogame. *See* Felipe Almeida Mendes, *It's Super Effective!*, KNOW YOUR MEME, http://knowyourmeme.com/memes/its-super-effective (last visited Dec. 31, 2013). Memes are cultural units that gain value through expression. *See Meme*, Wikipedia, http://en.wikipedia.org/w/index.php?title=Meme&oldid=586747709 (last updated Dec. 19, 2013, 5:56 AM). "It's Super Effective!" is an example of the cultural unit called the "meme," defined by Jack Balkin as a form of speech that competes for recognition and retention in the audience's memory. *See* Jack Balkin, Freedom of Speech, From a Meme's Point of View 2 (no date) (unpublished manuscript) (on file with author).

The section headings throughout this Note continue the theme of the Pokémon meme invoked in the title by drawing on other references from the original Pokémon television show, including the title of the first episode ("Pokémon, I Choose You!") and Team Rocket mottos ("Prepare for trouble!" "Make it double!" "To protect the world from devastation!"). *See Pokémon, I Choose You!*, WIKIPEDIA, http://en.wikipedia.org/w/index.php?title=Pok%C3%A9mon,_I_Choose_You!&oldid=566053545 (last updated July 27, 2013, 7:04 PM); *Team Rocket Mottos*, BULBAPEDIA, http://bulbapedia.bulbagarden.net/w/index.php?title=Team_Rocket_mottos&oldid=2047638 (last updated Dec. 30, 2013, 11:35 AM).

† J.D., 2013, New York University School of Law; B.A and M.A., 2010, Johns Hopkins University. Executive Articles Editor for the *N.Y.U. Annual Survey of American Law* 2012–13. Special thanks to Professor Diane Zimmerman, Professor Helen Hershkoff, Felicity Kohn, Stephen Elkind, Ava McAlpin, Sarah Siegel, Kelly Parker, and the N.Y.U. Annual Survey of American Law for all of their helpful comments and assistance. I would also like to thank my family for all of their support.

II. Prepare for Trouble and Make It Double: Analyzing Two Frameworks for Protecting User Rights on Social Networking Sites ..... 914 R
A. The First Amendment Does Not Provide Users with Direct Protection Against Censorship by Social Networking Sites ..... 915 R
B. Consumer Protection Rationales Provide a Basis for Intervention When Market Failures Exist ..... 925 R
C. Market Failures Arise in the Social Networking Site Market, Allowing Consumer Protection to Provide a Basis for Protecting User Rights of Expression Without the First Amendment ..... 929 R
D. Social Networking Sites Are More Aptly Characterized as Common Carriers Than Newspapers, Although in Reality They Fall Somewhere in Between Those Poles ..... 936 R
III. To Protect the World from Censorship: Evaluating Potential Regulations of Social Networking Sites Within a Consumer Protection Framework ..... 944 R
A. Establishing Clearer Content Regulation Standards ..... 946 R
B. Forcing Greater Disclosure on Censorship Policies ..... 947 R
C. Implementing Non-Litigation-Based Remedies ... 949 R
D. Creating a Cause of Action ..... 951 R
E. Eliminating Censorship Authority ..... 952 R
Conclusion ..... 955 R

INTRODUCTION

In October 2011, Facebook[1] deactivated Courtney Stodden's account.[2] The takedown e-mail alleged that Stodden, an eighteen-

1. This Note focuses on Facebook because it is the largest social networking website and the social networking site on which most people spend their time. *See* Alex Fitzpatrick, *Pew: Social Media Not Yet Driving News Traffic*, MASHABLE (Mar. 19, 2012), http://mashable.com/2012/03/19/pew-state-of-media-technology (indicating that social networking users spent, on average, 423 minutes on Facebook during the month of December 2011, in comparison to 151 minutes on Tumblr and 80 minutes on Pinterest); *see also infra* Part I.A. This Note will largely draw on the types of services that Facebook offers when considering social networking site issues. I recognize that other social networking sites are structured differently, but I would argue that they all seek the same goals with differences arising largely in the specific tools utilized.

2. *See* Sarah Ann Hughes, *Courtney Stodden's Facebook Taken Down, Restored*, WASH. POST BLOG (Oct. 14, 2011, 8:04 AM), http://www.washingtonpost.com/

year-old quasi-celebrity best known for marrying Doug Hutchison,[3] had posted "inappropriate sexual conduct" on her fan page.[4] This "conduct" consisted of self-portraits in which Stodden wore only a bikini and stood in sexually suggestive poses, a type of photo that is actually fairly common on the site.[5] When Facebook deleted Stodden's account, it cut off her ability to connect to other people on the platform and to express herself by sharing self-generated content. The social networking site took away Stodden's access without providing her with any notice or opportunity to contest the decision. In response, Stodden's mother drew media attention to the takedown, leading to the reactivation of Stodden's account and to an apology from Facebook claiming that the account deletion had occurred in error.[6]

While Stodden's scenario may appear trivial, it illustrates one way in which a social networking site can limit an individual's self-expression on the Internet.[7] Facebook and other social networking platforms allow users to post information and communicate with

blogs/celebritology/post/courtney-stoddens-facebook-taken-down-restored/2011/10/14/gIQAN12bjL_blog.html.

3. Stodden's marriage to Doug Hutchison was particularly notable due to the 35 year age difference. Angela Ellis & Sabina Ghebremedhin, *Exclusive: Doug Hutchison, 51, and Courtney Stodden, 16, Dish on Controversial Marriage*, ABC NEWS (July 15, 2011), http://abcnews.go.com/Entertainment/doug-hutchison-courtney-stodden-controversial-marriage-exclusive/story?id=14073130#.UImO2Ibme3w. Hutchison achieved fame for his roles in *The Green Mile*, *The X-Files*, *Lost* and *24*. *See Doug Hutchison*, WIKIPEDIA, http://en.wikipedia.org/w/index.php?title=Doug_Hutchison&oldid=586078143 (last updated Dec. 14, 2013, 6:35 PM).

4. *See* Hughes, *supra* note 2.

5. *See* Danica Daniel, *Teen Bride Courtney Stodden Kicked off Facebook for Being "Too Sexy." Mom Blames "Jealousy."*, DRJAYS.COM (Oct. 19, 2011, 2:05 PM), http://live.drjays.com/index.php/2011/10/19/teen-bride-courtney-stodden-kicked-off-facebook-for-being-too-sexy-mom-blames-jealousy.

6. *See id.* (describing Facebook's response as an "oops").

7. Social networking sites exist along a continuum of different models. *See* Tal Z. Zarsky, *Law and Online Social Networks: Mapping the Challenges and Promises of User-Generated Information Flows*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 741, 746–47 (2008) (describing various metrics for categorizing social networking sites and adopting the "strength of ties" metric); Alex Iskold, *Evolution of Communication: From Email to Twitter and Beyond*, READWRITEWEB (May 30, 2007), http://readwrite.com/2007/05/30/evolution_of_communication (providing a graphic breakdown of social media and Internet tools generally based on speed and type of communication); Dan Saffer, *The Continuum of Online Communication*, ADAPTIVE PATH (May 21, 2007), http://www.adaptivepath.com/ideas/the-continuum-of-online-communication (placing social media sites on a spectrum from dictation-based to conversation-based).

other account holders.[8] In addition to self-generated content, social networking sites also allow users to link content from other places on the Internet and to comment on other users' activity, creating a highly interactive online environment.[9] Users can post content in a variety of forms, including text and videos,[10] and preserve discussions for the future.[11] Social networking sites and other Internet media have changed the way that people communicate with each other by increasing the ease of long-distance communication, compressing communication into short entries, establishing the expectation of quicker responses, and allowing people to create virtual identities that use different social cues from real life.[12] Social networking sites have also changed the content of communication, as these platforms encourage users to share personal information, sometimes even information that would previously have been considered taboo or sensitive.[13]

8. *See Social Networking Service*, WIKIPEDIA, http://en.wikipedia.org/w/index.php?title=Social_networking_service&oldid=587124250 (last updated Dec. 21, 2013, 6:28 PM).

9. *See How to Post & Share*, FACEBOOK, http://www.facebook.com/help/sharing (last visited Dec. 31, 2013) (explaining how users can share information on Facebook). Other sites provide more limited types of sharing (e.g. Twitter) or focus on particular types of content, such as the professional connections offered by LinkedIn. *See The Beginner's Guide to LinkedIn*, MASHABLE, http://mashable.com/2012/05/23/linkedin-beginners/ (last updated Oct. 2013) (providing an overview of how to use LinkedIn and directing its focus to professional network development); *The Beginner's Guide to Twitter*, MASHABLE, http://mashable.com/2012/06/05/twitter-for-beginners/ (last updated Nov. 2013) (noting that Twitter focuses primarily on short, text-based information sharing).

10. Robert Young, *Social Networks are the New Media*, GIGAOM (May 29, 2006, 10:00 PM), http://gigaom.com/2006/05/29/social-networks-are-the-new-media (mentioning the existence of video-sharing platforms as far back as 2006).

11. *Cf.* Robert Young, *The Future of Social Networks – Communication*, GIGAOM (Oct. 9, 2006, 7:00 AM), http://gigaom.com/2006/10/09/the-future-of-social-networks-communication (describing Facebook's "wall" concept as a type of "bulletin board").

12. *See* THE NEW MEDIA CONSORTIUM, SOCIAL NETWORKING, THE "THIRD PLACE," AND THE EVOLUTION OF COMMUNICATION 1–3 (2007), *available at* http://www.nmc.org/pdf/Evolution-of-Communication.pdf. For more discussion of the characteristics of social networking sites, see *infra* Part I.A.

13. *See, e.g.*, Maureen Linke, *Women Turn to Social Media for Support After Pregnancy Loss*, USA TODAY (Mar. 28, 2012, 6:52 PM), http://yourlife.usatoday.com/parenting-family/story/2012-03-28/Women-turn-to-social-media-for-support-after-pregnancy-loss/53837714/1 (describing how more women are sharing the sensitive topic of pregnancy loss on online social networks to find support).

In order to communicate with others, users "friend" or follow other account holders, creating an online social network.[14] Users most often make connections with other people that they know from the real world.[15] Users can build a more extensive network both by finding new first-point contacts and by identifying second-point contacts from their existing friends.[16] In other instances, users connect to others based solely on common interests, such as politics or hobbies, or similar life experiences.[17] Occasionally, users even connect with complete strangers.[18] As membership in social networking sites grows, these websites are becoming more of a central means of communication.

Social networking sites have also gained greater importance by providing communication tools for more than just social interaction. Businesses use social networking sites to reach consumers through exclusive online-only deals,[19] and to increase brand expo-

14. *Social Network*, OXFORD DICTIONARIES, http://www.oxforddictionaries.com/definition/english/social-network (last visited Dec. 31, 2013); *Social Networking Services*, *supra* note 8.

15. *See, e.g.*, *Sign Up*, FACEBOOK, https://en-gb.facebook.com/ (last visited Dec. 31, 2013) ("Facebook helps you connect and share with the people in your life.").

16. *See* Danah M. Boyd & Nicole B. Ellison, *Social Network Sites: Definition, History, and Scholarship*, 13 J. OF COMPUTER-MEDIATED COMM. 210, 211–14 (2007), *available at* http://onlinelibrary.wiley.com/store/10.1111/j.1083-6101.2007.00393.x/asset/j.1083-6101.2007.00393.x.pdf?v=1&t=hpvfa786&s=76d95ae56f5d21c489446cadbd3f0916878716e6 (detailing how social networking websites help people create networks and find contacts).

17. *See, e.g.*, Linke, *supra* note 13 (noting that some online forums cater to women who have gone through pregnancy loss to connect them with each other for support); *see also What is Social Networking?*, WHATISSOCIALNETWORKING.COM, http://www.whatissocialnetworking.com (last visited Dec. 31, 2013) (suggesting that people connect based on common interests such as golfing or gardening).

18. *See, e.g.*, *Social Networking with Strangers*, LOUISGRAY.COM (Dec. 22, 2011), http://blog.louisgray.com/2011/12/social-networking-with-strangers.html (indicating how users can encounter strangers on social networking sites). Some social networking sites are even built on the idea of connecting users to complete strangers. *See, e.g.*, *Friends & Strangers*, TWITTER.COM, https://twitter.com/friend-strangers (last visited Dec. 31, 2013) (Twitter page of the former social networking site, Friends & Strangers, which explicitly promoted using the network to make "new connections").

19. *See, e.g.*, Macy's, *Exclusive Offer for Macy's Fans!*, FACEBOOK, http://www.facebook.com/events/151914784852726 (last visited Dec. 31, 2013) (giving Facebook users an extra 10% or 20% discount); The Resort at the Mountain, *May Facebook Exclusive Offer*, FACEBOOK (Apr. 29, 2011, 2:58 PM), http://www.facebook.com/note.php?note_id=10150164011661780 (offering a buy one, get one free offer for Facebook fans).

sure.[20] Employers have also started to conduct social networking background checks before tendering offers of employment.[21] Moreover, social networking sites play a role in political elections by providing candidates with a new medium to reach the electorate and by tending to increase users' levels of engagement with political discourse.[22] Twitter, in particular, allows users to react to political debates and discussions in real time, to share their thoughts on the candidates, and to engage more generally in the political process.[23] These functions illustrate that social networking sites provide individuals both with an important forum for self-expression and with a platform for connecting to a broad array of speech in society at large. Congruently, access to social networking sites is vital to an individual's ability to obtain information in an increasingly digital society.[24]

---

20. *See* Dan Schawbel, *Major Findings from the 2010 Social Media Marketing Industry Report*, SOCIAL MEDIA TODAY (May 1, 2010), [hereinafter Schawbel, *Major Findings*], http://socialmediatoday.com/danshwabel/101703/major-findings-2010-social-media-marketing-industry-report.

21. *See* Craig Kanalley, *The Growth of Social Media*, HUFFINGTON POST, http://www.huffingtonpost.com/2011/09/01/growth-social-media-infographic_n_945256.html (last updated Nov. 1, 2011, 6:12 AM) (infographic providing statistics on various uses of social networking sites).

Some states have proposed or enacted legislation that limits the ability of employers to require social networking background checks or release of social networking site login information from applicants. *See, e.g.*, 820 ILL. COMP. STAT. 55/10 (2013).

22. Harrison Weber, *Social Media Matters More Than Ever in This US Presidential Election*, THE NEXT WEB (Jan. 31, 2012), http://thenextweb.com/us/2012/01/31/social-media-matters-more-than-ever-in-this-us-presidential-election (stating that 62% of social networking site users expected the 2012 presidential candidates to have some social networking site presence); *see* Lauren Dugan, *Social Media in the 2010 US Midterm Election: What Worked (and What Didn't)*, SOCIAL TIMES (Nov. 3, 2010, 2:03 PM), http://socialtimes.com/social-media-in-the-2010-us-midterm-election-what-worked-and-what-didnt_b27242 (describing how social networking sites contributed to the 2010 midterm elections, most notably finding that numerous social networking sites focused on encouraging voter turnout).

23. *See* Joann Pan, *Twitter to Play Crucial Role in South Carolina Republican Debate*, MASHABLE (Jan. 16. 2012), http://mashable.com/2012/01/16/fox-twitter-republican-debate.

24. *See, e.g.*, Chris Godley, *THR's Social Media Poll: How Facebook and Twitter Impact the Entertainment Industry*, THE HOLLYWOOD REPORTER (Mar. 21, 2012, 11:53 AM), http://www.hollywoodreporter.com/gallery/facebook-twitter-social-media-study-302273#12 (citing a study by THR and Penn Schoen Berland in which 19% of respondents said that social networking sites are their primary source of breaking news); *see also* Fitzpatrick, *Pew: Social Media Not Yet Driving News Traffic*, *supra* note 1 (indicating that social media is not yet the primary source of news information for social media users but that these websites have the potential to become huge sources of information).

Despite their significant communication-enabling functions, social networking sites sometimes limit communication by filtering content or even users. In order to obtain access to a social networking site, users must agree to terms of service that give the platform certain types of control, including the right to take down material or to delete user accounts.[25] Although the ability to take down content or to delete accounts raises concerns about censorship, existing law is fairly settled on allowing social networking sites to engage in such behavior without consequence.[26] This censorial authority exists because the federal government has created a safe harbor for "interactive computer service" providers who voluntarily take down certain types of material from the Internet.[27] Section 230(c)(2) of the Communications Decency Act of 1996 states:

> No provider or user of an interactive computer service shall be held liable on account of . . . any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected . . . .[28]

This statute originally was enacted to resolve ambiguity on the question of online service provider liability for defamatory statements published by a user that harmed another private party and on what steps a website could take to insulate itself from legal action.[29] Prior to the statute, some scholars were concerned that web-

25. *See Facebook Community Standards*, FACEBOOK, http://www.facebook.com/communitystandards (last visited Dec. 31, 2013) (providing an overview of Facebook's content guidelines); *Statement of Rights and Responsibilities*, FACEBOOK (Nov. 15, 2013), http://www.facebook.com/legal/terms. The right to take down material is usually reserved in the terms of service to which users agree when creating an account. *See, e.g.*, *id.*

26. *See* Communications Decency Act of 1996, 47 U.S.C. § 230(c)(2) (2006) (providing interactive service providers with immunity for good faith takedowns of certain content); Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11, 11 (2006) (describing how the government has used private actors to regulate the Internet); *cf.* Christopher S. Yoo, *Free Speech and the Myth of the Internet as an Unintermediated Experience*, 78 GEO. WASH. L. REV. 697, 702 (2010) (indicating that Internet intermediary use is inevitable).

27. 47 U.S.C. § 230(c)(2).

28. *Id.*

29. *See* Susan Freiwald, *Comparative Institutional Analysis in Cyberspace: The Case of Intermediary Liability for Defamation*, 14 HARV. J.L. & TECH. 569, 588–96 (2001) (detailing pre-1996 liability for Internet entities and the enactment of the Communications Decency Act of 1996 as a response); *see also* Anthony Ciolli, *Chilling Ef-*

sites would attempt to avoid liability by over-censoring and taking down all content that appeared to be even marginally defamatory or problematic.[30] Internet provider industry lobbyists also presented legislators with the other extreme possibility of complete content non-discrimination, as an earlier district court case had found such non-intervention could avoid treatment as publishers of particular content.[31] The statute rectified this scenario by providing a safe harbor for takedown of certain types of content done in good faith on particular types of websites and by mandating that websites would not be treated as "publishers" of third-party content.[32] However, the broad statutory language allowed courts to expand the group of websites eligible for immunity, including online social networking sites, and to apply immunity to causes of action other than defamation.[33] The statute also permits websites to make their own content-regulation norms without external oversight.[34]

fects: The Communications Decency Act and the Online Marketplace of Ideas, 63 U. MIAMI. L. REV. 137, 146–47 (2008) (providing background for the *Cubby* and *Stratton Oaks* cases). *Compare* Cubby, Inc. v. CompuServe, Inc., 776 F. Supp. 135, 137–38, 140, 143 (S.D.N.Y. 1991) (finding no liability because CompuServe did not exercise editorial control over its content (i.e., it was a passive service provider)), *with* Stratton Oakmont, Inc. v. Prodigy Servs. Co., 1995 WL 323710, at *4–5, *7 (N.Y. Sup. Ct. May 24, 1995) (finding a message board liable as a publisher of defamatory statements because it actively monitored site content).

The language of the statute mentions "users" of interactive computer service providers as also having immunity. 47 U.S.C. § 230(c). However, court cases have generally focused on "providers" rather than "users." *See* sources cited *infra* note 33.

30. *See* Ciolli, *supra* note 29, at 148 (explaining that scholars at the time were concerned about the risk of over-censorship to prevent publication of any potentially harmful content).

31. *See* Cubby, 776 F. Supp. at 139–40; Freiwald, *supra* note 29, at 594 (noting that Internet service providers told Congressmen that they would adopt a policy of complete nondiscrimination in order to obtain "passive" website status).

32. *See* 47 U.S.C. § 230(c). Scholars did not completely agree on the specific purpose of the statutory immunity provision. *Compare* Ciolli, *supra* note 29, at 147–48 (stating that immunity was intended to discourage either or both under- or over-censorship), *with* Freiwald, *supra* note 29, at 595 (stating that the Communications Decency Act of 1996 was intended to allow intermediaries to choose their own level of monitoring, in the hopes of increasing it).

33. *See, e.g.*, Doe v. MySpace, Inc., 474 F. Supp. 2d 843, 849–50 (W.D. Tex. 2007) (finding that MySpace was entitled to application of the Communications Decency Act of 1996 immunity provision); *see also* H. Brian Holland, *In Defense of Online Intermediary Immunity: Facilitating Communities of Modified Exceptionalism*, 56 U. KAN. L. REV. 369, 373–75 (2008) (describing how courts have interpreted "interactive computer services" broadly and expanded the application of the immunity provisions beyond simple third-party defamation cases).

34. *See* Holland, *supra* note 33, at 369–70 (arguing that immunity provisions allow for online communities to develop their own standards).

