BALANCING THE PLEADING EQUATION

Paul Stancil*

I. INTRODUCTION 92
II. SALIENT FEATURES OF U.S. PRETRIAL PROCEDURE 95
A. Judicial Indifference and the Adversary System 96
B. Liberal Discovery 100
C. The American Rule 102
1. The General Standard 102
2. Courts' Authority to Cost-Shift 103
a. Post-Resolution Cost-Shifting 103
b. Interim Cost-Shifting 106
D. Summary Judgment Standards 107
E. Summary 109
III. NOTICE PLEADING 109
A. What is Pleading? 109
B. A Brief History of U.S. Pleading 109
C. A Partial Economic Defense of Liberal Pleading 114
1. Informational Asymmetry 114
2. Agency Costs 115
3. Countervailing Costs of Notice Pleading. 116
IV. A NARRATIVE DESCRIPTION OF PRETRIAL ECONOMICS 117
A. Key Assumptions 118
B. Building the Model 120
C. General Implications of the Model 124
1. A Note About Agency 125
2. Plaintiff's Costs 126
a. Internal Costs 126
b. External Costs 128
3. Defendant's Costs 129
a. Internal Costs 129
b. External Costs 131
4. Summary and Implications 132
D. The Type I/Type II Correlation Complication 133

V. A POSSIBLE PRESCRIPTIVE APPLICATION: BALANCING THE SCALES 134
A. Theories of Procedural Justice.....134
1. Descriptive Accuracy 135
2. The Irrelevance of Normative Preferences.....136
B. Existing Pleading Literature 137
C. Economic Literature.....141
D. Prescribing Balance.....146
1. Reimagining Transsubstantivity.....146
2. Notice Pleading for Low-Risk Claims 146
3. Strict Pleading for High-Risk Claims.....147
a. Strict Pleading Lowers Three Forms of Social Cost 148
i. Inefficient Filing and Settlement: The Type I Problem 148
ii. Administrative and Other Social Costs Also Decrease.....148
4. Wrongful Dismissals: The Type II Problem 149
a. Balancing the Equation: A Bond Requirement.....150
i. Economics of the Bond Requirement 150
ii. Practical Challenges and Solutions.....152
iii. Differing Expectations.....152
iv. Difficult Calculations.....153
v. Case-by-Case or Claim Type? 154
E. The Bond Hearing 155
1. Practically.....155
a. Political Feasibility.....155
b. Implementation 156
2. Procedure.....158
a. General Contours of The Process.....158
b. Specific Considerations 160
i. How much Disparity, and Measuring External