Once a social networking site takes down content or deletes an account, the user has limited recourse. These limitations are usually incorporated into the platform's terms of service. For instance, Facebook's terms of service state that claims against it must be litigated in "state or federal court located in Santa Clara County."[35] Due to the high costs of litigation, a user is more often limited to an informal appeal to the social networking site. (The media pressure that appears to have caused the reversal by Facebook in Courtney Stodden's case is probably not an avenue of response available to non-celebrities.) Such an appeal may be effective in certain cases, such as when Facebook deletes a user's account on the belief that the user has assumed a false name,[36] but not in more difficult cases when the issue involves a judgment about whether certain content falls into the amorphous category of "otherwise objectionable" subject matter. These appeals are not subject to any type of external review, making it difficult to check the social network's decision-making process for consistency and fairness.

For constitutionally unprotected speech, the social networking site's censorial power seems acceptable. For instance, if a user were to upload hardcore pornographic images to his or her Facebook

---

Many people agree that social networking sites should have the ability to take down defamatory or obscene material, particularly because of the risk that children can come across such content. *See, e.g.*, Ciolli, *supra* note 29, at 265 (stating that "immunity from vicarious liability in [defamation] tort actions . . . remains necessary"); Usman Qazi, The Internet Censorship Controversy (May 9, 1996) (unpublished manuscript), *available at* http://courses.cs.vt.edu/cs3604/lib/Censorship/notes.html (noting concerns about children's access to pornography and violent material through the Internet).

Some level of censorship may provide positive benefits in helping to filter the amount of information available on the Internet. That discussion is beyond the scope of this Note, and I will assume that some level of censorship will exist for the online social networking context due to the existing case law. For a further discussion on the debate over censorship, see Holland, *supra* note 33, at 369–70, 391 (immunity to censor allows online communities to develop their own standards and engage in collaborative production); Zarsky, *supra* note 7, at 778–79 (stating that filtration of content helps users sift through massive amounts of information and limits manipulation).

35. *Statement of Rights and Responsibilities*, *supra* note 25. Terms may also restrict users to non-litigation-based remedies. *See generally* Amalia D. Kessler, *Stuck in Arbitration*, N.Y. TIMES, Mar. 6, 2012, http://www.nytimes.com/2012/03/07/opinion/stuck-in-arbitration.html (noting that many commercial contracts include arbitration provisions).

36. *See, e.g.*, *Facebook Asks Aaditya Thackeray for ID Proof*, NDTV.COM (Feb. 24, 2012), http://www.ndtv.com/article/cities/facebook-asks-aaditya-thackeray-for-id-proof-179437 [hereinafter *Facebook Asks*] (detailing the deletion and restoration by Facebook of an Indian politician's account).

profile, neither the courts nor the general population is likely to deny that Facebook has a general right to remove such content, and perhaps even the user's account. However, social networking sites take down even constitutionally protected forms of speech.[37] In those cases, despite contractually limited options, users may believe that the First Amendment protections for freedom of expression provide a basis to challenge a social networking site's censorship decision. However, the First Amendment applies to only state action.[38] Therefore it will only apply to a private entity's actions in limited situations, such as when its actions meet the state action doctrine, which requires "the conduct allegedly causing the deprivation of a federal right [to be] fairly attributable to the State."[39] Under the current state of free speech doctrine, it is unlikely that a private social networking site would be considered a state actor.[40]

Alternatively, a user could attempt to invoke the "public forum doctrine," which limits speech restrictions for particular types of fora.[41] However, that doctrine would also be insufficient to provide a basis for applying First Amendment protections to social networking sites, as the websites clearly are not traditional public fora and

37. *Compare* Ashcroft v. Free Speech Coalition, 535 U.S. 234, 245–46 (2002) (outlining the categories of unprotected speech for purposes of the First Amendment), *with Facebook Community Standards*, *supra* note 25 (providing general categories of objectionable content). In particular, terms in the Facebook Community Standards such as "bullying," "abusive behavior," "harassment," "hate speech," or "graphic content" are not traditionally used to describe unprotected speech. *See Ashcroft*, 535 U.S. at 245–46; *Facebook Community Standards*, *supra* note 25. For an example of constitutionally protected speech that has been taken down or punished, see *supra* notes 2–4 and accompanying text regarding Courtney Stodden's nonobscene self-portraits.

For both the *Ashcroft* and Facebook categories, the line between acceptable and unacceptable content appears to be clear. However, the case law suggests that in reality the lines are difficult to determine. *See, e.g.*, Miller v. California, 413 U.S. 15, 20–25 (1973) (outlining the Supreme Court's difficulties in defining "obscenity" and declaring a new multi-part test). At least part of Facebook's inconsistency in applying its own terms of service (discussed *infra* Part I) stems from this problem.

38. *See* U.S. CONST. amend. I ("*Congress* shall make no law . . . ") (emphasis added); Gitlow v. New York, 268 U.S. 652, 666 (1925) (incorporating the First Amendment against the states via the Fourteenth Amendment).

39. Lugar v. Edmondson Oil Co., 457 U.S. 922, 936–37 (1982).

40. *See* discussion *infra* Part II.A.

41. The public forum doctrine can attach government control to certain types of locations, which then implicates government responsibilities and constitutional protections. *See, e.g.*, Perry Educ. Ass'n v. Perry Local Educators' Ass'n, 460 U.S. 37, 45–46 (1983) (discussing the distinctions between traditional public fora, designated public fora, and nonpublic fora).

do not have the traits of designated public fora.[42] Even if the First Amendment were to apply under either of these doctrines, there would be an issue of competing speech rights between the individual user and the social networking site.[43] The social networking site would claim that its right to expression should give it control over the entire platform and should allow it to determine what content is permissible and what content should be taken down.[44] The competing claims to free expression and the lack of direct First Amendment protections thus raise an important question: How can the free speech interests of users on social networking sites be protected?

This Note contributes to the scholarship on Internet regulation[45] by engaging in two discussions. First, I argue that the First

42. *See id.* at 39, 45–46 (distinguishing between traditional public fora—such as streets and parks, designated public fora—such as university meeting facilities, and nonpublic fora—such as public school mail facilities); *see also* discussion *infra* Part II.A.

43. *See, e.g.*, Jack M. Balkin, *Virtual Liberty: Freedom to Design and Freedom to Play in Virtual Worlds*, 90 VA. L. REV. 2043, 2043, 2074 (2004) [hereinafter Balkin, *Virtual Liberty*] (noting the conflicts between rights of users and online game providers).

The issue of competing rights is not unique to this scenario and arises in other First Amendment contexts, such as that of art, where an artist's interests can come into conflict with those of other private entities (galleries or curators) while not being able to invoke the First Amendment. S*ee* Richard Woodward, *Color Bind*, VILLAGE VOICE, June 25, 1996, at 78 (discussing a scenario where David Levinthal's Polaroids of "blackface memorabilia" show was censored by the Institute of Contemporary Art in Philadelphia for being "controversial").

44. Alternatively social networking sites could invoke their property rights to the platform to argue that they have the right to determine who may enter or use their property. However, property rights can give way to free speech issues in some cases. *See, e.g.*, Marsh v. Alabama, 326 U.S. 501, 503, 508–09 (1946) (a private company town invoked its property right to exclude a Jehovah's Witness, but the Supreme Court held that the individual's free speech interest outweighed this property right).

The more problematic conflict, in my view, would be between the competing First Amendment rights of the sites and the users. Thus I will focus on that tension.

45. The overarching question of how the Internet should be regulated has been the subject of much academic debate. One issue surrounds the "network neutrality" debate for Internet Service Providers. "Network Neutrality" refers to the concept that Internet service providers should not be able to censor any content that passes through their servers. *See* Christopher S. Yoo, *Network Neutrality and the Economics of Congestion*, 94 GEO. L.J. 1847, 1847, 1855–60 (2006) (providing a brief history of the network neutrality debate). For further information on the network neutrality debate, see generally Rob Frieden, *Assessing the Merits of Network Neutrality Obligations*, 115 PENN ST. L. REV. 49, 53 (2010) (conceptually dividing the Internet into three layers and analyzing the applicability of network neutrality for each level); Frank Pasquale, *Beyond Innovation and Competition: The Need for Qualified*

Amendment does not provide users of social networking sites with free speech protections because neither the state action doctrine nor the public forum doctrine clearly applies to social networking sites. Then, because of the importance of protecting user rights of expression, I recommend an alternative analytic model—a consumer protection framework[46]—to achieve the same goal of protecting user free-expression rights without adding any new distortions to First Amendment doctrine. This framework allows us to protect free speech values by focusing on the particular market failures that exist for social networking sites, which may then justify governmental intervention to protect user rights of expression independently of constitutional rights. At the same time, social networking sites may be able to invoke First Amendment protections to challenge governmental action to limit the sites' censorial power. The effect of this latter issue will depend on how social networking sites are characterized in terms of First Amendment media, as courts provide different levels of editorial rights to different

Transparency in Internet Intermediaries, 104 NW. U. L. REV. 105, 109 (2010) (suggesting regulations increasing transparency as the first step to addressing intermediation issues); Christopher S. Yoo, *Network Neutrality and the Economics of Congestion*, 94 GEO. L.J. 1847, 1904 (2006) (proposing a new framework for analyzing network neutrality restrictions and suggesting that a network diversity principle might be effective in obtaining policy objectives). A second point of much debate has been the FCC's recent regulations regarding whether Internet access provisions fall into the same category as broadband and cable. *See generally* Samuel L. Feder & Luke C. Platzer, *FCC Open Internet Order: Is Net Neutrality Itself Problematic for Free Speech?*, COMM. LAW., June 2011, at 1 (2011) (outlining the development of the FCC order and raising First Amendment issues of the order related to government purpose and chilling effects); Hannibal Travis, *The FCC's New Theory of the First Amendment*, 51 SANTA CLARA L. REV. 417, 421 (2011) (discussing FCC regulatory history and theorizing four new articulations of the First Amendment based on this history). Additionally, scholars have also discussed search engine bias. *See generally* Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149, 1151 (2008) (arguing that search engines should be subject to some nondiscrimination regulation to prevent search engine manipulation); Jennifer A. Chandler, *A Right to Reach an Audience: An Approach to Intermediary Bias on the Internet*, 35 HOFSTRA L. REV. 1095, 1096–97 (2007) (positing that search engine and other Internet regulations should be based on the relationship between speakers and listeners).

In addition, scholarly literature on Internet regulation in the context of social networking sites has largely focused on third-party torts and liability for Internet intermediaries under the CDA. *See, e.g.*, Ciolli, *supra* note 29; Freiwald, *supra* note 29.

46. *See* Douglas A. Kysar, *The Expectations of Consumers*, 103 COLUM. L. REV. 1700, 1747–50 (2003) (consumer protection rationales focus on how consumers act in particular markets to protect consumer interests).

types of media.[47] Determining whether social networking sites are more similar to newspapers or telephones will affect what rights a social networking site could invoke against regulation and how the site's rights would be balanced against a user's interests.

The structure of this Note is as follows: Part I discusses the nature of social networking sites, including their importance in social interaction today and their role as censors. Part II outlines my suggested framework for analyzing social networking sites. This Part begins by explaining why social network users probably will not be able to invoke First Amendment protections against social networking sites. Then I articulate consumer protection principles as an alternate basis for regulating social network censorship. I describe general consumer protection rationales, followed by an explanation of how these rationales apply to the social networking site market. This Part concludes with a discussion characterizing social networking sites in the context of existing First Amendment media categories. I argue that social networking sites generally do not exercise the same type of editorial discretion over content as newspapers, and should be characterized more as general pipelines of communication, resulting in a smaller amount of editorial discretion, if any. In Part III, I suggest and evaluate potential types of regulations or legal protections that could promote user rights on social networking sites within the backdrop of consumer protection rationales. This Part should not be seen as advocating for any of these specific "solutions," but rather as a call for more transparency from social networking sites, and an argument as to why a consumer protection framework that incorporates First Amendment principles should be used in discussing free speech issues related to the Internet.

I.
FACEBOOK, I CHOOSE YOU: A DESCRIPTION OF SOCIAL NETWORKING SITES

This Part provides the background information necessary to engage in the First Amendment and consumer protection analysis in Parts II and III. This Part begins by describing the growth of social networking sites and then demonstrates their increasing societal importance through examples of the myriad services they now provide. Then, this Part illustrates the role of social networking sites as censors by surveying a number of recent examples of censorship.

47. *See* discussion *infra* Part II.D.

A. Social Networking Sites Have Become Important Sites for Communication in Contemporary Society Due to Rapid Growth

Over the past few years, social networking sites have grown to become significant actors in society and interpersonal communication. Recent studies estimate that Facebook, the top social networking site based on user traffic, has over 750 million unique visitors per month.[48] As of October 2012, Facebook counted over 1 billion individuals as active monthly users globally, with approximately 190 million in the United States and Canada.[49] Twitter, the second most popular social networking site,[50] has an estimated 250 million unique monthly visitors and 500 million registered users globally.[51] These sites have grown at an accelerated rate. In 2009, Facebook became the top social networking site based on user traffic, just five years after its creation.[52] Between November 2009 and November 2011, the number of hits by unique visitors increased by nearly 50%

48. Albert Rox, *Top Social Networking Sites*, SOCIAL NETWORKS PLANET (Sept. 9, 2012), http://www.affilatenetworking.com/top-social-networking-sites/; *Top 10 Social Networking Sites*, DISCOVERY NEWS (Dec. 12, 2012), http://news.discovery.com/tech/top-ten-social-networking-sites.html; *Top 15 Most Popular Social Networking Sites*, EBIZMBA, http://www.ebizmba.com/articles/social-networking-websites (last updated Dec. 1, 2013); *Top Sites*, ALEXA, http://www.alexa.com/topsites (last visited Dec. 30, 2013) (listing Facebook as the number two website globally).

However, social networking statistics should be viewed with some amount of skepticism. *See* Matt Rhodes, *93% of the World is Not on Facebook*, SOCIAL MEDIA TODAY (July 21, 2010), http://socialmediatoday.com/mattrhodes/149640/93-world-not-facebook (arguing that Facebook's impressive growth statistics often overlook the fact that most of the world is not on Facebook, and that in some countries where Facebook is used, it is not the most popular way for people to connect online); Zoe Siskos, *Where Social Media Measurement Falls Short*, SOCIAL MEDIA TODAY (June 4, 2010), http://socialmediatoday.com/zoesiskos/110625/where-social-media-measurement-falls-short (noting that statistics regarding social media, while they can be useful, are often not used in the most beneficial manner and that they ignore the human aspects behind the number).

49. *See Key Facts*, FACEBOOK, http://newsroom.fb.com/content/default.aspx?NewsAreaId=22 (last visited Dec. 31, 2013) ("1.19 billion monthly active users as of September 30, 2013 . . . . Approximately 80% of our monthly active users are outside the U.S. and Canada.").

50. *See Top 15 Most Popular Social Networking Sites*, *supra* note 48.

51. *Id.*; Lisa O'Carroll, *Twitter Active Users Pass 200 Million*, THE GUARDIAN (Dec. 18, 2012, 12:51 PM), http://www.guardian.co.uk/technology/2012/dec/18/twitter-users-pass-200-million.

52. *See* Andy Kazeniac, *Social Networks: Facebook Takes Over Top Spot, Twitter Climbs*, COMPLETE PULSE (Feb. 9, 2009, 2:01 PM), http://blog.compete.com/2009/02/09/facebook-myspace-twitter-social-network (noting when Facebook overtook MySpace as the top social networking site, as measured by unique visitors and monthly visits).

on Facebook and nearly doubled on Twitter.[53] Users are also spending more time on social networking sites, with an average increase from 4.6 to 6.3 hours per month among U.S. users between 2010 and 2011.[54] Social networking site users also skew toward the young (under fifty), relatively affluent (annual income of $90,000+), and educated (college graduates).[55] In fact, from September 2005 to May 2010, the largest rates of growth have come from the eighteen to twenty-nine and the thirty to fourty-nine age groups, respectively increasing from 16% to 86% and from 12% to 61%.[56]

These statistics indicate that more and more people are using social networking sites as a means of communicating with others. Rather than calling, sending a letter, or even writing an e-mail, people are choosing to engage with others through a social networking site. Thus social networking sites operate as an infrastructure of communication in a similar way to the telephone and post office. They provide the same services as these other entities, just for the online environment. The increasing growth of account holders suggests that social networking sites may become the primary mode of communication for the general public in the future, especially because of the exponential growth rate of social networking site usage among younger users. Moreover, social networking sites are largely built on positive externalities—the addition of a user to a social network in turn spurs other members of that user's non-Internet-based

53. *See* Alyssa Maine, *Are 20-Somethings* Too *Connected? Or Not Connected Enough?*, COMPLETE PULSE (Dec. 19, 2011, 4:11 PM), http://blog.compete.com/2011/12/19/are-20-somethings-too-connected-or-not-connected-enough (publishing a graph showing unique visitors per month, with Facebook increasing from approximately 110 million to approximately 160 million and Twitter increasing from approximately 20 million to approximately 40 million).

54. *See* Dan Nelms, *Social Networking Growth Stats and Patterns*, SOCIAL MEDIA TODAY (June 16, 2011), http://socialmediatoday.com/amzini/306252/social-networking-growth-stats-and-patterns.

Part of this increased time spent on social networking sites may be attributed to the growth of smart phone use generally and, in particular, to the increase in mobile access to social networking sites, which grew by more than 200% from 2010 to 2011. *See* Kanalley, *supra* note 21.

55. *See* Lymari Morales, *Google and Facebook Users Skew Young, Affluent, and Educated*, GALLUP ECONOMY (Feb. 17, 2011), http://www.gallup.com/poll/146159/facebook-google-users-skew-young-affluent-educated.aspx.

56. *See* Kanalley, *supra* note 21. Another study estimates that 96% of 18–35 year olds are using a social networking site. *See Social Media Growth Statistics*, KISSMETRICS, http://blog.kissmetrics.com/social-media-statistics/ (last visited Jan. 11, 2013).

social network to join that specific online network;[57] this then increases the benefits for other users as more members of physical social networks become part of online social networking sites. Thus as entire friend groups or industries come to rely on particular sites, users may become locked into one platform.[58]

Social networking sites also provide services beyond a simple platform for communication. Today, businesses, employers, educational institutions, medical institutions, news organizations, and political figures all use social networking sites to expand their reach.

Social networking sites allow businesses to communicate with consumers in a faster and more efficient way.[59] A 2011 report revealed that 71% of companies used Facebook and 59% used Twitter.[60] Companies have learned that they can use social media to increase exposure for their company or brand, gain business partnerships,[61] and optimize their web presence.[62] On Facebook, for example, companies can set up an official company page to provide users with regular updates and create online events.[63] Companies often create offers that are only available on a specific social networking site.[64] Many companies also create software applications

57. *Cf. The Power of Social Networks*, GALLUP BUSINESS JOURNAL (Mar. 8, 2007), http://businessjournal.gallup.com/content/26770/power-social-networks.aspx (discussing the power of friends in social networking site dynamics).

58. The lock-in effect would occur due to the nature of social networking sites: users tend to join a social networking site based on access to their physical social network. *See id.*

59. *See* Bryant Ott, *Marketing to Tweeters and Their Facebook Friends,* GALLUP BUSINESS JOURNAL (Apr. 11, 2011), http://businessjournal.gallup.com/content/146990/marketing-tweeters-facebook-friends.aspx (discussing how social media provides a way for companies to respond to buyers' desire for speed).

60. Kanalley, *supra* note 21.

The growth of marketing events and conferences related to digital media suggests that the percentage of companies using social networking sites as part of business practices will continue to grow. *See, e.g.*, Jennifer Shore, *Facebook Marketing Strategy and 60+ More Events in Digital Media*, MASHABLE (Sept. 27, 2012), http://mashable.com/2012/09/27/events-9-27.

61. *See* Schawbel, *Major Findings*, *supra* note 20.

62. *See, e.g.*, Deborah M. Todd, *Small Business Owners Get Free Help Improving Their Web Presence*, PITTSBURGH POST-GAZETTE, Apr. 11, 2012, http://www.post-gazette.com/stories/business/news/small-business-owners-get-free-help-improving-their-web-presence-630830 (describing Google-sponsored events that provide advice for small business owners).

63. John Rampton, *5 Tips to get More Exposure for Business Using Facebook*, MAPLE NORTH (Feb. 28, 2012), http://www.maplenorth.com/2012/02/28/5-tips-to-get-more-exposure-for-business-using-facebook.

64. *See, e.g.*, Antioquia, *Just 2 More Sleeps! Facebook Exclusive Offer. . . Read. . .on*, FACEBOOK (Apr. 1, 2010, 7:08 PM), http://www.facebook.com/note.php?note_id=395086852368; Macy's, *supra* note 19; The Resort at the Mountain, *supra* note 19.

(known as "apps") that use social networking platforms in an attempt to reach more consumers.[65]

Employers also use social networking sites when considering applicants for purposes such as verifying credentials or identifying potentially problematic information about prospective hires.[66] Examples of issues that have arisen from social network background checks include: using Craigslist to search for OxyContin, putting up naked photos on an image-sharing site, and making anti-Semitic comments online.[67] A social networking background check has become so important that there are even companies, such as Social Intelligence, whose purpose is to provide this service.[68] Additionally, law schools are providing students with information on how to protect their online personas, precisely to prevent any social networking information from affecting a student's job search.[69]

Additionally, social networking sites play a role in medicine. The sites have been used to raise funds for medical projects, conduct digital diagnoses of patients in foreign countries, and expand awareness about diseases.[70]

---

65. *See* Jon Swartz, *Facebook Rolls out 60 Apps for Timeline*, USA TODAY, http://usatoday30.usatoday.com/tech/news/story/2012-01-18/facebook-lifestyle-apps/52653014/1 (last updated Jan. 18, 2012, 9:12 PM); *Facebook Adding More Than 60 New Apps for Its Timeline*, PEREZHILTON.COM (Jan. 22, 2012, 10:40 AM), http://perezhilton.com/2012-01-22-facebook-announces-that-they-are-adding-more-than-sixty-new-apps-for-their-timeline#.T1bCfHJSQU9 (providing examples of Facebook apps, including a fitness app from Nike and a movies app from Rotten Tomatoes). Information also flows in other directions, such as when social networking sites allow consumers to communicate with each other and to respond to companies. *See* Ott, *supra* note 59 (discussing how companies engage in discussions with "Millenials").

66. *See* Kanalley, *supra* note 21.

67. Jennifer Preston, *Social Media History Becomes a New Job Hurdle*, N.Y. TIMES, July 20, 2011, http://www.nytimes.com/2011/07/21/technology/social-media-history-becomes-a-new-job-hurdle.html?pagewanted=all.

68. *See id.* (stating that Social Intelligence only uses publicly available information when conducting a background check). The legality of these searches is no longer a question; in 2011, the FTC approved Social Intelligence's actions as being in compliance with the Fair Credit Reporting Act. *See* Vivian Luckiewicz, *Could You Survive a Social Media Background Check?*, OI PARTNERS (Oct. 2011), http://aka-oi.com/newsroom/newsletter/october-2011-newsletter/11-10-14/Could_You_Survive_a_Social_Media_Background_Check.aspx.

However, as mentioned earlier, some states are placing limits on the use of social networking information in employment contexts. *See, e.g.*, 820 ILL. COMP. STAT. 55/10 (2013).

69. *See, e.g.*, Office of Career Services, New York University School of Law, Protecting Your Online Persona (no date) (on file with author).

70. *See* Catherine A. Brownstein et al., *The Power of Social Networking in Medicine*, 27 NATURE BIOTECHNOLOGY 888, 888–89 (2009) (discussing development

For educational institutions, social networking sites provide services such as announcements, blogs about student life, online message or bulletin boards for classes, and professional development.[71] Social networking sites also allow education providers to communicate with each other to discuss teaching techniques and to share resources.[72] In addition, just as social networking may be changing societal views of communication, social networking may be changing the way that students learn, encouraging educators to experiment with and develop their teaching methods.[73]

---

of the "PatientsLikeMe," a medical social networking site); Lauren Hockenson, *How a Team of Doctors Uses Social Media to Drive Awareness and Save Lives*, MASHABLE (Feb. 27, 2012), http://mashable.com/2012/02/27/doctors-twitter-social-media.

In some cases, medical institutions or organizations develop their own purpose-specific social media or networking sites. *See, e.g.*, Brownstein et al., *supra*, at 888–89. These websites have allowed doctors to obtain data about patient populations and speed up patient recruitment for clinical research trials. *See id.* at 890. Some primary care facilities have also started experimenting with web-based social media such as weblogs, instant messaging platforms, and social networking sites to provide low-cost services. *See* Carleen Hawn, *Take Two Aspirin and Tweet Me in the Morning: How Twitter, Facebook, and Other Social Media Are Reshaping Health Care*, 28 HEALTH AFF. 361, 361, 363 (2009).

71. *See* Bryan Alexander, *Social Networking in Higher Education*, *in* THE TOWER AND THE CLOUD: HIGHER EDUCATION IN THE AGE OF CLOUD COMPUTING 197, 199 (Richard N. Katz ed., 2008), *available at* http://net.educause.edu/ir/library/pdf/PUB7202s.pdf (describing examples of social media uses in higher education, such as blogging or wiki projects); Matt Silverman, *How Higher Education Uses Social Media*, MASHABLE (Feb. 3, 2012), http://mashable.com/2012/02/03/higher-education-social-media (infographic surveying uses of social media in higher education, providing examples such as tweeting class announcements and creating a school Facebook page). As in the medical context, educational institutions use both pre-existing social networking sites and self-developed websites. *See* Alexander, *supra*, at 198–200 (discussing examples including educational uses of *Wikipedia* and course-specific blogs and podcasts).

72. *See List of Networks*, EDUCATIONAL NETWORKING, http://www.educationalnetworking.com/List+of+Networks (last visited Dec. 31, 2013) (providing lists of educational resource pages on Facebook and LinkedIn, and of separate education-specific pages and networks).

73. *See* Lorenzo Franceschi-Bicchierai, *Twitter Boosts College Grades and Class Engagement*, MASHABLE (Oct. 20, 2012), http://mashable.com/2012/10/20/twitter-students-writers; *cf. NMC Initiatives*, THE NEW MEDIA CONSORTIUM, http://www.nmc.org/about/initiatives (last visited Jan. 12, 2013) (describing the core initiatives of the New Media Consortium, including how technology contributes to dynamic knowledge production and initiating collaborations with other organizations to encourage digital literacy).

At least one study has found that use of Twitter in education led to students who "[were] more engaged and [had] higher grades." Franceschi-Bicchierai, *supra*.

Journalists and news organizations use social networking as a means of obtaining, verifying, and disseminating up-to-date information about ongoing events.[74] For example, newspapers post headline feeds in Twitter, create online events to attract readers, and customize delivery of information to a reader's interests.[75] *The Washington Post* and *The New York Times* have gone so far as creating social readers compatible with Facebook that encourage Facebook users to read articles.[76] Yet perhaps the "Trending Articles" feature on Facebook is the most telling example of how news and social networking sites are connecting, as it allows a user to see news articles based on what the user's friends are reading.[77]

Finally, social media and social networking sites play a role in politics. These websites played a significant role in the "Arab Spring" revolutions, particularly in Egypt and Tunisia.[78] Protestors utilized social networking sites like Facebook and Twitter to accelerate protests,[79] publicly discuss political issues,[80] and link activist

74. *See* Dan Schawbel, *How a Journalist Uses Social Media*, FORBES (Oct. 24, 2011, 10:04 AM), http://www.forbes.com/sites/danschawbel/2011/10/24/how-a-journalist-uses-social-media.

75. *See* Woody Lewis, *10 Ways Newspapers are Using Social Media to Save the Industry*, MASHABLE (Mar. 11, 2009), http://mashable.com/2009/03/11/newspaper-industry.

76. *See* Josh Catone, *New York Times Launches Facebook App*, READWRITEWEB (Sept. 12, 2007), http://www.readwriteweb.com/archives/new_york_times_launches_facebook_app.php (describing *The New York Times*' daily short news quiz Facebook app, which encourages quiz-takers to open news articles on a *New York Times* webpage, unlike *The Washington Post* app, which allows users to read articles without leaving Facebook); Jennifer Van Grove, *The Washington Post Launches Social Reader as a Newspaper for Facebook*, MASHABLE (Sept. 22, 2011), http://mashable.com/2011/09/22/social-reader.

77. *See* Lauren Indvik, *Facebook Tests Smaller Version of 'Trending Articles' in Newsfeed*, MASHABLE (Apr. 26, 2012), http://mashable.com/2012/04/26/facebook-trending-articles-test.

78. *See* Raymond Schillinger, *Social Media and the Arab Spring: What Have We Learned?*, HUFFINGTON POST (Sept. 20, 2011, 3:59 PM), http://www.huffingtonpost.com/raymond-schillinger/arab-spring-social-media_b_970165.html.

79. *Twitter, Facebook, and YouTube's Role in Arab Spring (Middle East Uprisings)*, SOCIAL CAPITAL BLOG, http://socialcapital.wordpress.com/2011/01/26/twitter-facebook-and-youtubes-role-in-tunisia-uprising/ (last updated Oct. 12, 2012).

80. *See* Kate Taylor, *Arab Spring Really Was Social Media Revolution*, TG DAILY (Sept. 13, 2011, 5:00 AM), http://www.tgdaily.com/software-features/58426-arab-spring-really-was-social-media-revolution; Robert F. Worth, *Twitter Gives Saudi Arabia a Revolution of Its Own*, N.Y. TIMES, Oct. 20, 2012, http://www.nytimes.com/2012/10/21/world/middleeast/twitter-gives-saudi-arabia-a-revolution-of-its-own.html?pagewanted=all&_r=0.

groups in different locations.[81] Social networking sites also played a role in the 2008 U.S. presidential election as a tool for candidates to reach greater portions of the U.S. population, and they featured prominently in the 2012 presidential election.[82]

All of these examples demonstrate the extent to which social networking sites pervade areas of social activity beyond communication or self-expression.[83] The expansive growth of social networking sites into areas from business to medicine to politics suggest that a user's ability to access social networking sites is becoming more important and that either having content taken down or an account deleted would have serious negative ramifications for a user's social experience. Further, I argue that this consolidation of the social experience gives rise to user rights, which while not protectable under the U.S. Constitution are, as will be further explicated, protectable under a consumer protection framework.

81. *See* Natana J. DeLong-Bas, *The New Social Media and the Arab Spring*, OXFORD ISLAMIC STUDIES ONLINE, http://www.oxfordislamicstudies.com/Public/focus/essay0611_social_media.html (last visited Dec. 31, 2013).

82. *See* Alex Fitzpatrick, *Second Presidential Debate: Less Twitter, More Facebook*, MASHABLE (Oct. 17, 2012), http://mashable.com/2012/10/17/hofstra-debate-twitter-facebook; Electron Libre, *The Role of Social Networks in the 2012 Presidential Election*, FRANCE 24 http://www.france24.com/en/20121107-2012-11-07-1319-wb-en-webnews (last updated Nov. 8, 2012); Weber, *supra* note 22.

The 2010 U.S. Midterm Elections demonstrated the use of Twitter to remind people to vote, monitor the voting process, and communicate about electoral issues. *See* Dugan, *supra* note 22.

83. This Note does not rely on any assumption that social networking sites are a net positive phenomenon. Instead, the purpose of this Section is to demonstrate that social networking sites have become a pervasive element of social interaction. Though it is beyond the scope of this Note, there is much debate over the effects of social networking sites on communication and social interaction. *Compare* James Gurd, *Does Social Media Kill Communication Skills?*, ECONSULTANCY (Sept. 18, 2009, 10:52 AM), http://econsultancy.com/us/blog/4636-does-social-media-kill-communication-and-people-skills (arguing that social media enhances communication), *with* Melissa Bell & Elizabeth Flock, *'A Gay Girl in Damascus' Comes Clean*, WASH. POST, June 12, 2011, http://www.washingtonpost.com/lifestyle/style/a-gay-girl-in-damascus-comes-clean/2011/06/12/AGkyH0RH_story.html (reporting the worries of gay bloggers that their ability to use pseudonyms could be jeopardized after the revelation that a beloved Syrian lesbian blogger's identity was completely fabricated—the true blogger was an American man), *and* Megan Puglisi, *Social Networking Hurts the Communication Skills of College Students*, THE DAILY ATHENAEUM (Oct. 13, 2010), http://www.thedaonline.com/opinion/social-networking-hurts-the-communication-skills-of-college-students-1.1689315 (arguing that social networking sites are ruining the communication skills of college students), *and* Nick Stamoulis, *Is Social Networking Slowing Down the Generational Lines of Communication?*, MARKETING PILGRIM (Mar. 20, 2009), http://www.marketingpilgrim.com/2009/03/social-networking-generations.html (arguing that rapidly evolving social networking sites are exacerbating generational divides).

B. Social Networking Sites Act as Censors and Filter Both Content and Users

Although social networking sites provide many important services and serve expressive purposes, they engage in censorship when they take down content or delete user accounts.[84] As mentioned in the Introduction, it is relatively settled law that social networking sites have some ability to censor due to the fact that they now qualify as "interactive computer service" providers and thus obtain the immunity protections of § 230 of the Communications Decency Act of 1996.[85] Although the immunity provision was originally intended to provide a small exception to Internet service provider liability, courts have expanded the immunity to apply to a greater variety of actors, including employers or people who repost the material of others.[86]

One style of censorship is account deletion. The opening illustration of the deletion of Courtney Stodden's fan page provides an example of when Facebook took such action against a user for posting allegedly inappropriate self-portraits.[87] A seemingly less problematic example of account deletion is that of Aaditya Thackeray, whose account was taken down because Facebook believed that he was not using his real name and was therefore violating Facebook's terms of use.[88] Mr. Thackeray was able to prove his identity, and his account was reactivated after an internal appeal to Facebook.[89]

However, full account deletion is not the only means that social networking sites use to censor user content. Another perhaps more common form of censorship is when a social networking site takes down particular content rather than delete a user's account.

84. The right to take down material is usually reserved in the terms of service that users agree to when creating an account. *See, e.g.*, *Statement of Rights and Responsibilities*, *supra* note 25.

85. *See, e.g.*, Zeran v. America Online, Inc., 129 F.3d 327, 330, 332 (4th Cir. 1997) (interpreting provisions of the CDA and finding that the statute applied to early multi-purpose sites like AOL); Doe v. MySpace, Inc., 474 F. Supp. 2d 843, 849–50 (W.D. Tex. 2007) (finding MySpace to be entitled to immunity under the CDA).

86. Mark A. Lemley, *Rationalizing Internet Safe Harbors*, 6 J. ON TELECOMM. & HIGH TECH. L. 101, 102–03 (2007). However, the actual scope of the immunity provision remains unclear. *Id.* at 106–07.

87. *See* Hughes, *supra* note 2.

88. *See Facebook Asks*, *supra* note 36; *Statement of Rights and Responsibilities*, *supra* note 25.

89. *See Facebook Asks*, *supra* note 36.

For example, in 2011 Facebook took down a set of pro-rape pages.[90] Although Facebook initially declined to remove the content, stating that "[i]t is very important to point out that what one person finds offensive another person can find entertaining, just as telling a rude joke won't get you thrown out of your local pub, it won't get you thrown off Facebook," an online petition garnering 186,000 signatures and a Twitter campaign pressured the site into removing some, but not all, of the sexual assault pages.[91] Facebook has also removed user content that is arguably less offensive, such as pictures of breastfeeding mothers.[92] The breastfeeding picture issue also created a large online movement in support of allowing these pictures, however this movement did not cause an immediate change in Facebook's community standards.[93] It is only recently that Facebook has changed its position regarding these images.[94]

Part of the content take-down guidelines that Facebook provides to external censoring companies was leaked to the public, allowing the online community a glimpse into what content Facebook finds objectionable and what content the website permits.[95] The guidelines demonstrate many oddities. For example, Facebook does not allow any form of sexual activity, even simulated, whereas depictions of deep wounds and excessive blood are allowed, "as long as no insides are showing."[96] The guidelines provided to the external company that Facebook employs to review flagged content are much more specific and in-depth than the standards available to users on Facebook's help section.[97] Yet even these

90. *See* Adrian Chen, *Facebook Removes Pro-Rape Pages, Kicking and Screaming*, GAWKER (Nov. 9, 2011, 4:21 PM), http://gawker.com/5858000/facebook-removes-pro+rape-pages-kicking-and-screaming.

91. *See id.*

92. *See* Jenna Wortham, *Facebook Won't Budge on Breastfeeding Photos*, N.Y. TIMES (Jan. 2, 2009, 6:34 PM), http://bits.blogs.nytimes.com/2009/01/02/breastfeeding-facebook-photos.

93. *See id.*

94. *Facebook Community Standards*, *supra* note 25 ("We aspire to respect people's right to share content of personal importance, whether those are photos of a sculpture like Michelangelo's David or family photos of a child breastfeeding.").

95. *See No Sex, but Crushed Heads Are OK. Leaked Facebook Document Reveals Website's Secretive and Bizarre "Graphic Content" Policy*, DAILY MAIL ONLINE (Feb. 21, 2012, 7:56 PM), [hereinafter *No Sex, But Crushed Heads are OK.*], http://www.dailymail.co.uk/sciencetech/article-2104424/Facebooks-bizarre-secretive-graphic-content-policy-revealed-leaked-document.html

96. *See id.*

97. *Compare No Sex, but Crushed Heads are OK.*, *supra* note 95 (revealing Facebook's guidelines to the external company that prohibit "[c]ontent showing Poster's delight in/involvement in/promoting of/encouraging of violence against humans or animals for sadistic purposes (e.g. torture, staged animal fights, animal

"specific guidelines" are not enforced directly or consistently, which complicates a user's ability to predict what content will be considered problematic and what consequences may attach.[98]

Facebook is not alone in engaging in censorship. In addition to taking down material that infringes on copyrights or depicts child pornography, Twitter has a policy of "block[ing] certain messages in countries where they [are] deemed illegal."[99] This Twitter policy provides further complications for the site's users. Rather than setting a uniform standard, under this policy, Twitter may withhold content from users in a country where the government sends Twitter a request stating that a Twitter account is illegal.[100] In response to this censorship policy, many Twitter users

starvation, obvious neglect, etc.)" and content "[d]epicting the mutilation of people or animals, or decapitated, dismembered, charred, or burning humans"), *with Facebook Community Standards*, *supra* note 25 (stating that "[s]haring any graphic content for sadistic pleasure is prohibited").

98. *See, e.g.*, Katie Weiss, *Online Censorship: No Sex Please, We're Facebook*, CRIKEY (Aug. 2, 2011, 1:03 PM), http://www.crikey.com.au/2011/08/02/online-censorship-no-s-x-please-were-facebook (noting that a literal interpretation of the Facebook guidelines resulted in a takedown of Nirvana's *Nevermind* album cover, which depicted a naked baby, but that the album art was later restored as not actually being a violation of Facebook's policies).

99. *See* Somini Sengupta, *Censoring of Tweets Sets Off #Outrage*, N.Y. TIMES, Jan. 27, 2012, http://www.nytimes.com/2012/01/28/technology/when-twitter-blocks-tweets-its-outrage.html?_r=1.

The question of which country's or community's laws should be used to govern Twitter or Facebook is a difficult issue, especially because of the trans-border nature of the Internet. Such issues are beyond the scope of this Note, and I will assume that U.S. laws will apply to social media use within the territory of the United States.

100. *See* Nicholas Kulish, *Twitter Blocks Germans' Access to Neo-Nazi Group*, N.Y. TIMES, Oct. 18, 2012, http://www.nytimes.com/2012/10/19/world/europe/twitter-blocks-access-to-neo-nazi-group-in-germany.html?pagewanted=all; Hayley Tsukayama, *Twitter's Country-Specific Censorship Tool Prompts User Protest*, WASH.POST, Jan. 27, 2012, [hereinafter Tsukayama, *Twitter's Country-Specific Censorship Tool*], http://www.washingtonpost.com/business/technology/twitters-country-specific-censorship-tool-prompts-user-protest/2012/01/27/gIQALWoMVQ_story.html?wpisrc=nl_headlines. For example, Twitter recently banned a Neo-Nazi account in Germany because Nazi symbols and slogans are banned under German law. *See* Kulish, *supra*. However, Twitter did not block the account in other countries. *Id.*

started a protest.[101] This protest has not yet led to a change in Twitter's policy.[102]

These policies raise the question of what level of censorship authority social networking sites should have over user content. To resolve this question, it is necessary to discuss what legal doctrines and rights would apply to online social networking sites.

II.
PREPARE FOR TROUBLE AND MAKE IT DOUBLE: ANALYZING TWO FRAMEWORKS FOR PROTECTING USER RIGHTS OF EXPRESSION ON SOCIAL NETWORKING SITES

This Part determines what framework of analysis will best protect user rights of expression in order to evaluate possible governmental interventions for protecting user rights of free expression discussed in Part III. First, I explain why users will likely be unsuccessful in invoking First Amendment protections against social networking sites. The biggest obstacles for individuals in this framework are the state action and public forum doctrines. I therefore suggest a consumer protection framework as an alternative.[103] I provide an overview of consumer protection rationales and turn to how traditional market failures exist in the social networking site context. This conclusion invites government intervention as a solution to these market failures. However, application of the consumer protection framework may run into problems in practice. Specifically, social networking sites could invoke their First Amendment rights against attempts to limit their censorial authority. Yet the

101. Hayley Tsukayama, *Twitter Faces Accusations of Censorship; Users Plan Saturday Boycott*, WASH. POST, Jan. 27, 2012, http://www.washingtonpost.com/business/technology/twitter-faces-accusations-of-censorship-users-plan-saturday-boycott/2012/01/27/gIQAJ8dHWQ_story.html?wpisrc=nl_headlines; Tsukayama, *Twitter's Country-Specific Censorship Tool*, *supra* note 100.

102. *Country Withheld Content*, TWITTER, http://support.twitter.com/groups/33-report-abuse-or-policy-violations/topics/148-policy-information/articles/20169222-country-withheld-content# (last visited Dec. 31, 2013).

103. My discussion will focus on existing doctrines used to apply the First Amendment against private entities. For alternate views on how the First Amendment could be applied to extend free speech protections to prevent censorship by private entities, see Chandler, *supra* note 46, at 1097–98 (suggesting that the right to reach an audience could provide for First Amendment protections to be applied against Internet intermediaries); Hannibal Travis, *Of Blogs, eBooks, and Broadband: Access to Digital Media as a First Amendment Right*, 35 HOFSTRA L. REV. 1519, 1526 (2007) (arguing that an originalist reading of the First Amendment would allow courts to apply First Amendment protections against private entities that hold government-sponsored monopolies).

amount of editorial or censorial power given to a particular form of media varies depending on the type of media. The final Section of this Part will discuss how social networking sites should be characterized to determine what amount of editorial power and First Amendment protection they warrant.

A. The First Amendment Does Not Provide Users with Direct Protection Against Censorship by Social Networking Sites

The First Amendment states: "Congress shall make no law . . . abridging the freedom of speech, or of the press . . . ."[104] In effect, the First Amendment prevents the government from taking any action that would restrict an individual's speech or expression.[105] However, the First Amendment generally does not proscribe actions by a private non-state actor that affect another private non-state actor's expressive ability.[106] There are two main exceptions to this general rule. The First Amendment can apply to private actors, including social networking sites, if the state action requirement is met[107] or if the forum is characterized as a public forum.[108] I address each doctrine in turn.

---

104. U.S. CONST. amend. I.

I will restrict my analysis to the free speech protections of the federal Constitution. State constitutions may provide broader free speech protections and may not contain a state action requirement. *See, e.g.*, Pruneyard Shopping Ctr. v. Robins, 447 U.S. 74, 80–81 (1980); Melvin v. Reid, 297 P. 91, 91, 93–94 (Cal. Dist. Ct. App. 1931) (applying California state constitution provisions against nongovernmental actors).

105. *See* R.A.V. v. City of St. Paul, 505 U.S. 377, 382 (1992) ("The First Amendment generally prevents government from proscribing speech, or even expressive conduct, because of disapproval of the ideas expresed.") (citations omitted).

106. *See* Helen Hershkoff, *Horizontality and the "Spooky" Doctrines of American Law*, 59 BUFF. L. REV. 455, 455 (2011) [hereinafter Hershkoff, *Horizontality*] ("That the federal Constitution does not apply to private relations ranks among the most entrenched principles of American law."); Michael J. Perry, *Freedom of Expression: An Essay on Theory and Doctrine*, 78 NW. U. L. REV. 1137, 1138 (1984).

107. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 265 (1964) (affirming the state action requirement when invoking the First Amendment).

108. *See, e.g.*, Amalgamated Food Emp. Union Local 590 v. Logan Valley Plaza, Inc., 391 U.S. 308, 315–19 (1968) (finding private shopping center to be public for the purposes of the First Amendment and applying early "public forum" doctrine cases, although the specific term "public forum" is not used). *But see* Lloyd Corp. v. Tanner, 407 U.S. 551, 570 (1972), *extended by* Hudgens v. NLRB, 424 U.S. 507, 514–21 (1976) (recognizing that *Lloyd* overturned *Logan Valley*). The issue of whether the public forum doctrine actually can be applied to private actors is discussed later in this Part. For further information on the public forum doctrine, see Nancy J. Whitmore, *First Amendment Showdown: Intellectual Diversity Mandates and the Academic Marketplace*, 13 COMM. L. & POL'Y 321, 339 (2008)

The essential inquiry for state action is whether "the conduct allegedly causing the deprivation of a federal right [is] fairly attributable to the State."[109] The case law governing the issue of what conduct meets the state action doctrine is unclear, and no court has formulated any generally accepted metric.[110] In the most obvious cases, government officials acting within the scope of their authority meet the state action requirement.[111] Yet even indirect government involvement can trigger state action.[112] In perhaps one of the least obvious state action cases, the Supreme Court found that the mere application of state law in a libel lawsuit between an individual and The New York Times met the state action requirement and allowed the newspaper to bring a First Amendment defense.[113] How-

[hereinafter Whitmore, *First Amendment Showdown*] (discussing the public forum doctrine and how it restricts government control over speech).

109. Lugar v. Edmondson Oil Co., 457 U.S. 922, 937 (1982).

Some scholars argue that the doctrine as a whole should be scrapped. However, courts are likely to continue applying the state action doctrine in order to determine when constitutional protections can be applied against nonstate actors. For criticisms of the state action doctrine, see generally Dilan A. Esper, Note, *Some Thoughts on the Puzzle of State Action*, 68 S. CAL. L. REV. 663, 670–77 (1995) (describing various criticisms of the state action doctrine, such as the rise of the corporation and the arguable incoherence of the public/private distinction).

110. *See* Erwin Chemerinsky, *Rethinking State Action*, 80 NW. U. L. REV. 503, 503–05 (1985) (quoting other scholars describing state action doctrine as a "conceptual disaster" and citing sources that point to the incoherence of the state action doctrine).

111. *See, e.g.*, Harris v. City of Roseburg, 664 F.2d 1121, 1127 (9th Cir. 1981) (indicating that police intervention in a repossession is state action).

112. *See, e.g.*, Soldal v. Cook Cnty., 506 U.S. 56, 72 (1992) (finding state action when alleged debtor's mobile home was seized with support of deputy sheriffs). Government-created entities can also meet the state action requirement if the entity is established in furtherance of governmental objectives and the government retains some degree of control over the entity. *See* Lebron v. Nat'l R.R. Passenger Corp., 513 U.S. 374, 399 (1995).

In at least one extreme case, a court held that government inaction—in that case, a county policy that prevented unauthorized citizens from saving a drowning child—violated constitutional protections. *See* Ross v. United States, 910 F.2d 1422, 1425, 1430 (7th Cir. 1990) (holding that the policy violated the Fourteenth Amendment).

113. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 262–65 (1964) (finding that the First Amendment applied because of state enforcement and judicial proceedings). In another case involving rights of publicity, the Eighth Circuit also found that court enforcement of state-created obligations met the state action requirement. *See* C.B.C. Distribution & Mktg., Inc. v. Major League Baseball Advanced Media, L.P., 505 F.3d 818, 823 (8th Cir. 2007).

ever, courts have not been consistent in finding state action in every case where similar facts are alleged.[114]

More problematic cases arise when the action involves only private entities. In fact, the state action doctrine was created precisely to protect private individuals from having their rights and liberty subject to governmental obligations when engaging with other private parties.[115] Despite that concern, in some cases private actions are attributed to the state.[116]

---

114. *See, e.g.*, Flagg Bros. v. Brooks, 436 U.S. 149, 153–55, 163 (1978) (holding that the proposed sale of respondent's property by a creditor was not state action, even though the sale was authorized by a state statute); Harley v. Oliver, 539 F.2d 1143, 1144–46 (8th Cir. 1976) (denying recovery to a mother who temporarily lost custody of her child pursuant to a court order because the father's actions in seeking court action did not constitute a "scintilla of state action").

115. *See* Lugar v. Edmondson Oil Co., 457 U.S. 922, 936–37 (1982) (stating that the state action limitation "preserves an area of individual freedom by limiting the reach of federal law"); William M. Burke & David J. Reber, *State Action, Congressional Power, and Creditors' Rights: An Essay on the Fourteenth Amendment*, 46 S. CAL. L. REV. 1003, 1012–14, 1016–17 (1973) (arguing that the state action doctrine was intended to protect individual autonomy); Chemerinsky, *supra* note 110, at 506.

For an alternate view of state action doctrine, see Christopher D. Stone, *Corporate Vices and Corporate Virtues: Do Public/Private Distinctions Matter?*, 130 U. PA. L. REV. 1441, 1492–1506 (1982) (arguing that the theory underlying state action in cases involving "hybrid actors" should be a "moral exemplar" theory—that public organizations must meet a higher standard because they set an example for others).

116. *See, e.g.*, Burton v. Wilmington Parking Auth., 365 U.S. 715, 716–17, 723–26 (1961) (finding state action when a restaurant leasing governmental property engaged in discrimination).

My view of state action focuses on two categories of state action, public function and entanglement, which are supported by multiple scholars. *See, e.g.*, Hershkoff, *Horizontality*, *supra* note 106, at 499–500 (discussing the public function and "entanglement" exceptions); *cf.* Peter M. Shane, *The Rust that Corrodes: State Action, Free Speech, and Responsibility*, 52 LA. L. REV. 1585, 1586–87, 1589–90 (1992) (noting that cases tend to fall into one of three versions of state action, and using the example of *Edmonson v. Leesville Concrete Co.*, 500 U.S. 614 (1991), to illustrate the Supreme Court's focus on concepts such as "significant participation of the government" and "traditional function of the government" to support finding state action (internal quotation marks omitted)). My reason for focusing on these two categories is that they appear to encompass the majority of state action cases. The more nuanced distinctions in other discussions subdivide these major categories into more specific groupings. For alternate and additional conceptions of when private action becomes state action, see, e.g., G. Sidney Buchanan, *A Conceptual History of the State Action Doctrine: The Search for Governmental Responsibility,* 34 HOUS. L. REV. 333, 344–354 (1997) (dividing state action into six "issues"); Esper, *supra* note 109, at 709–13 (describing three situations where state action is found: (1) when the right at issue is "so important to the functioning of the government that the state has a . . . duty to protect it;" (2) when "the state has delegated significant authority to a private entity;" or (3) when private actors become too powerful).

One type of private-entity state action occurs when a governmental entity delegates authority over a public function to a private actor. While the point at which governmental delegation creates state action is not clear, one general rule of thumb has been that where the function provided by the private entity serves the general public, the entity's action is considered a public function,[117] and therefore state action.[118] Courts have used this doctrine to hold that the delegation of public functions was sufficient to confer state action in contexts such as company towns,[119] but not for shopping malls.[120]

However, both of these sources—and many others—include both categories in their discussions of state action that I focus on here.

Finally, for a specific discussion of private networks revolving around the congressionally created National Research and Education Network ("NREN"), see Michael I. Meyerson, *Virtual Constitutions: The Creation of Rules for Governing Private Networks*, 8 HARV. J.L. & TECH. 129, 134–39 (1994) (discussing the NREN government-sponsored network in the context of a variety of state action cases). NREN is not a social networking site, but Meyerson's article may provide some useful insight on state action and Internet infrastructure. However, as NREN is a government-sponsored program, it does not present the same scenario as a purely private entity.

117. The actual definition of "public function" is not clear from the case law. Some cases have defined "public function" as a function traditionally *exclusively* performed by the state. *See, e.g.*, S.F. Arts & Athletics, Inc. v. U.S. Olympic Comm., 483 U.S. 522, 544 (1987). However, other cases have found a public function for actions such as a pre-primary election, which are not exclusive or traditional functions performed by the state. *See, e.g.*, Terry v. Adams, 345 U.S. 461, 462–63, 468–69 (1953). For a broader discussion of the problems defining public function in the case law, see Esper, *supra* note 109, at 690 n.132, 692–708.

Additionally, the public function issue within the state action doctrine closely relates to the question of whether a location is a public forum, which will be discussed further below.

118. *See* Buchanan, *supra* note 116, at 346 (discussing how serving a public function establishes state action).

119. *See* Marsh v. Alabama, 326 U.S. 501, 502–03, 508–10 (1946). In *Marsh*, the Supreme Court found that a private company town was subject to First Amendment limitations without identifying the town as a state actor. *Id.* at 502–03. Further cases define the scope of this decision and explain that the case stands as a "paradigmatic" example of the state action question despite no language in the case itself referring to the state action doctrine. *See Marsh*, 326 U.S. 501; Steven Siegel, *The Constitution and Private Government: Toward the Recognition of Constitutional Rights in Private Residential Communities Fifty Years after* Marsh v. Alabama, 6 WM. & MARY BILL RTS. J. 461, 472–74 (1998). The Supreme Court also noted that the private town performed a public function. *Marsh*, 326 U.S. at 506.

120. *See* Lloyd Corp. v. Tanner, 407 U.S. 551, 570 (1972), *extended by* Hudgens v. NLRB, 424 U.S. 507, 514–21 (1976) (recognizing that *Lloyd* overturned *Amalgamated Food Employees Union Local 590 v. Logan Valley Plaza, Inc.*, 391 U.S. 308 (1968)—by extension, limiting the scope of *Marsh*).

Alternatively, courts have sometimes applied an entanglement or nexus test.[121] For example, in *Burton v. Wilmington Parking Authority*,[122] the Supreme Court found a lease of private property from the government could require enforcement of constitutional rights against the private leaseholder because of the relationship between the state and the private entity.[123] Once again, the amount or quality of contacts sufficient to establish state action is unclear. However, cases suggest that some type of contract or legally recognized relationship may be sufficient for a nexus to exist.[124]

Although the government does not directly act when a private website takes down content or deletes a user's account, individuals who desire to invoke the First Amendment could make various arguments as to how censorship by a social networking site meets the state action doctrine. In the public function version of state action, a user could argue that social networking sites provide a communication infrastructure—a public function like that provided by the postal service. Thus the government has delegated control of this public communication function to private entities. This argument may be augmented by the fact that social networking websites, like Facebook or Twitter, now allow anyone to join their networks and publish content on the Internet for consumption by the general public. For the nexus test, a user could argue that social networking sites are using the government-granted immunity provisions[125] to engage in censorship. Therefore these sites function as proxies for government action. Finally, a user could argue that the immunity legislation can be conceived of as a contract or agreement between the government and interactive service providers regarding censorship policies.

Given the uncertainty in the state action doctrine, it would be difficult to say that any of these arguments would convince a court

121. *See generally* Samuel Estreicher, Note, *Federal Power to Regulate Private Discrimination: The Revival of the Enforcement Clauses of the Reconstruction Era Amendments*, 74 COLUM. L. REV. 449, 451–61 (1974) (examining the development of the entanglement test in relation to private discrimination actions).

122. 365 U.S. 715 (1961).

123. *Id.* at 716–17, 723–26.

124. *See* Evans v. Newton, 382 U.S. 296, 299–302 (1966) (city operation of a park "for whites only" was tainted by state action both when the state acted as trustee and when private trustees were substituted); *Burton*, 365 U.S. at 723–24 (discussing how leased areas were "a physically and financially integral and, indeed, indispensable part of the State's plan"). However, the *Evans* Court also believed that the "public function" test was met. *See Evans*, 382 U.S. at 302.

125. *See* Communications Decency Act of 1996, 47 U.S.C. § 230(c)(2) (protecting from liability providers or users of interactive computer services who, in good faith, restrict access to materials that are considered offensive).

to apply the First Amendment, especially since there are compelling arguments on the opposite side.[126] Although social networking sites provide a platform of communication to the general public, these websites can choose to restrict access to particular groups of users, as Facebook did when it started,[127] and change membership criteria at will. Additionally, users must create an account and agree to terms of service that give the website control over content, among other things.[128] The private contractual relationships between users and social networking sites undermine the argument that Facebook provides a public function akin to the post office. In addition, whether posts are open to the entire Internet public or not results from the individual user's choice, partially mitigating the argument that Facebook connects users to the public sphere.[129] Finally, the government did not explicitly delegate control of these platforms to the social networking sites, and social networking sites are private entities. If an implied delegation of a public function to a private entity were accepted, many other scenarios where the gov-

126. *Cf.* Balkin, *Virtual Liberty*, *supra* note 44, at 2074–75 (discussing conflicts between gamers and game designers for virtual games and suggesting that the First Amendment might not apply). *See generally* Neil Weinstock Netanel, *Cyberspace Self-Governance: A Skeptical View from Liberal Democratic Theory*, 88 CALIF. L. REV. 395, 459 (2000) (suggesting that the private nature of regulation on the Internet calls into question the availability of First Amendment protections).

Additionally, some scholars note that the most expansive applications of state action to private entities have come in the context of racial discrimination, which is not an issue in the case of online social networking sites. *See* Terri Peretti, *Constructing the State Action Doctrine, 1940–1990*, 35 LAW & SOC. INQUIRY 273, 287 (2010) (suggesting that, in some cases, the presence of a racial discrimination claim was an important element in finding state action); David A. Strauss, *State Action After the Civil Rights Era*, 10 CONST. COMMENT. 409, 409 (1993) (emphasizing the role of state action in racial discrimination cases). For examples of outlier cases where racial discrimination may have led to a finding of state action, see *Burton*, 365 U.S. at 716–17, 723–26 (finding state action when restaurant leasing governmental property engaged in discrimination) and *Shelley v. Kraemer*, 334 U.S. 1, 19 (1948) (finding state action in state enforcement of a racially restrictive covenant). If true, this would be another reason why courts would be unwilling to find state action in this context.

127. *See* Adam Pash, *Facebook Opens Registration to All*, LIFEHACKER (Sept. 26, 2006, 3:00 PM), http://lifehacker.com/203315/facebook-opens-registration-to-all (announcing that Facebook opened membership to people who were not students or corporations); Rachel Rosmarin, *Open Facebook*, FORBES (Sept. 11, 2006, 5:30 PM), http://www.forbes.com/2006/09/11/facebook-opens-up-cx_rr_0911facebook.html (announcing the preliminary decision by Facebook to open the site to the general public).

128. *Statement of Rights and Responsibilities*, *supra* note 25.

129. *See How to Post & Share*, *supra* note 9 (including specific information on how to limit who can see a particular post).

ernment does not act, but allows for private action, would also become state action.

As for the nexus test, governmental immunity provisions do not constitute an actual contract or any other type of legal relationship between the government and social networking sites. Moreover, the backdrop of the immunity provisions indicates that the statute was enacted amid concerns of both non- and overcensorship.[130] The government did not create these provisions to obtain any kind of agreement from social networking sites, but rather gifted certain Internet entities with immunity from civil liability. Treating this relationship as sufficient to satisfy the nexus test would mean that all websites falling within the § 230 immunity provisions could be state actors for purposes of First Amendment analysis, as could any other private entity that receives some sort of government protection, incentive, or subsidy.[131] Extending this position to its extreme conclusion, any copyright holder or patentee could also be considered a state actor because they too benefit from government protection and incentives to partake in certain actions, namely legal monopolies in exchange for innovation. Such a broad interpretation of state action would effectively apply the First Amendment and other constitutional protections to a huge number of private interactions and defeat the purpose of the doctrine as a limit on the U.S. Constitution.

Even if the state action argument fails, users could next argue that social networking sites are public fora, which the government must generally keep open for expressive activities.[132] Free speech fora generally fall into one of three categories: traditional public fora, designated public fora, or non–public fora.[133] Traditional public fora are those locations "which by long tradition or government fiat have been devoted to assembly and debate."[134] In these

130. *See* discussion *supra* notes 29–32 suggesting that the law may have been intended to provide interactive computer service providers with space to develop their own policies rather than to create government intervention.

131. The immunity provision serves as an incentive for social networking sites to engage in some level of censorship and functionally subsidizes social networking site action because it prevents the websites from having to payout liability in civil suits.

132. *See* Whitmore, *First Amendment Showdown*, *supra* note 108, at 339 (discussing the public forum doctrine).

133. Perry Educ. Ass'n v. Perry Local Educators' Ass'n, 460 U.S. 37, 45–46 (1983).

134. *See id.* at 45 (providing streets and parks as examples); *see also* Hague v. Comm. for Indus. Org., 307 U.S. 496, 515 (1939) ("[S]treets and parks . . . have immemorially been held in trust for the use of the public and . . . have been used

fora, the government's ability to restrict speech is very limited and subject to strict or heightened scrutiny, depending on whether the regulation is content-based or content-neutral.[135] Designated public fora are "property which the State has opened for use by the public as a place for expressive activity."[136] As long as the government keeps these areas open—and it is not required to do so—these fora are subject to the same conditions as traditional public fora.[137] Non–public fora consist of "[p]ublic property which is not by tradition or designation a forum for public communication."[138] For these fora, the government has essentially the same rights as a private owner to limit activity and speech, as long as any government regulation of speech meets the less stringent rational basis test.[139]

Social networking sites are clearly not traditional public fora because they did not exist until very recently and thus cannot have the long "tradition" of existing as a space for communication and expression. However, users could argue that social networking sites are designated public fora.[140] In *Marsh v. Alabama*, the Supreme Court held that an individual could invoke First Amendment protections against a private company town because the company town looked and acted like any other municipal entity.[141] The Supreme

for purposes of assembly, communicating thoughts between citizens, and discussing public questions.").

135. *See Perry Educ. Ass'n*, 460 U.S. at 45 ("For the State to enforce a content-based exclusion [in a traditional public forum] it must show that its regulation is necessary to serve a compelling state interest and that it is narrowly drawn to achieve that end. . . . The State may also enforce regulations of the time, place, and manner of expression which are content-neutral, are narrowly tailored to serve a significant government interest, and leave open ample alternative channels of communication." (citation omitted)).

136. *See id.* at 45–46 (providing examples such as university meeting facilities, school board meetings, and municipal theaters).

137. *Id.* at 46.

138. *See id.* at 46; *see, e.g.*, U.S. Postal Serv. v. Council of Greenburgh Civic Ass'ns, 453 U.S. 114, 128–29 (1981) (finding that postal service mailboxes are not public fora); Greer v. Spock, 424 U.S. 828, 838 (1976) (holding that a military base is not a public forum).

139. *See id.* (indicating that government regulations for nonpublic fora need only be reasonable and not aimed at suppression speech).

140. One possible argument users could make is to analogize social networking sites to video games, and then borrow Jack Balkin's comparison of video game platform owners to the company town in *Marsh*. *See* Balkin, *Virtual Liberty*, *supra* note 43, at 2076–77, 2081.

141. 326 U.S. 501, 507–09 (1946).

As discussed, *Marsh* did not clearly state whether the case was decided under the state action doctrine or the public forum doctrine. Given the similarity of the

Court focused on the importance to an individual's free speech rights of being able to speak in traditionally public areas, like a street or city center.[142] Even though in *Marsh* the streets were owned by a private entity, the significant similarity between the private company town and traditional public fora pushed the Supreme Court to find that the private entity violated the First Amendment when it arrested a Jehovah's Witness for passing out brochures.[143] The Supreme Court later extended the metaphor of the city center in other cases to find designated public fora in other private locations, including private malls,[144] although the application of this line of cases to private malls was later overturned.[145]

To invoke the public forum doctrine, a user could argue that social networking sites are becoming the new city center precisely because these sites are a central location for communication.[146] Social networking sites allow people to communicate with their friends, as well as the general public, about any topic of their choice. As mentioned in Part I, these sites are already being used to disseminate information about commerce, politics, the news, and everyday events. Social networking sites are comparable to other designated public fora, such as school board meetings or municipal

public function prong of the state action doctrine and the public forum doctrine, any resulting confusion is not surprising. For example, some scholars characterize the "public function" prong of state action as being the same as the "public forum" test. *See, e.g.*, Michael L. Taviss, Editorial Comment, *Dueling Forums: The Public Forum Doctrine's Failure to Protect the Electronic Forum*, 60 U. CIN. L. REV. 757, 766 (1992) ("The public property-only restriction on the public forum doctrine is merely an application of the state action doctrine . . . .").

However, later cases that drew on *Marsh* narrowed the definition of commercial private entities in the physical world against which First Amendment protections could apply, eventually allowing restrictions on speech by private entities. *See, e.g.*, Hudgens v. NLRB, 424 U.S. 507, 520–21 (1976) (recognizing that *Lloyd* limited the scope of *Marsh* to apply only when "a private enterprise" takes on "all of the attributes of a state-created municipality and the exercise by that enterprise of semi-official municipal functions" (quoting Lloyd Corp. v. Tanner, 407 U.S. 551, 569 (1972))).

For information on the confused and uncertain legacy of *Marsh*, see Esper, *supra* note 109, at 692.

142. *See* Marsh v. Alabama, 326 U.S. 501, 503–04, 507–09 (1946).

143. *Id.*

144. *See* Amalgamated Food Employees Union Local 590 v. Logan Valley Plaza, Inc., 391 U.S. 308, 325 (1968).

145. *See Hudgens*, 424 U.S. at 514–21 (recognizing that *Lloyd* overturned *Logan Valley*).

146. *See supra* Part I.A; *cf.* Balkin, *Virtual Liberty*, *supra* note 43, at 2076–77 (focusing on communication related concerns as an important point for analogizing virtual cities to the company town in *Marsh*).

theaters, precisely because they are necessary for engaging in commerce and obtaining information in today's world. Social networking sites fit even more closely with the idea of the new city center because they provide more opportunities for speech than the other examples of designated public fora. If social networking sites were designated as public fora, the sites may have only limited and indirect ability to regulate speech.[147]

However, the public forum cases refer to when *government* property can be characterized into one of the public forum categories based on various criteria.[148] Social networking sites are not government property. This means that the government does not have the ability to control social networking sites as a forum. Thus as a threshold matter, the public forum doctrine cannot be applied to social networking sites without changing the existing doctrine. Moreover, the public forum doctrine's applicability to private property now appears much more restricted and unlikely given that the Supreme Court has shifted away from *Marsh*, having found that various forms of private property are not public fora.[149]

Even if social networking sites were somehow considered government property, individuals invoking the First Amendment in the context of public fora generally challenge *government attempts* to restrict access.[150] In other words, in order for their private actions to be attributed to the government, social networking sites must also be state actors.[151] As discussed earlier, censorship decisions by private social networking sites are not made by the government, limiting access to the state action doctrine.

Thus a social networking site user who attempts to invoke either the state action or the public forum doctrine would run into quite a few obstacles, suggesting that the First Amendment cannot

147. For public and designated fora, the regulating entity cannot engage in content- or viewpoint-based discrimination. *See* Dawn C. Nunziato, *The First Amendment Issue of Our Time*, 29 YALE L. & POL'Y REV. INTER ALIA 1, 14–16 (2010) (discussing Rosenberger v. Rector & Visitors of the Univ. of Va., 515 U.S. 819 (1995)).

148. *See* Warren v. Fairfax Cnty., 196 F.3d 186, 190–91 (4th Cir. 1999) (noting the use of the following criteria in distinguishing between the different types of fora: "the physical characteristics of the property, including its location; the objective use and purposes of the property; and government intent and policy with respect to the property" (citations omitted)).

149. *See, e.g.*, *Hudgens*, 424 U.S. at 518 (recognizing that a prior case applying public forum doctrine to privately-owned shopping malls had been overturned).

150. *See, e.g.*, Ward v. Rock Against Racism, 491 U.S. 781, 788, 791 (1989).

151. The problems of applying state action were mentioned *supra* Part II.A. The state action and public forum questions can be very closely intertwined, and a court may reach an outcome without clearly stating which doctrine it is using.

successfully be invoked in this context. Given the irregularity in application of both of these doctrines, it is not easy to predict how a court would decide these issues. However, for the aforementioned reasons, it is unlikely that a user would prevail under either a state action or a public forum framework.

B. Consumer Protection Rationales Provide a Basis for Intervention When Market Failures Exist[152]

Consumer protection analysis can provide an alternate basis for finding that government intervention into the social networking site market is appropriate. The most cited view of consumers depicts them as rational calculators seeking an optimal mix of goods and services in the marketplace.[153] In an ideal market, rational suppliers, market competition, and rational consumers with rational preferences lead to efficiency and to an economically optimal outcome.[154] In this view, government intervention is justified only when a market failure exists.[155] There are two common types of market failure: information disparity and power disparity.

152. This Part is only intended to provide a brief overview of consumer protection rationales. For a more in depth analysis of these general principles, see the discussion *infra* notes 153–55.

153. Kysar, *supra* note 46, at 1747–49 (this conclusion relies on a number of assumptions about consumers including the idea that consumers rationally pursue their self-interest and that social welfare is maximized when individual consumers are allowed to pursue their own interests). *See generally* George J. Stigler & Gary S. Becker, *De Gustibus Non Est Disputandum*, 67 AM. ECON. REV. 76, 76 (1977) (proposing that "widespread and/or persistent human behavior can be explained by a generalized calculus of utility-maximizing behavior, without introducing the qualification 'tastes remaining the same'").

I will assume this model is accurate; the discussion of whether consumers act rationally or not goes beyond the scope of this Note. For a critique of this view of people as rational actors, see generally Thorstein Veblen, *Why is Economics not an Evolutionary Science?*, 12 Q.J. ECON. 373, 392–93 (1898) (arguing that a purely economic account of consumer actions ignores psychological and other factors).

154. *See* Kysar, *supra* note 46, at 1747–50; W. Kip Vicusi, *Using Warnings to Extend the Boundaries of Consumer Sovereignty*, 23 HARV. J.L. & PUB. POL'Y 211, 212 (1999).

155. *See* Oren Bar-Gill, *The Behavioral Economics of Consumer Contracts*, 92 MINN. L. REV. 749, 801–02 (2008).

Most scholars agree that these types of market failures exist and instead debate the issue of what form market intervention should take. *See* Richard A. Epstein, *The Neoclassical Economics of Consumer Contracts*, 92 MINN. L. REV. 803, 806 & n.16 (2008) [hereinafter Epstein, *The Neoclassical Economics of Consumer Contracts*]. For those who believe that market forces such as consumer learning or seller education efforts minimize information and power imbalances, intervention should be modest. *See, e.g.*, *id.* at 810–17. In contrast, scholars who see these market forces as

An information disparity exists when one group within the market has more information than other groups.[156] Information is essential to the functioning of a market because it is the basis on which actors make rational decisions.[157] For example, when consumers do not have enough information, they cannot make decisions that accurately reflect their preferences.[158] Information disparities may exist for a number of different reasons, including manipulative marketing and a lack of incentives for suppliers.[159]

---

less powerful advocate for steps that go beyond simple disclosure mechanisms or nonintervention. *See, e.g.*, Bar-Gill, *supra*, at 753–54.

For discussion of forms of consumer protection other than government intervention, see Robert Pitofsky, *Beyond Nader: Consumer Protection and the Regulation of Advertising*, 90 HARV. L. REV. 661, 669–70 (1977).

156. *See* ALAN STONE, REGULATION AND ITS ALTERNATIVES 153 (1982) (discussing what information disparities are and how they develop); *cf.* Pitofsky, *supra* note 155, at 663–67 (describing situations where sellers and buyers in the marketplace have differing amounts of information about products).

157. *See* STEPHEN BREYER, REGULATION AND ITS REFORM 26 (1982); Howard Beales, Richard Craswell & Steven C. Salop, *The Efficient Regulation of Consumer Information*, 24 J.L. & ECON. 491, 492 (1981).

158. Steven P. Croley & Jon D. Hanson, *Rescuing the Revolution: The Revived Case for Enterprise Liability*, 91 MICH. L. REV. 683, 770 (1993); *see* Pitofsky, *supra* note 155, at 664. These authors acknowledge that consumers will never have complete information but suggest that there is certain information that is essential to consumer decision making. *See* Croley & Hanson, *supra*, at 770–71 (indicating that it would be "very unlikely . . . for consumers to be perfectly informed" because rational consumers will only obtain information when the marginal gains are greater than the marginal costs of gathering information); Pitofsky, *supra* note 155, at 663 (noting that constantly changing products and the self-interest of sellers will likely prevent "perfect information").

Information disparities can arise on the other side when producers are unaware of consumer preferences and therefore unable to respond by providing the appropriate goods or services. *See* Croley & Hanson, *supra*, at 779. In some situations, companies may rely on models of "average" customers. *See* A. Michael Spence, *Monopoly, Quality, and Regulation*, 6 BELL J. ECON. 417, 417–18 (1975). Since companies only obtain a profit margin based on the decisions of marginal consumers, the "average consumer" assumption may not result in products that appeal to the margin, creating less than optimal outcomes. *See id.* at 418. Moreover, even average consumers may deviate from "average" preference levels, thereby affecting consumption preferences. *Cf.* Croley & Hanson, *supra*, at 783–84 (discussing marginal and average consumers). If Croley and Hanson are correct, then modeling of "average" consumers would not necessarily result in accurate information for the majority of consumers.

159. *See* Pitofsky, *supra* note 155, at 663–67. In some cases, companies may engage in marketing and advertising campaigns that manipulate consumers into believing certain ideas, even possibly manufacturing demand. *See* JOHN KENNETH GALBRAITH, THE AFFLUENT SOCIETY 124–31 (4th ed. 1998); Thomas A. Cowan, *Some Policy Bases of Products Liability*, 17 STAN. L. REV. 1077, 1087 (1965). *But see* Stigler & Becker, *supra* note 153, at 83–84 (indicating that advertising does not change pref-

When such an information disparity exists, government intervention may be appropriate to force disclosure, which then improves the efficiency of the market because consumers have a better foundation for choosing among products.[160]

Power disparity, the second type of market failure, exists when consumers and producers are not symmetrically situated.[161] When one side/party has superior bargaining power or control over the market, they can take advantage of the other side/party through contracting tactics.[162] For example, standard form contracts often require the consumer to take or leave the entire contract "as is."[163] Consumers cannot negotiate to obtain their preferred terms, lead-

erences by affecting consumer tastes, but rather by affecting the price of goods). In other cases, producers may simply not have an incentive to provide full information to prevent competition or discourage new market entrants. *See* Pitofsky, *supra* note 155, at 665.

Consumer information disparities can arise from consumer missteps as well. In some cases, consumers misunderstand or misperceive accurate information, resulting in choices that do not accurately reflect rational preferences. *See* Alan Schwartz & Louis L. Wilde, *Imperfect Information in Markets for Contract Terms: The Examples of Warranties and Security Interests*, 69 VA. L. REV. 1387, 1431–45 (1983) [hereinafter Schwartz & Wilde, *Imperfect Information*]. In other situations, consumers may not be able to overcome the cost of obtaining information or may choose not to do so because the rational consumer may believe that the marginal utility of gaining the information does not outweigh the cost. *See* Croley & Hanson, *supra* note 159, at 770–71.

160. *See* Pitofsky, *supra* note 155, at 664, 674–75, 701 (providing numerous examples of information that was not disclosed until the government intervened); Schwartz & Wilde, *Imperfect*, *supra* note 160, at 1456.

161. *Cf.* Epstein, *The Neoclassical Economics of Consumer Contracts*, *supra* note 156, at 805 (discussing monopolies). *See generally* Daniel D. Barnhizer, *Inequality of Bargaining Power*, 76 U. COLO. L. REV. 139, 192–223 (2005) (discussing the legal concept of inequalities of bargaining power).

162. *See* Bar-Gill, *supra* note 156, at 791. For example, producers may strategically redesign their products to play off of consumer misperceptions. *Id.* at 801.

163. *See, e.g.*, *Statement of Rights and Responsibilities*, *supra* note 25. These types of standard form contracts are often referred to as "contracts of adhesion." *See* Andrew A. Schwartz, *Consumer Contract Exchanges and the Problem of Adhesion*, 28 YALE J. ON REG. 313, 346 (2011) [hereinafter Schwartz, *Consumer Contract Exchanges*] (providing the example of Google Terms of Service as a contract of adhesion).

Standard form contracts are not bargained for and usually contain terms that are significantly more advantageous for the more powerful party, resulting in less than socially optimal outcomes. *See* Friedrich Kessler, *Contracts of Adhesion—Some Thoughts About Freedom of Contract*, 43 COLUM. L. REV. 629, 632 (1943). *But see* Alan Schwartz, *A Reexamination of Nonsubstantive Unconscionability*, 63 VA. L. REV. 1053, 1057–58 (1977) (arguing that exploitative contracts may be in line with consumer preferences).

ing to a less than socially optimal outcome.[164] This problem is compounded further when producers have the ability to unilaterally change the terms of a standard form contract[165] or to dominate a thin market.[166] As a result, many courts and legislators have viewed standard form contracts with some degree of skepticism and have intervened to prevent enforcement.[167] However, other courts have upheld standard form contracts, including in the context of new technology,[168] except where the contracts have been found to be

164. *See* Schwartz, *Consumer Contract Exchanges*, *supra* note 163, at 346–47, 351 (discussing the lack of bargaining for standard form contracts and that such contracts "force[ ] consumers to submit to organizational domination" (internal quotation marks omitted) (alteration in original)).

165. *See* David Horton, *The Shadow Terms: Contract Procedure and Unilateral Amendments*, 57 UCLA L. REV. 605, 608–09 (2010).

166. Thin markets are situations where consumers do not have adequate alternatives for comparison-shopping such that spot prices are not readily available. *See* Ian Ayres & F. Clayton Miller, *"I'll Sell It to You at Cost": Legal Methods to Promote Retail Markup Disclosure*, 84 NW. U. L. REV. 1047, 1057–61 (1990) (discussing thick and thin markets). In such markets, monopolists can make markups resulting in less than socially optimal allocations of goods and services, and likely contract terms. *See generally* Spence, *supra* note 158, at 420–21. In contrast, markets where consumers can comparison shop and obtain competitive reactions from sellers allow for contract terms to closely align with consumer preferences. *See* Alan Schwartz & Louis L. Wilde, *Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis*, 127 U. PA. L. REV. 630, 638 (1979).

But see Richard A. Epstein, *Behavioral Economics: Human Errors and Market Corrections*, 73 U. CHI. L. REV. 111, 127 (2006) [hereinafter Epstein, *Behavioral Economics*] (arguing that efficient markets will result in standard form contracts that accurately reflect market preferences); Jason Scott Johnston, *The Return of Bargain: An Economic Theory of How Standard-Form Contracts Enable Cooperative Negotiation Between Businesses and Consumers*, 104 MICH. L REV 857, 858 (2006) (discussing how standard form contracts provide benefits beyond lowering the cost of contracting and noting that companies will forgive many violations of terms).

167. *See, e.g.*, Comb v. Paypal, Inc., 218 F. Supp. 2d 1165, 1177 (N.D. Cal. 2002); *see also* Larry A. DiMatteo & Bruce Louis Rich, *A Consent Theory of Unconscionability: An Empirical Study of Law in Action*, 33 FLA. ST. U. L. REV. 1067, 1097–98 (2006) (finding that successful unconscionability challenges more often involved a standard form contract than a negotiated one); Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. CHI. L. REV. 1203, 1204 & n.9 (2003) (indicating that many scholars object to the enforcement of standard form contracts); Arthur Allen Leff, *Unconscionability and the Code—The Emperor's New Clause*, 115 U. PA. L. REV. 485, 553 (1967) (arguing that the disfavor of contracts of adhesion may stem from the view that nonnegotiability and unequal bargaining power are involved); Schwartz, *Consumer Contract Exchanges and the Problem of Adhesion*, *supra* note 163, at 347–48.

168. *See, e.g.*, ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1448–49 (7th Cir. 1996) (upholding a shrink wrap license as binding on software purchaser); Nw. Nat'l Ins. Co. v. Donovan, 916 F.2d 372, 377–78 (7th Cir. 1990) (defending stan-

too one-sided and therefore unconscionable.[169] Where power disparities exist, government intervention could take the form of broadening the doctrine of unconscionability to prevent unequal bargaining power from creating inefficient outcomes.[170] Alternatively, equity-based default rules could be created to reduce instances of unequal bargaining power.[171]

C. Market Failures Arise in the Social Networking Site Market, Allowing Consumer Protection to Provide a Basis for Protecting User Rights of Expression Without the First Amendment

As mentioned in Part I, social networking sites create a platform for communication and provide tools for users to express themselves. Despite intuitive concerns that communicative media do not function in accordance with traditional economic principles,[172] the Federal Communications Commission ("FCC") nevertheless treats media infrastructure as an economic market when

dard form contracts because they decrease the cost of contracting); Feldman v. Google, Inc., 513 F. Supp. 2d 229, 235 (E.D. Pa. 2007).

169. *See* 1 E. ALLAN FARNSWORTH, FARNSWORTH ON CONTRACTS § 4.28, at 581 (3d ed. 2004); Schwartz, *Consumer Contract Exchanges*, *supra* note 163, at 354–55 (discussing current court unconscionability doctrine).

170. Alternatively, courts could also apply an "objective" theory of contracting. *See* Michael I. Meyerson, *The Reunification of Contract Law: The Objective Theory of Consumer Form Contracts*, 47 U. MIAMI L. REV. 1263, 1265–66 (1993) (stating that courts applying an objective view will not assume there has been consent merely based on a signature).

171. *See* W. David Slawson, *Standard Form Contracts and Democratic Control of Lawmaking Power*, 84 HARV. L. REV. 529, 530, 554–56 (1971) (suggesting that beneficial default rules can be derived from the parties' conduct and broader equity considerations); *cf.* Kevin E. Davis & Helen Hershkoff, *Contracting for Procedure*, 53 WM. & MARY L. REV. 507, 549–50, 560 (2011) (indicating that parties might exploit the system to their advantage by seeking short-term rents when opportunities allow, and that empirical review of contracting could allow for the identification of what rules can adequately serve as default procedures and what rules would have to be changed).

172. Some scholars argue that economic analyses of free speech are inappropriate because the premises of economic actions are not the same as the premises which motivate free speech decisions. *See* Whitmore, *First Amendment Showdown*, *supra* note 108, at 334–35. However, even if economic analysis may not fit free speech models generally, it may be appropriate within the context of social networking. *Cf.* Ellen P. Goodman, *Media Policy Out of the Box: Content Abundance, Attention Scarcity, and the Failures of Digital Markets*, 19 BERKELEY TECH. L.J. 1389, 1421–61 (2004) (discussing how media market failures in analog markets exist and apply to digital media).

promulgating regulations.[173] In doing so, the FCC appears to have adopted the "marketplace of ideas" theory of free speech, where the goal of the First Amendment is to provide "an open and competitive market that can supply consumers with the content they want."[174] Although social networking sites often do not charge a membership fee,[175] the relationship between the user and the site is still governed by a contract, and these sites provide many of the same purposes and fee models as other forms of media.[176] Thus as a threshold issue, traditional consumer protection concerns appear relevant and appropriate for social networking sites as a method of identifying market failures and justifying government intervention, particularly with respect to issues of content takedown and user ac-

173. *See, e.g.*, Report and Order and Notice of Proposed Rulemaking, 18 FCC Rcd. 13,620, 13,631 (2003) (stating that the FCC's "core policy objective" of democratic discourse is based on the idea that "the free flow of ideas under-grids and sustains our system of government").

The FCC's view of communications regulation is relevant as it would most likely be the government entity with the authority and ability to regulate Internet media. Although the FCC's views are not controlling, they suggest that economic analysis can be applied.

174. *See* Goodman, *supra* note 172, at 1400.

In addition, although the First Amendment may not directly apply in this scenario, the values established through constitutional free speech doctrine may influence decisions by legislators and courts. *See* Helen Hershkoff, *"Just Words": Common Law and the Enforcement of State Constitutional Social and Economic Rights*, 62 STAN. L. REV. 1521, 1526, 1547 (2010) (discussing the effect of constitutional norms outside of constitutional jurisprudence).

Also, the marketplace of ideas theory is widely accepted in First Amendment scholarship. *See, e.g.*, JOHN STUART MILL, ON LIBERTY ch. 2 (Longmans, Green, & Co. 1921) (1859); Thomas I. Emerson, *Toward a General Theory of the First Amendment*, 72 YALE L.J. 877, 878–79, 902 (1963) (outlining four values in free speech theory, including the marketplace of ideas); Eugene Volokh, *In Defense of the Marketplace of Ideas/Search for Truth as a Theory of Free Speech Protection*, 97 VA. L. REV. 595, 595 (2011) (defending the marketplace of ideas); Whitmore, *First Amendment Showdown*, *supra* note 108, at 326.

175. *See* Jonathan Strickland, *How Do Social Networking Sites Make Money?*, HOW-STUFFWORKS, http://computer.howstuffworks.com/internet/social-networking/information/how-social-networking-sites-make-money.htm (last visited Dec. 31, 2013) (noting that most social networking sites do not use a membership fee model). The membership fee concern probably does not pose an obstacle to applying economic analysis to social networking markets because the market consensus is to provide these services for free. *See id.*

176. *See Statement of Rights and Responsibilities*, *supra* note 25. For a discussion of the purposes of other media and social networking sites, see Strickland, *supra* note 175 (noting that social networking sites function primarily by collecting ad revenue and developer fees); *infra* Part II.D.

count deletion.[177] When viewed in the consumer protection framework, social networking sites present both types of market failures—information and power disparities—described in Part II.B.

Facebook's censorship policies exhibit the information disparity market failure. For example, Facebook did not publicly share its specific guidelines for content takedown until the guidelines were leaked to the public.[178] Until then, users only had access to the general community standards.[179] The examples outlined in Part I demonstrate that Facebook users did not and still do not have a clear understanding of what content would violate the general community standards, signaling an information disparity.[180] For example, Courtney Stodden was surprised when her bikini photos led not only to content takedown but also to full account deletion.[181] Moreover, bikini pictures are relatively common on the site, and Facebook clearly does not remove all similar images.[182] The breastfeeding mothers encountered a similar issue where they too did not expect their pictures to be taken down.[183] In contrast, Facebook allowed pro-rape pages to stay active, even after receiving notice that many people were disturbed, only deciding to take down some of this content later on.[184] Also when Facebook took action against the breastfeeding pictures or pro-rape pages, it appears that the site only deleted the content, not the individual user accounts. When juxtaposed with the pro-rape page, Stodden's punishments are disproportionate; the pro-rape page contained more

177. Traditional consumer protection models may need to be modified to include some fairness concept to account for the fact that platforms both provide services and act as governors. *See* Balkin, *Virtual Liberty*, *supra* note 43, at 2082–83 (discussing dual roles in context of online video game platform owners).

178. *See No Sex, But Crushed Heads O.K.*, *supra* note 98. The leak only provided a portion of Facebook's content guidelines. *Id.*

179. *See supra* notes 91–95 and accompanying text.

180. The particular reason why Facebook, or any other social networking site, does not share specific content regulation information is not important for this discussion. However, see Part II.C for a discussion of various reasons why information disparities may exist.

181. *Cf.* Hughes, *supra* note 2 (Stodden described herself as the victim of "cyber-bullying" when her Facebook fan page—primarily comprised of bikini pictures—was removed for "inappropriate sexual conduct").

182. *See, e.g.*, *Bikini Babes Photos*, FACEBOOK, http://www.facebook.com/pages/Bikini-babes-photos/188537394505550 (last visited Dec. 31, 2013) (a Facebook page dedicated to pictures similar to those Stodden had taken down by the social networking site).

183. *See* Wortham, *supra* note 92.

184. *See* Chen, *supra* note 90.

disturbing content than a scantily clad girl. Facebook's censorship decisions do not provide any coherent pattern for users to predict when their materials will be taken down versus when they will face account deletion.

The unpredictability of Facebook content takedowns and account deletions may be further complicated by a risk of error. In Stodden's case, Facebook claimed that it had taken down the account in error.[185] In another case, Facebook took down the cover art of Nirvana's iconic album, *Nevermind*—which shows a naked baby swimming—for violating the site's "obscenity" policy; Facebook later restored the image and alleged a mistake.[186] If both of these claims are true, then Facebook has explicitly recognized that in at least some cases, its censorship policy leads to mistakes, resulting in the takedown of nonproblematic content and perhaps even the deletion of innocent users' accounts. The existence of error could provide one explanation for some of the irregularities in Facebook's censorship decisions. However, as the breastfeeding and pro-rape pages show, in other situations Facebook intentionally makes decisions to allow or disallow content or user access, and it is unclear which censorship decisions are made upon close review and which are the result of error.

Facebook enjoys a power imbalance relative to its users, which compounds the information disparity market failure. Facebook has an extremely strong market position, as it controls a significantly larger share of the social networking market than its competitors.[187] Social networking sites are more prone to monopolies than other forms of media due to the fact that users connect to other people through their social circles.[188] This network phenomenon creates an additional obstacle to competition because the value of a particular social networking site to an individual user is often tied to the other users present on that site.[189] Facebook's users likely have no sufficient alternative platforms because other social networking sites do not contain all of the same friends or services that are available on Facebook. As the largest social networking site, Facebook controls the bottleneck to a platform of mass communi-

185. Hughes, *supra* note 2.

186. *See* Weiss, *supra* note 98.

187. *See supra* notes 48–53 and accompanying text. For additional information on power imbalances in free speech markets, see Whitmore, *First Amendment Showdown*, *supra* note 108, at 327–34 (describing the market imbalance in higher education).

188. *See* discussion *supra* notes 57–58.

189. *See* Dwight R. Lee & Richard B. McKenzie, *A Case for Letting a Firm Take Advantage of "Locked-In" Customers*, 52 HASTINGS L.J. 795, 796 (2001).

cation, both to a particular social network and to the public more generally because of its market control.[190] The site can thus take advantage of users through the creation of terms of service that give Facebook an even higher degree of censorial power. Complicating matters further, Facebook retains the ability to change terms post hoc, and the site has unilaterally changed its terms of service on multiple occasions.[191]

Government intervention could correct the information disparity by providing users with more information about what content social networking sites are likely to take down and about which speech actions could result in user account deletion. Additionally, intervention could help remedy the existing power imbalances by providing users with more tools to use as leverage against a social network's decisions to censor. However, it should be noted that courts have struck down many attempted regulations of media by the government for insufficient empirical findings or analytic argumentation.[192] Thus even when governmental regulation may be warranted, intervention still needs to be narrowly tailored to withstand judicial scrutiny.

In addition to basic market failures, the free speech policy considerations that underlie First Amendment protections also arise in the social networking context and bolster arguments for government intervention in media markets, even when the First Amendment itself is not applicable. The use of these media is closely related to the First Amendment goal of a diverse public discourse and empowers various sectors of society to speak.[193] Communica-

190. *Cf.* Zarsky, *supra* note 7, at 765 ("[E]xisting media concentration rules are justified . . . to prevent instances in which very few entities control the crucial bottlenecks to the public's attention.").

191. *See, e.g.*, Elinor Mills, *Facebook Changes "Privacy Policy" to "Data Use"*, CBS NEWS (Mar. 23, 2012, 10:44 AM), http://www.cbsnews.com/8301-501465_162-57403181-501465/facebook-changes-privacy-policy-to-data-use/; *Facebook Changes Terms of Use Over Privacy Concerns*, BUSINESS TODAY (Apr. 21, 2012, 12:00 AM), http://businesstoday.intoday.in/story/facebook-changes-terms-of-use-over-privacy-concerns/1/24222.html (indicating that Facebook takes user input but retains the final decision regarding any changes to terms).

192. *See* Howard A. Shelanski, *Antitrust Law as Mass Media Regulation: Can Mergers Standards Protect the Public Interest?*, 94 CAL. L. REV. 371, 391, 419 (2006) (discussing cases striking down governmental regulations); *see also infra* note 204 and accompanying text.

193. *See* Yochai Benkler, *From Consumers to Users: Shifting the Deeper Structures of Regulation Toward Sustainable Commons and User Access*, 52 FED. COMM. L.J. 561, 563 (2000); *cf.* Associated Press v. United States, 326 U.S. 1, 20 (1945) (finding that the First Amendment is best served by "the widest possible dissemination of information from diverse and antagonistic sources").

tion platforms also implicate issues such as "truth and understanding" that are not at issue in purely economic markets.[194] If a social network discriminated against particular users or types of content, the result would be a biased dissemination of certain kinds of information and speech over others. Government intervention would be important to ensure both that people could express themselves and that the social networking user base would receive relatively unbiased information.[195]

Free speech values within the First Amendment are often articulated along three metrics: (1) assuring individual self-fulfillment, (2) as a means of attaining truth, and (3) contributing to political participation and good governance. *See* Emerson, *supra* note 174, at 878–79. Emerson divides the values into four categories. *Id.*

However, more recent theorists have turned the discussion to these three categories and tend to argue that one particular value among these three is the only or the most important free speech value rather than focusing on all of them as Emerson does. *Compare* C. Edwin Baker, *Autonomy and Free Speech*, 27 CONST. COMMENT. 251, 251, 265 (2011) (arguing for the centrality of the autonomy and self-realization rationale), C. Edwin Baker, *Scope of the First Amendment Freedom of Speech*, 25 UCLA L. REV. 964, 964, 966 (1978) (arguing for the liberty model), *and* David A. J. Richards, *Free Speech and Obscenity Law: Toward a Moral Theory of the First Amendment*, 123 U. PA. L. REV. 45, 62 (1974) (adopting a moral theory of the First Amendment that rests the value of free speech "on its deep relation to self-respect arising from autonomous self-determination"), *with* Robert Post, *Participatory Democracy and Free Speech*, 97 VA. L. REV. 477 (2011) (advocating for the participatory democracy theory), James Weinstein, *Free Speech and Political Legitimacy: A Response to Ed Baker*, 27 CONST. COMMENT. 361, 361 (2011) (emphasizing participatory democracy), James Weinstein, *Participatory Democracy as the Basis of American Free Speech Doctrine: A Reply*, 97 VA. L. REV. 633, 633 (2011) (same), Nancy J. Whitmore, *Facing the Fear: A Free Market Approach for Economic Expression*, 17 COMM. L. & POL'Y 21, 27–28 (2012) (advocating for the participatory democracy theory, noting that "expression of popular political sentiment" is necessary for representative democracy), MILL, *supra* note 174, at ch. 2 (arguing that the marketplace of ideas allows truth to emerge), *and* Volokh, *supra* note 174 (defending marketplace of ideas).

194. *See* Goodman, *supra* note 172, at 1422 (internal quotation marks omitted).

195. I recognize that this statement draws on an assumption that objective information or truth is somehow attainable. *See generally* FREDERICK SCHAUER, FREE SPEECH: A PHILOSOPHICAL INQUIRY 19–30 (1982) (noting that the marketplace of ideas theory assumes that some type of objective truth exists and can be found). However this is not a universally agreed upon assumption as some scholars posit that objective truth through the marketplace of ideas simply is unattainable. *See, e.g.*, CATHARINE A. MACKINNON, FEMINISM UNMODIFIED: DISCOURSES ON LIFE AND LAW 155 (1987) (arguing that the marketplace theory allows the powerful to win, not necessarily the truth).

An alternative conception is that the government should help ensure a variety of perspectives, even if not an "unbiased" form of information.

In addition, at least some First Amendment case law suggests that the Supreme Court may not always be persuaded by an anti-distortion position with re-

Some scholars argue that the traditional market failures of information disparities and consumer exploitation either do not exist or are less prominent in the case of the Internet because the Internet facilitates more dialogue and lowers costs of transacting and gathering information.[196] The Internet also has eliminated many problems of scarcity that justified intervention in earlier media markets.[197] These scholars argue that the markets would be able to approach efficient levels on their own,[198] therefore eliminating the traditional bases of government intervention in neoclassical economics discussed in Section B.

However, these claims are not sufficient to eliminate the relevance of consumer protection rationales to social networking sites/speech platforms. The argument that information is efficiently and evenly dispersed holds true only if the platforms are impartial and rational.[199] For this to be the case, social networking sites would have to use some type of rational decision-making calculus to determine that taking down content or eliminating a user outweighs the free speech benefits of that particular content or user. Given the importance of social networking sites as a medium of communication, the Supreme Court's recognition of the importance of commercial information,[200] and the sites' continuing growth in society

spect to freedom of expression. *See* Citizens United v. Fed. Election Comm'n, 130 S. Ct. 876, 903–04, 917 (2010) (striking down laws limiting political spending by corporations, even though other cases upheld similar laws based on an anti-distortion rationale).

196. *See, e.g.*, YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM 30, 99–116 (2006) [hereinafter BENKLER, THE WEALTH OF NETWORKS], *available at* http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf (lauding the informational benefits of modern "social transactional frameworks"); Zarsky, *supra* note 7, at 762–63 ("Within [social] networks data are provided by experienced consumers, advisors and public-interest groups, and even by the vendors themselves (regarding their and their competitors' product) . . . . [C]onsumers in e-commerce . . . tend to consult search engines, forums, and social networks.").

197. *See* Goodman, *supra* note 172, at 1392–93.

198. *See, e.g.*, BENKLER, THE WEALTH OF NETWORKS, *supra* note 196, at 114–16; Zarsky, *supra* note 7, at 763.

199. *See* BENKLER, THE WEALTH OF NETWORKS, *supra* note 196, at 399–408 (discussing Internet service providers and net neutrality). For a critique of the idea that speech can ever be impartial, see MACKINNON, *supra* note 195, at 155.

200. The Supreme Court's more recent cases on "commercial speech" have recognized its importance within the context of the First Amendment. *See* Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc., 425 U.S. 748, 763 (1976) ("As to the particular consumer's interest in the free flow of commercial information, that interest may be as keen, if not keener by far, than his interest in the day's most urgent political debate.").

overall, it does not seem appropriate to defer to a private entity's decisions about when the costs of speech outweigh the benefits, especially when the social networking sites have not demonstrated a rational pattern of action.[201] Also while scarcity may not exist for the Internet in general, social networking markets demonstrate first-mover gains, such as monopolistic tendencies and high costs of switching.[202] As mentioned earlier, social networking sites are partially premised on the ability to connect to other people and create a network. Once a network has been established, however, a user would not gain the same benefits by switching to another platform unless a large portion of the same people were present and the alternate platform provided the same or similar services to the initial platform. The nature of social networks artificially creates scarcity in the market by locking users into a particular website even when there may be alternatives.

D. Social Networking Sites Are More Aptly Characterized as Common Carriers Than Newspapers, Although in Reality They Fall Somewhere in Between Those Poles

Although consumer protection provides a useful foundation for government intervention in social networking site markets, regulation under this framework would still run into some obstacles. Most significantly, social networking sites could invoke First Amendment protections to resist government intervention and counter free speech claims by users.[203] In fact, the Supreme Court has

201. The examples from Part I provide strong evidence that social networking sites have not been following any pattern of censorship that measures the value of speech against censorship.

In addition, some court decisions could be interpreted as rejecting deference to private entity decision calculus. For example, in *Marsh*, the Supreme Court did not want a company town to be able to decide what speech residents should hear, specifically mentioning the importance of listeners in deciding that a Jehovah's Witness' free speech interests outweighed a company town's property rights. Marsh v. Alabama, 326 U.S. 501, 503, 507–09 (1946).

202. *See supra* notes 187–91. First-mover gains refer to when the first entrant within a market gains advantages against later entrants due to various factors including the nature of the market, resource preemption, and switching costs. *See* Marvin B. Lieberman & David B. Montgomery, *First-Mover Advantages*, 9 STRATEGIC MGMT. J. 41, 41–42 (1988). In this case, the nature of social networks provides a first-mover advantage because social networking sites rely on the connections between users. *See* Pasquale, *supra* note 45, at 153 (noting the existence of "switching costs" on social networking sites). The high switching costs lead to the development of social networking site monopolies. See discussion *supra* notes 187–91.

203. As a general matter, it appears that the issue of competing rights between two private entities arises in cases where both entities engage in expression.

struck down many government regulations of the Internet as violative of the First Amendment.[204]

Within traditional First Amendment doctrine, different types of media are given different scopes of protection and deference with respect to censorship and content control.[205] On one end of the spectrum, newspapers are given a high degree of editorial discretion in choosing what content they publish.[206] These actions tend not to raise concerns of censorship. At the other end, telephone companies—characterized as common carriers—cannot regulate the content that passes through their lines at all.[207] Broadcast, cable, and other media are placed between these two extremes and receive some amount of discretionary control.[208]

I suggest that social networking sites also fall somewhere in between the two poles. Just as broadcast and cable media have special

See, e.g., Hurley v. Irish-Am. Gay, Lesbian and Bisexual Grp. of Bos., Inc., 515 U.S. 557, 559–62, 581 (1995) (resolving free speech conflict that arose when a private organization authorized to put together a St. Patrick's day parade excluded the Irish-American Gay, Lesbian and Bisexual Group of Boston).

In this case, it is not true that social networking sites actually speak or engage in expression. Thus a court might perhaps find that there is no actual conflict. However, it is difficult to predict whether a court would view the social networking site as speaking in some way or simply acting as a conduit, and various scholars have found such conflicting rights scenarios to exist in analogous cases. *See* discussion *supra* note 43.

204. *See, e.g.*, Ashcroft v. Free Speech Coalition, 535 U.S. 234, 239–40, 244, 258 (2002) (finding portions of the Child Pornography Prevention Act to be overbroad and violative of the First Amendment); Reno v. Am. Civil Liberties Union, 521 U.S. 844, 849, 858–61 (1997) (striking down portions of the Communications Decency Act of 1996, which sought to protect minors from harmful material on the Internet); Am. Civil Liberties Union v. Mukasey, 534 F.3d 181, 184 (3d Cir. 2008) (affirming injunction against enforcement of Child Online Protection Act), *cert denied*, 555 U.S. 1137 (2009). *But see* United States v. Am. Library Ass'n, 539 U.S. 194, 198–99 (2003) (upholding Children's Internet Protection Act).

205. *See, e.g.*, Turner Broad. Sys., Inc. v. Fed. Commc'ns Comm'n, 512 U.S. 622, 656 (1994) (distinguishing discretion of cable operators from that of newspapers); Columbia Broad. Sys., Inc. v. Democratic Nat'l Comm., 412 U.S. 94, 110, 117–18 (1973) (explicitly recognizing that broadcasters do not receive the same amount of editorial discretion as a private newspaper).

206. *See* Pittsburgh Press Co. v. Pittsburgh Comm. on Human Relations, 413 U.S. 376, 391 (1973) ("reaffirm[ing] unequivocally the protection afforded to editorial judgment" for newspapers).

207. *See generally* James B. Speta, *A Common Carrier Approach to Internet Interconnection*, 54 FED. COMM. L.J. 225, 254–55, 261–68 (2002) (describing common law "common carrier" duties and the application of such duties to telephone companies).

208. *Columbia Broad. Sys.*, 412 U.S. at 110, 117–18 (recognizing that broadcasters do not receive the same amount of editorial discretion as a private newspaper but that they still exercise a "large measure of journalistic freedom").

considerations that require treating them differently from newspapers and telephones, social networking sites also have concerns that require an intermediate approach. This characterization implies that social networking sites should receive less than full editorial discretion but greater than a complete ban on content regulation and therefore that some forms of government intervention should not be struck down as infringing on the social networking sites' First Amendment rights.

Newspapers have traditionally been allowed to exercise editorial discretion in choosing what stories to publish and what content to withhold or censor. More specifically:

> The power of a privately owned newspaper to advance its own political, social, and economic views is bounded by only two factors: first, the acceptance of a sufficient number of readers—and hence advertisers—to assure financial success; and, second, the journalistic integrity of its editors and publishers.[209]

The wide discretion given to newspapers was primarily grounded in the newspapers' importance as a medium for political discussion and critique of the government.[210] Providing newspapers with a wide berth of discretion in what to publish was necessary to foster political discussion and to prevent individual actors from being able to force newspapers to self-censor potentially contentious subject matter.[211] Historically an individual denied access to one newspaper could go to another because there were so many to choose from.[212]

In contrast to newspapers, broadcast media presented a different problem because of the physical restriction of limited frequen-

209. *Id.* at 117.

210. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 266, 272, 278 (1964) (protecting a newspaper that published an inaccurate civil rights advertisement against liability). The wide discretion given to newspapers is also derived, at least in part, from the history of prior restraints in England, which had been used by the crown to silence dissenting speech. *See generally* David A. Anderson, *The Origins of the Press Clause*, 30 UCLA L. REV. 455, 463–64, 488–93 (1983).

211. *See Sullivan*, 376 U.S. at 300 (Goldberg, J., concurring) ("And if newspapers, publishing advertisements dealing with public issues, thereby risk liability, there can be little doubt that the ability of minority groups to secure publication of their views on public affairs and to seek support for their causes will be greatly diminished.").

212. *See* Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 251 (1974). Even when the number of newspapers decreased, the Supreme Court continued to uphold broad editorial discretion for newspapers. *See id.* at 251, 258.

cies.[213] As such, not all people who desired to share content with others through television and radio could do so, regardless of their financial resources. The Supreme Court's concern over monopolies in the context of broadcasting resulted in government intervention through practices such as the fairness doctrine.[214] The fairness doctrine recognized that broadcasters have more limited First Amendment protections because the "right of the viewers and listeners . . . is paramount."[215] Both the rights of viewers and the scarcity of available frequencies justified the placement of the government as a public trustee of the airwaves, giving rise to the FCC's authority to regulate broadcast media more than newspapers.[216] In the context of cable, the Supreme Court found that regulations requiring cable companies to carry broadcast frequencies did not violate the First Amendment because the regulations were content-neutral (implicating lesser scrutiny)[217] and served a substantial government interest (Congress had acted in order to ensure the viability of broadcast

213. *See* Red Lion Broad. Co. v. Fed. Commc'ns Comm'n, 395 U.S. 367, 376 (1969) (recognizing that broadcast frequencies were physically limited).

214. *See, e.g.*, *id.* at 375–86 (outlining the historical development of the fairness doctrine amidst concerns of the public interest and ineffective control by the private sector).

However, in 2011, the FCC formally deleted the language allowing the fairness doctrine from media industry rules, finally solidifying the repeal of the doctrine begun in 1987. *See* Brooks Boliek, *FCC Finally Kills off Fairness Doctrine*, POLITICO (Aug. 22, 2011, 3:22 PM), http://www.politico.com/news/stories/0811/61851.html.

215. *See Red Lion*, 395 U.S. at 390; *see also* Columbia Broad. Sys., Inc. v. Fed. Commc'ns Comm'n, 453 U.S. 367, 377–79, 397 (1981) (upholding statutory right of access under the Federal Election Campaigns Act allowing FCC to revoke licenses from stations that refused to allow access to qualified candidates).

In addition, regulation of broadcast media has also been justified by its "pervasive" nature. *See* Christopher S. Yoo, *The Rise and Demise of the Technology-Specific Approach to the First Amendment*, 91. GEO. L.J. 245, 293–95 (2003) (noting that one concern about broadcast was its ability to intrude upon citizens, in contrast to newspapers, which a reader had to actively seek out). This concern about media intrusion provided a second justification for government intervention, especially as the scarcity arguments became less valid. *See* Fed. Commc'ns Comm'n v. Pacifica Found., 438 U.S. 726, 729, 748–49 (1978) (upholding FCC regulation of George Carlin's "Filthy Words" monologue, in part because of the "pervasive" nature of broadcast media).

216. *See Columbia Broad. Sys., Inc.*, 453 U.S. at 394–95 (citing earlier decisions indicating that the government is the "ultimate arbiter and guardian of the public interest" and that the government's role is to ensure that the rights of the public are protected (internal quotation marks omitted)).

217. Turner Broad. Sys., Inc. v. Fed. Commc'ns Comm'n, 512 U.S. 622, 643–45, 662 (1994).

media for people who did not have cable access).[218] These special considerations grounded the Supreme Court's decision to allow regulation of cable media. However, *TBS v. FCC*[219] noted that regulations singling out one type of media would normally be subject to heightened scrutiny.[220] Thus even though the Supreme Court allowed regulation in these contexts, it determined that Congress' purpose was still to encourage these media to develop with the "widest journalistic freedom consistent with its public obligations" and refused to characterize these media as common carriers, thereby limiting the scope of governmental regulatory power.[221]

Telephones have a long history of being categorized as common carriers.[222] At common law, a common carrier was initially defined as "one who holds himself out to the public . . . offering his services to the public generally."[223] "The distinctive characteristic of a common carrier [was] that he under[took] to carry for all people indifferently, and hence [was] regarded in some respects as a pub-

218. *Id.* at 661–62 (explaining that the must-carry provisions were justified because of two special characteristics: "the bottleneck monopoly power exercised by cable operators and the dangers this power poses to the viability of broadcast television").

219. *Id.*

220. *See id.* at 640–41 (noting that "laws that single out the press, or certain elements thereof . . . 'pose a particular danger of abuse by the State' . . . and so are always subject to at least some degree of heightened First Amendment scrutiny" (citation omitted)).

221. *See* Fed. Commc'ns Comm'n v. League of Women Voters of Cal., 468 U.S. 364, 366, 402 (1984) (invalidating a law that prevented any noncommercial educational broadcasting station from exercising editorial discretion); Columbia Broad. Sys., Inc. v. Democratic Nat'l Comm., 412 U.S. 110, 117–18 (1973).

222. *See* Fed. Trade Comm'n v. Verity Int'l, Ltd., 443 F.3d 48, 57 (2d Cir. 2006) (noting that Congress passed the Mann-Elkins Act in 1910 designating telephones as common carriers).

Common carrier arguments also require some discussion of Equal Protection and/or Due Process, which is beyond the scope of this Note. For further discussion, see David S. Bogen, *Why the Supreme Court Lied in Plessy*, 52 VILL. L. REV. 411, 413 (2007) (discussing cases where the Supreme Court encountered Equal Protection issues in common carrier situations).

223. Lone Star Steel Co. v. McGee, 380 F.2d 640, 643 (5th Cir. 1967) (quoting Kelly v. Gen. Elec. Co., 110 F. Supp. 4, 6 (E.D. Pa.), *aff'd*, 204 F.2d 692 (3d Cir. 1953), *cert. denied*, 346 U.S. 868 (1953) (internal quotation marks omitted). In contrast, private carriers were those who provided services by engaging in individualized contracts with each customer. *See* Kieronski v. Wyandotte Terminal R.R., 806 F.2d 107, 109 (6th Cir. 1986) ("Another category of carriers that are not considered to be 'common carriers,' is that of private carriers . . . . Private carriers haul for others, but only pursuant to individual contracts, entered into separately with each customer." (citation omitted)).

lic servant."[224] The Supreme Court then stepped in, providing a definition of a common carrier as "one that makes a public offering to provide communications facilities whereby all members of the public who choose to employ such facilities may communicate or transmit intelligence of their own design and choosing," essentially serving as a pipeline of communication.[225] As such, the government has broad authority to regulate common carriers and ban content discrimination.[226]

Social networking sites are somewhat difficult to categorize because they have characteristics that could place them under any of these labels. First, social networking sites are open to the public. Anyone who wants an account can create one, the only condition being that the prospective user has an e-mail account.[227] Even if the

224. *Kieronski*, 806 F.2d at 108.

225. Fed. Commc'ns Comm'n v. Midwest Video Corp., 440 U.S. 689, 701, 703 (1979) (internal quotation marks omitted) (noting that Congress did not intend to treat broadcast companies as common carriers and allowed those entities to engage in access and content selection to some extent).

However, these definitions should not be taken as being clear statements of the law. The FCC, for example, has had much difficulty defining "common carrier" for the purpose of the Communications Decency Act. *See* Phil Nichols, Note, *Redefining "Common Carrier": The FCC's Attempt at Deregulation by Redefinition*, 1987 DUKE L.J. 501, 512–13 (1987). In some instances, the FCC has adopted the common law definition of common carriers mentioned above, but the FCC has also sought to create a definition that focuses on the "situation of the provider within the market." *See id.* (stating that the FCC's more economic definition looks at whether a particular entity has competition in the market or is largely a monopoly). Also the FCC "has long held that all those who provide some form of transmission services are not necessarily common carriers" and "did not subject to common-carrier regulations those service providers that offered enhanced services over telecommunications facilities, but that did not themselves own the underlying facilities." Nat'l Cable & Telecomm. Ass'n v. Brand X Internet Servs., 545 U.S. 967, 993 (2005) (internal quotation marks omitted). The FCC's difficulty in creating a definition is due in part to the fact that the FCC's definition must fit with legislative directives that designate certain entities as common carriers and others as not. *See Midwest Video*, 440 U.S. at 702–04 (discussing the limitations on the FCC's attempts to treat certain entities as common carriers due to restrictions embedded in Congressional legislation). Moreover, attempting to change the status of a technology can incite serious debate. *See supra* note 45. The inclusion of new technology within the ambit of the FCC's competence further complicates the agency's search for a definition. *See supra* note 45.

226. Speta, *supra* note 207, at 110, 117–18.

227. *See Create an Account*, FACEBOOK, http://www.facebook.com/help/create account (follow "How do I sign up for Facebook?" hyperlink) (last visited Dec. 31, 2013) ("To sign up for a brand new account, enter your name, birthday, gender, and email address . . . ."); *id.* (follow "How old do you have to be to sign up for Facebook?) (noting the minimum age of thirteen); *How to Create an Account on LinkedIn*, EHOW, http://www.ehow.com/how_2030914_account-linkedin.html (last

sites had more restrictive conditions in the past, they have changed their practices to be more open, essentially serving the public.[228] Second, social networking sites generally post content exactly as the user submits it.[229] The sites provide a platform on which users can communicate with each other, appearing to "serve as conduits for the speech of others."[230]

These factors might suggest that social networking sites should be treated as common carriers. However, the relationship between users and a site is governed by private contracts, a factor pointing against common carrier status.[231] Moreover, most modern instances of common carrier status have resulted from Congressional action.[232] Finally, social networking sites do take some steps to censor content or account holders, distinguishing these sites from traditional common law common carriers.[233]

visited Dec. 31, 2013) (noting that LinkedIn only requires a name and an email address); *How to Sign Up on Twitter*, TWITTER, http://support.twitter.com/articles/100990-how-to-sign-up-on-twitter (last visited Dec. 31, 2013) (only requiring a name and an e-mail address). Although Facebook was originally limited to university students, the website is now open to anyone over the age of thirteen. *See* sources cited *supra* note 127.

228. I realize that my discussion of a service to the public in the context of the common carrier doctrine may appear to be in tension with my discussion of public function for the purposes of state action in Part II. However, these two concepts are not the same; an entity may be open to the public to the extent that it provides services to a significant number of people (the general public) while at the same time not providing a "public function." As discussed earlier, "public function" is a term of art within state action discourse, distinct from my use of the term "public" here in its more general sense.

229. *See How to Post a Tweet*, TWITTER, http://support.twitter.com/groups/31-twitter-basics/topics/109-tweets-messages/articles/15367-how-to-post-a-tweet (last visited Dec. 31, 2013) (stating that when a user clicks the "Tweet" button, the user "will immediately see [the] Tweet in the timeline on [the user's] homepage"); *How to Post & Share*, *supra* note 9; *Share, Star and Hide Stories*, FACEBOOK, http://www.facebook.com/help/106105072867502/ (last visited Feb. 19, 2013) (describing Facebook Timeline and what happens when a user posts content).

230. Denver Area Educ. Telecomms. Consortium, Inc. v. Fed. Commc'ns Comm'n, 518 U.S. 727, 793 (1996) (Kennedy, J., concurring) (describing cable providers as conduits for the speech of others when they provided service for public access channels under franchise agreements).

231. *See* Kieronski v. Wyandotte Terminal R.R., 806 F.2d 107, 109 (6th Cir. 1986) (noting that carriers who carried for others based on separate contracts were not considered common carriers).

232. *See* Speta, *supra* note 207, at 258–68.

233. *See id.* at 254, 261 (noting that the common carrier doctrine "required those engaged in serving the public to, in fact, serve the entire public with reasonable care" and as applied to communication carriers like telephone companies, this included a duty of nondiscrimination (noncensorship)).

Social networking sites' censorship authority might then suggest that these sites should most appropriately be treated like newspapers. However, the basic premise of social networking sites is that users can share whatever information they want with their network. And it appears that the vast majority of content is posted directly and is not subject to any kind of censorship.[234] Moreover, even when the platforms take down material, that content often is unprotected speech under the First Amendment.[235] As such, social networking sites do not have the same history of broad editorial discretion of only accepting limited content among protected speech, as newspapers have traditionally done. Instead, all of these factors point to the conclusion that social networking sites should be treated similarly to cable and broadcast media. Although the Supreme Court has held that the special circumstances justifying intervention in the broadcast and cable context do not apply in the context of the Internet generally,[236] the largest social networking sites control a significant part of the user market such that transferring to another network is difficult and switching to another site would not realize the same benefits unless a substantial portion of a user's network also moved to the new network.[237] As a result, the social networking site market may have an artificially created scarcity of suppliers. Additionally, Internet services implicate the same communications concerns that pushed telephony to be categorized as a common carrier in the first place: an important connection to

234. This claim cannot be supported by direct evidence because social networking sites do not publicize every instance in which they take down material or delete a user's account for violating the content guidelines. Additionally, given the sheer number of users who have accounts on these websites, the number of instances of takedown or account deletion that are publicly available is very low.

235. *See Facebook Community Standards*, *supra* note 25; *see, e.g.*, Weiss, *supra* note 98 (listing examples of material removed by Facebook for alleged obscenity). Facebook is not alone in making these allegations, as other social networking sites such as Twitter also claim to only take down illegal content, which would not be protected under the First Amendment. *See* Alex Howard, *On Twitter, Censorship and Internet Freedom*, GOV20.GOVFRESH (Jan. 26, 2012, 7:12 PM), http://gov20.govfresh.com/on-twitter-censorship-and-internet-freedom (Twitter's new censorship policy claims to only take down content that is illegal in other countries); George Stahl, *Twitter CEO: New Policy for Transparency, Not Censoring*, WALL ST. J., Jan. 31, 2012, http://online.wsj.com/article/SB10001424052970204740904577194021894304072.html (same).

236. *See* Reno v. Am. Civil Liberties Union, 521 U.S. 844, 868 (1997) (noting that scarcity and invasiveness concerns "are not present in cyberspace").

237. *See* Pasquale, *supra* note 45, at 153 (noting the existence of "switching costs" on social networking sites); *cf. The Power of Social Networks*, *supra* note 57 (identifying the role of friends as an important element in seeking membership on social networking sites).

economic growth, popular support for basic access to communication, and the difficulty of using post hoc solutions to resolve disputes over content control.[238] Some of these communication concerns also came up in the context of broadcast and cable, serving as at least part of the justification for restricting editorial discretion for those media.[239] In the case of social networking sites, these concerns are more prevalent because of the sizeable portion of the population obtaining news through these sites, the depth of users' reliance on these sites as platforms for communicating and expressing themselves, and the growing importance of access to commercial information.

Thus the overall balance of the current nature of social networking sites, with particular emphasis on the First Amendment goal of ensuring free speech, firmly distinguishes social networking sites from newspapers, placing the sites between broadcast/cable and common carriers. As a result, some amount of regulation affecting censorship by social networking sites may be able to withstand judicial scrutiny.

III.
TO PROTECT THE WORLD FROM CENSORSHIP: EVALUATING POTENTIAL REGULATIONS OF SOCIAL NETWORKING SITES WITHIN A CONSUMER PROTECTION FRAMEWORK

This Part will analyze the following types of potential government intervention: establishing clearer content regulation standards, forcing greater disclosure of censorship policies, implementing non-litigation-based remedies, creating a cause of action, and eliminating censorship authority entirely. All of these eval-

238. *See* Susan P. Crawford, *Transporting Communications*, 89 B.U. L. Rev. 871, 884–86, 913–14 (2009) (identifying traditional principles for government involvement in communication and infrastructure and demonstrating how they arise in the context of the Internet).

239. *See, e.g.*, Turner Broad. Sys., Inc. v. Fed. Commc'ns Comm'n, 512 U.S. 622, 663–64 (1994) (justifying cable regulations in part because "broadcasting is demonstrably a principal source of information and entertainment for a great part of the Nation's population" and "assuring that the public has access to a multiplicity of information sources . . . promotes values central to the First Amendment" (internal quotation marks omitted)); *cf.* Columbia Broad. Sys., Inc. v. Democratic Nat'l Comm., 412 U.S. 94, 112–13 (1973) (indicating that the "Fairness Doctrine" was premised on the importance of "the right of the public to be informed").

uations will take place in a theoretical context, as there is not yet any comprehensive data on social network censorship.[240]

The first two interventions—establishing clearer content censorship standards and forcing disclosure—address the issue of information disparities by creating tools that would provide users with more information about what content the social networking sites find objectionable and the resulting penalties, if any, for posting that content. Additionally, the highly limited nature of these government interventions makes them more feasible from a practical standpoint. However, these solutions do not change the nature of user rights with respect to social networking sites.

The third and fourth types of intervention—creating a nonlitigation-based remedy and creating a cause of action—respond to the issue of unequal bargaining power. By providing users with some mechanism of dispute resolution external to the social networking site, there would be more oversight of the relationship between users and the sites. Moreover, the threat of external oversight could put pressure on social networking sites to be more responsive to users' concerns about speech restrictions. While these interventions would give users more power, they would import more cost to the social networking site model and would be much more difficult to enact than the first two interventions.

The final suggestion—completely eliminating censorship power—would in essence treat social networking sites as a common carrier, like the telephone. Social networking sites would not be able to censor any content that is posted on their platform and would only serve as a conduit for communication. The websites would not need to be concerned with liability, as the Communications Decency Act of 1996 immunity provision would still apply, preventing social networking sites from being treated as a speaker or publisher of user-generated content. However, this intervention would negate whatever benefits there are to allowing social networking sites to censor[241] and would be the most difficult to enact.

240. At least one of these suggestions already exists to a limited degree. Google and Twitter voluntarily post some but not all of the removal requests they receive on Chilling Effects (located at http://www.chillingeffects.org). *See* Sengupta, *supra* note 99. However, it is not clear what criteria these sites use when deciding what requests to post, and other sites do not post their takedowns. For further discussion of Chilling Effects, see *infra* notes 246–50 and accompanying text.

241. *See* Holland, *supra* note 33, at 369–70, 391 (describing service provider community developments which could result from censorship immunity); Zarsky,

A. Establishing Clearer Content Regulation Standards

Of all of the potential interventions considered, establishing clearer standards is the least disruptive to the existing systems. Social networking sites would simply be required to provide fairly comprehensive, although perhaps not exhaustive, guidelines for which types of content are permissible and which penalties would be imposed for violating content standards. Social networking sites would not necessarily have to change their content takedown policies as long as users knew the bounds of these policies. In terms of feasibility, even individuals who oppose government intervention would be willing to accept clarification or disclosure of content regulation terms.[242]

Clarification or disclosure of content regulation terms would not be difficult for Facebook. As shown by the recent leak of the website's content restriction guidelines,[243] Facebook already has fairly specific standards in place that its censors should follow when deciding whether content violates Facebook's terms of use. Additionally, Facebook has provided clear information in other contexts, such as in its privacy policies.[244] For new or potential users, these standards will provide more information as they decide whether to even create an account on Facebook. For some existing users, the standards would help them decide whether to stay on the site or to try to shift their social network to a different website.[245] For other existing users, the clearer standards would simply help them make decisions about whether to share certain content. However, this government intervention would not change the ability of users to influence the social networking site's censorship policies,

supra note 7, at 778 (describing the benefits of and need for filtration of information by online social networks by methods including accreditation).

242. *See, e.g.*, Epstein, *Behavioral Economics*, *supra* note 166, at 128 (approving of the Truth in Lending Act disclosure provisions for credit card companies but unwilling to advocate any further steps).

243. *See supra* notes 95–98 and accompanying text.

244. *See* Kevin J. O'Brien, *Facebook Offers More Disclosure to Users*, N.Y. TIMES, Apr. 12, 2012, http://www.nytimes.com/2012/04/13/technology/facebook-offers-more-disclosure-to-users.html?_r=2 (providing users with information about what data Facebook collects from users).

245. Although I discuss the lock-in potential for social networking sites, at least some recent cases suggest that both individual users and sometimes entire social circles have moved social networking sites in response to changes in social networking site policies. *See, e.g.*, Peter Panchal, *Why Facebook is Losing U.S. Users*, PCMAG.COM (June 14, 2011), http://www.pcmag.com/article2/0,2817,2386884,00.asp (one possible explanation for users leaving Facebook may be the website's changing privacy policies). However, it does not appear that this is a normal response to changes in social networking site policies. *See id.*

and therefore the benefits would be limited. Rather it would provide information for users and regulators to engage in a dialogue about the actual content standards applied by social networking sites. This would benefit users if, as a result of this dialogue, social networking sites changed their policies to align with either general community standards or underlying free speech values.

B. Forcing Greater Disclosure on Censorship Policies

Requiring social networking sites to disclose certain types of information regarding their censorship activities is a slightly stronger form of intervention. Social networking sites could be forced to provide some kind of information report when they take down material or delete a user account.[246] One model would require social networking sites to enter information into a censorship database; for each censorship decision, the site would create a database entry noting the date, time, user name, a description of the content, and reason for taking down the content.

Some could argue that such a censorship database already exists. The reporting website Chilling Effects attempts to gather and catalogue censorship decisions by social networking sites.[247] As mentioned earlier, some websites such as Google and Twitter voluntarily provide information about some of their takedown requests to Chilling Effects.[248] However, a quick view of Chilling Effects shows that the reports are based on cease-and-desist letters, which Facebook does not use. Reports for takedown of user-generated content or account deletion for reasons such as indecency or violence by social networking sites would probably fall into the "uncategorized" entries, which are not easily searchable.[249] Therefore the information provided by this website is not very useful for analyzing social networking site censorship behavior. Instead the Chilling Effects website appears to be focused more on takedowns related to

246. *Cf.* Urs Gasser, *Regulating Search Engines: Taking Stock and Looking Ahead,* 8 YALE J.L. & TECH. 201, 233–34 (2006) (suggesting disclosure notification requirements when search engines block or remove content).

247. *See* CHILLING EFFECTS, http://chillingeffects.org (last visited Dec. 31, 2013).

248. *See* Sengupta, *supra* note 99.

249. *See Report Receiving a Cease & Desist Notice*, CHILLING EFFECTS, http://chillingeffects.org/input.cgi (last visited Dec. 31, 2013). The designation as "uncategorized" is based on the classification structure currently used on Chilling Effects, which does not have a designated category for content takedown by social networking sites or other Internet platforms. *Id.*

copyright or trademark concerns.[250] In addition, this website appears to be mostly driven by voluntary submissions.[251] Thus a centralized list of takedowns and user account deletions due to social networking site content regulations would provide benefits beyond existing resources. By compiling all the relevant information in one place with an organizational structure specifically designed to optimize censorship analysis, a mandatory censorship database would overcome the shortcomings of Chilling Effects.[252]

Forcing disclosure of content takedown and account deletions would provide scholars with the data necessary to identify trends and industry practices regarding how social networking sites approach sensitive or controversial topics.[253] As with clearer standards of content regulation, this intervention would provide users with more information about which types of content would be likely to be taken down.[254] In contrast to clearer guidelines, however, empirical data could also provide legislators or agencies with a strong foundation to create social networking site regulations that could withstand judicial scrutiny.[255]

250. *Id.* The search function on Chilling Effects focuses on categorizing cease-and-desist letters for copyright and trademark issues. *Search the Database*, CHILLING EFFECTS, http://chillingeffects.org/search.cgi (last visited Dec. 31, 2013).

251. *See* CHILLING EFFECTS, *supra* note 246 ("In addition, we want your help. We are gathering a searchable database of Cease and Desist notices sent to Internet users like you. We invite you to input Cease and Desist letters that you've received into our database, to document the chill.")

252. *Cf.* Chandler, *supra* note 45, at 1117 (suggesting a central reporter for search engine content removal).

The FCC is moving toward creating such a centralized registry for television political ad spending, suggesting that it could do the same for website censorship. *See* Brian Stelter, *F.C.C. Pushes for Web Site on TV Political Ad Spending*, N.Y. TIMES, Apr. 8, 2012, http://www.nytimes.com/2012/04/09/business/media/fcc-pushes-for-web-site-on-political-ad-spending-on-tv.html?_r=3 (suggesting that the registry would be approved by the FCC later in April 2012).

253. *Cf.* Alexander Reynolds, *Enforcing Transparency: A Data-Driven Alternative for Open Internet Regulation*, 19 COMMLAW CONSPECTUS 517, 520–21 (2011) (advocating for information collection mechanisms from Internet service providers (ISPs) to learn more about ISP practices).

254. *Cf.* Alexander Reicher, *Redefining Net Neutrality After* Comcast v. FCC, 26 BERKELEY TECH. L.J. 733, 736 (2011) (suggesting that transparency in broadband service would empower consumers).

255. An empirical basis could potentially provide enough support to avoid the overbreadth problems that have plagued previous government regulations of the Internet. *See* cases cited *supra* note 204. *But see* Nunziato, *supra* note 147, at 10–11 (arguing that a disclosure mandate without a nondiscrimination mandate would not sufficiently protect Internet users' free speech rights and that transparency provisions do not protect unpopular speech).

A reporting system would have to take into account additional considerations to prevent harm to the platforms or other parties. For example, social networking sites may need immunity from being held liable as speakers if the reports are made public. A user whose content was taken down as being obscene could arguably bring suit against the social networking site for defamation if the report falsely stated that the user's content was obscene. Although some immunity from defamation actions is provided by the Communications Decency Act of 1996 § 230, this provision may not apply to database entries because the content of a report would not have been "provided by another information content provider" but rather by the "interactive computer service" itself.[256] Access to the reporting system may also need to be limited to regulators or to another subset of actors if there are legitimate needs for secrecy.[257]

C. Implementing Non-Litigation-Based Remedies

A third option would be to create some type of non-litigation-based remedy subject to review. There are two primary models for this kind of alternative dispute resolution ("ADR") strategy that could work for social networking site censorship. On the one hand, social networking sites could provide private ADR. eBay provideees an example this type of remedy. eBay users who have grievances against other users can opt for the services of SquareTrade, an online dispute resolution provider.[258] In adopting an online dispute resolution provider, eBay has recognized that the costs of litigation, and even traditional ADR, are often prohibitive and would leave users with no other means of obtaining a remedy.[259] The main benefit of online dispute resolution services is that all of the communication takes place on the Internet, eliminating many of the costs associated with litigation and traditional ADR, which both require physical presence.[260] An online dispute resolution provider should be external to the social networking site; this would be better than an internal review due to a lesser risk of bias. However, there would

256. *See* Communications Decency Act of 1996, 47 U.S.C. § 230(c)(1). Websites that create their own content cannot access the immunity provisions of § 230(c)(1). *See id.*

257. Pasquale, *supra* note 45, at 109–10 (discussing how to tailor reporting systems, providing Internet carrier networks and search engines as examples). Greater need for privacy could arise in cases such as when proprietary material or trade secrets were at risk for disclosure. *See id.* at 109.

258. Orna Rabinovich-Einy, *Technology's Impact: The Quest for a New Paradigm for Accountability in Mediation*, 11 HARV. NEGOT. L. REV. 253, 253 (2006).

259. *Id.* at 254.

260. *See id.* at 255.

probably need to be some degree of government oversight to ensure that the decisions of private ADR bodies are fair.[261]

On the other hand, ADR could occur through a governmental organization. Public ADR could be structured along the same lines as existing government-provided mediation services. The Office of the Attorney General of Maryland, for example, provides mediation for consumer disputes between individual consumers and businesses.[262] An aggrieved customer files a complaint with the Office of the Attorney General, who reviews the complaint and determines whether the issue is one that can be mediated by the office.[263] However, this procedure is voluntary and cannot bind the parties to any particular resolution unless both sides agree.[264] As such, if the government were to provide some kind of ADR for aggrieved social networking site users, it may need to be mandatory and binding, thereby having greater legal effect.

This type of non-litigation-based remedy could affect the amount of information users have regarding the practices a social networking site uses in regulating content by providing particular cases of censorship as well as creating patterns of censorship over time. More directly, ADR would provide users with more leverage and bargaining power when content takedown or account deletion occurs. Since the final decision regarding the takedown or deletion would be left to a (presumptively) neutral third-party, users would have a tool to check censorship decisions by social networking sites. At the same time, however, imposing this additional process would raise the cost of maintaining the platform, and the social networking site might choose to shift the costs to the users by turning social networking sites into primarily pay sites or by increasing commercial content. Such an outcome could be harmful to user rights of

261. The risk then would be that the close connection between the repeat player website and the dispute resolution body could foster bias in favor of the repeat player. *Cf.* Marc Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, *in* LAW AND SOCIETY 165 (R. Cotterrell ed., 1994), *available at* http://www.marcgalanter.net/Documents/papers/WhytheHavesComeOutAhead.pdf (discussing how repeat players can gain advantages in the legal system). *But cf.* Rabinovich-Einy, *supra* note 257, at 256, 269 (suggesting that experience can produce accountability without external review).

262. *File a Consumer Complaint*, MARYLAND ATTORNEY GENERAL, http://www.oag.state.md.us/Consumer/complaint.htm (last visited Dec. 31, 2013).

Government arbitration is another option, which may have fewer problems regarding legal enforceability.

263. *About the Complaint Mediation Process*, MARYLAND ATTORNEY GENERAL, http://www.oag.state.md.us/Consumer/complaintmediation.htm (last visited Dec. 31, 2013).

264. *See id.*

expression because it would restrict an important and effective medium of communication currently available to the large majority of Internet users for free.

D. Creating a Cause of Action

Moving along the spectrum of remedies, the step beyond nonlitigation-based remedies would be to create a cause of action that allows users to bring lawsuits against social networking sites for content takedowns or account deletions. Perhaps the simplest way of establishing a cause of action would be to legislatively decree or judicially find that social networking sites are engaged in state action, thereby allowing users to invoke First Amendment protections directly against a social networking site. Users could then challenge individual instances of censorship or a site's broader content regulation policies under doctrines such as overbreadth[265] or as content- or viewpoint-based regulations.[266] Given the aforementioned problems with this approach, courts or the legislature could alternatively create a sui generis law giving users a cause of action against social networking sites. One example of this type of claim is the right of publicity, which was essentially created by courts and then adopted through statutes to protect against invasions of privacy.[267] In this case, courts could recognize that users should have some type of expressive or property right that would allow suits against social networking sites (or other Internet entities) that take down content or delete user accounts.[268]

A cause of action sounds appealing because it would provide users with the greatest amount of procedural protections in com-

265. *See* United States v. Stevens, 559 U.S. 460, 473 (2010) ("[A] law may be invalidated as overbroad if a substantial number of its applications are unconstitutional, judged in relation to the statute's plainly legitimate sweep." (internal quotation marks omitted)).

266. *See* Am. Booksellers Ass'n v. Hudnut, 771 F.2d 323, 325 (7th Cir. 1985) (striking down anti-pornography law as an invalid viewpoint restriction).

267. *See* Haelan Labs., Inc. v. Topps Chewing Gum, Inc., 202 F.2d 866, 868–69 (2d Cir. 1953). In *Haelan*, Judge Frank essentially created the right to publicity as a way of allowing people who became public figures to still have a means of protecting themselves against privacy invasions when their image was misappropriated. *See* Diane Leenheer Zimmerman, *Who Put the Right in the Right of Publicity?*, 9 DEPAUL-LCA J. ART & ENT. L. & POL'Y 35, 43–44 (1998).

States then adopted the right of publicity through statutory enactments. *See, e.g.*, McKinney's Civil Rights Law § 50, NY CIV RTS § 50.

268. For a discussion of issues that may arise due to "legislative imposition of free speech norms to private parties," see Julian N. Eule & Jonathan D. Varat, *Transporting First Amendment Norms to the Private Sector: With Every Wish There Comes a Curse*, 45 UCLA L. REV. 1537, 1539, 1580–1600 (1998).

parison to all of the other remedies previously discussed. However, a cause of action would also come with the most significant costs because it invokes the legal system.[269] Moreover, even if a cause of action existed, there may also be a problem of insufficient incentives to bring suit. Although a violation of First Amendment rights may be considered an injury for the purposes of obtaining a preliminary (and potentially permanent) injunction,[270] it is unclear whether courts would also find damages substantial enough to induce users to incur the costs of litigation. Thus creating a cause of action might realistically create a right in name with no actual method of enforcement. To remedy this problem, legislatures could also create a statutory damages scheme. The scheme would have to allow for relatively high damage awards and possibly attorneys fees in order to overcome the cost barriers.

E. Eliminating Censorship Authority

The final and most extreme solution would be to prevent social networking sites from censoring any posted content. This intervention would functionally treat social networking sites as common carriers, which cannot discriminate based on content at all.[271] Current discussions about nondiscrimination on the Internet occur in the context of the "net neutrality" debate—the issue of whether Internet connectivity providers may discriminate amongst consumers with regard to costs of carriage and preferences for carriage.[272] Although the point of tension for "net neutrality" partisans is not

269. *See generally* JAMES S. KAKALIK & NICHOLAS M. PACE, THE RAND CORP., COSTS AND COMPENSATION PAID IN TORT LITIGATION, vii–x (1986), *available at* http://www.rand.org/content/dam/rand/pubs/reports/2006/R3391.pdf (discussing the sources of high litigation costs in tort cases). *But see generally* Danya Shocair Reda, *The Cost-and-Delay Narrative in Civil Justice Reform: Its Fallacies and Functions*, 90 OR. L. REV. 1085, 1085–86 (2012) (arguing that the costs of litigation are not as high as "common wisdom" claims).

270. *See, e.g.,* Elrod v. Burns, 427 U.S. 347, 373 (1976) (finding that "[t]he loss of First Amendment freedoms, for even minimal periods of time, unquestionably constitutes irreparable injury"). However, this argument assumes that the courts have found some way of adopting the First Amendment in the social networking site context or that they are willing to find a similar harm without explicitly invoking the First Amendment.

271. *See supra* Part II.D.

272. *See* Christopher S. Yoo, *Network Neutrality and the Economics of Congestion*, *supra* note 45, at 1855–60 (providing an overview of the net neutrality debate); *see, e.g.*, Carol M. Hayes, Note, *Content Discrimination on the Internet: Calls for Regulation of Net Neutrality*, 2009 U. ILL. J.L. TECH. & POL'Y 493, 499–500 (2009) (discussing net neutrality issues such as content-based censorship).

based on content, some of the arguments from the access nondiscrimination debate resonate with content censorship issues.[273]

Proponents of "net neutrality" argue that the benefits of communication and discussion on the Internet can only be attained if Internet intermediaries do not discriminate in granting access to the Internet.[274] In the same vein, preventing social networking sites from restricting user speech may further communication and free speech by ensuring that all views are heard. Proponents also argue that Internet intermediaries function as communication conduits and therefore should be treated similarly to other common carriers that cannot engage in any discrimination.[275] As described in Part II, social networking sites usually post user content without modification and also appear to function as conduits of communication, suggesting that there would be some basis in the nature of social networking sites to suggest that they too should not be allowed to discriminate based on content.

Opponents, on the other hand, argue that the market will provide sufficient incentives to facilitate communication.[276] For social networking sites, the desire to draw the largest number of individual users, businesses, and public traffic could incentivize the sites to put forward policies that most closely reflect the interests of users. Moreover, opponents of "net neutrality" claim that intervention in the market could result in negative consequences, such as forcing Internet service providers ("ISPs") to engage in wasteful activities, such as bureaucratic checkpoints or expansion of regulation beyond simple assurance of nondiscrimination due to administrative mission creep.[277] Analogously, preventing social networking sites

273. At least one scholar has found analogies between the two contexts and started to include social networking sites in access discrimination discussions. *See* Pasquale, *supra* note 45, at 152.

274. *See* Tim Wu, *The Broadband Debate, A User's Guide*, 3 J. ON TELECOMM. & HIGH TECH. L. 69, 72–73 (2004) (describing the Openist view that suggests openness as the basis for innovation and positive externalities).

275. *See* Nunziato, *supra* note 147, at 2–3 (discussing how broadband providers are common carriers and conduits for communication and that free speech issues should not be left to the market).

276. *See* Wu, *supra* note 273, at 76 (outlining the Deregulationist view that believes the Internet developed well in part due to little intervention by the government).

277. *See* TIMOTHY B. LEE, CATO INST., POLICY ANALYSIS NO. 626: THE DURABLE INTERNET: PRESERVING NETWORK NEUTRALITY WITHOUT REGULATION 30–32 (2008), *available at* http://www.cato.org/sites/cato.org/files/pubs/pdf/pa-626.pdf (providing examples of bureaucratic cost and mission creep—the expansion of agency power beyond the purpose for which it was given—in the context of the Interstate

from being able to censor any content would undermine the value of censorship[278] and could add similar costs.

As mentioned earlier, we do not have full information about how social networking sites engage in censorship. At least one net neutrality scholar has argued that the resolution of whether or how to apply net neutrality should be decided after enforcing transparency so that regulators know how networks are discriminating.[279] The information-sharing interventions mentioned above may be prerequisites to determining whether nondiscrimination is desirable or even necessary. Assuming that content nondiscrimination was proposed, it would encounter implementation issues.[280] Additionally, since social networking sites have access to existing immunity provisions, it may be difficult to fully remove social network censorship powers, as the Communications Decency Act of 1996 would also have to be amended.

Commerce Commission's regulation of railroad rates and the FCC telephone regulations).

278. Reducing or eliminating social networking site censorship capabilities would undermine the benefits of censorship. *See* Zarsky, *supra* note 7, at 778 (describing the benefits of and need for accreditation of information by online social networks).

Social networking site censorship may even help prevent physical harm in some cases. *Cf.* Amy Summers, *Is Social Networking More Dangerous to Teens Than "Stranger Danger"?*, SOCIAL TIMES (Jan. 7, 2011, 9:46 AM), http://socialtimes.com/is-social-networking-more-dangerous-to-teens-than-stranger-danger_b33468 (noting certain dangers on social networking sites such as cyber-bullying and incidents of murders that have occurred from sharing information on social networking sites).

279. *See* Reicher, *supra* note 253, at 734–35.

280. The net neutrality debate demonstrates that there is conflicting authority on which government actor is the most appropriate to regulate the Internet, which may also arise in the context of regulating social networking sites. *See* Hayes, *supra* note 271, at 494. Congress has previously considered net neutrality regulation, although it did not enact any of the proposals. *See id.* at 501–02. Courts have also stepped in to the extent that the Supreme Court has reviewed the FCC's categorizations, which affected the scope of the FCC's authority to regulate different types of services. *See, e.g.*, Nat'l Cable & Telecomm. Ass'n v. Brand X Internet Servs., 545 U.S. 967, 984–97 (2005) (finding that the *Chevron* doctrine applied and analyzing FCC regulations against that backdrop). Federal agencies, the FCC in particular, have taken somewhat active steps that gesture toward network neutrality. *See, e.g.*, Adelphia Commc'ns Corp., 21 FCC Rcd. 8203, 8299 para. 223 (2006) (indicating that the FCC would consider net neutrality concerns); Appropriate Framework for Broadband Access to the Internet Over Wireline Facilities, Policy Statement, 20 FCC Rcd. 14,986, 14,987–88 para. 4 (2005) (adopting a stance that broadband networks should be "widely deployed, open, affordable, and accessible to all consumers").

CONCLUSION

The continuing publication of stories outlining social networking site censorship suggests that there is a valid concern about this issue, especially as these sites become more integral to commercial, social, and political activities. Moreover, the unclear availability of First Amendment protections due to the threshold eligibility issues and the fairly clear market failures in the social networking site market, including information and power disparities, leave users largely vulnerable to the whims of private entities. Thus a consumer protection-based approach may be an appropriate foundation for justifying governmental intervention to protect user rights.

Additionally, my discussions regarding consumer protection are not limited to social networking sites. The factors that make social networking sites amenable to consumer protection theory arise in other online contexts where individuals also sign user agreements on expressive platforms. While this view can potentially include any Internet website, as most are governed by a user agreement and involve expression, the consumer protection rationale appears to fit best in cases where the user creates a profile and the purpose of the website is to function as a site of communication (as opposed to providing games or other services). Examples of Internet activity comparable to Facebook could include Pinterest,[281] Blogger,[282] or Tumblr.[283]

Once a consumer protection framework becomes applicable, it opens up a range of possible government interventions from those that are minimally invasive, such as requiring social networking sites to provide clear content regulation guidelines, to those that are heavily invasive, such as mandating complete content nondiscrimination. Each of the interventions discussed in Part III have advantages and drawbacks, and are targeted to respond to slightly different problems. Yet it is not possible to propose any of these interventions as an appropriate solution without further information about how social networking sites (or other Internet entities)

281. Pinterest is basically an online cork board where users can 'pin' images, videos, and other objects to their pin board. PINTEREST, www.pinterest.com (last visited Dec. 31, 2013).

282. Blogger is a Google-owned website that allows users to create, write, and share blogs on the Internet. *See* BLOGGER, http://www.blogger.com (last visited Dec. 31, 2013); *Blogger Tour*, BLOGGER, https://www.blogger.com/tour_start.g (last visited Dec. 31, 2013).

283. Tumblr is another blog-hosting platform that allows users to "effortlessly share anything." *About*, TUMBLR, http://www.tumblr.com/about (last visited Dec. 31, 2013).

engage in censorship practices. Thus the first step will be to obtain more information on these practices to obtain a better view of the situation. Then, based on that information, we can determine what regulatory approaches, if any, are necessary to protect user rights of expression.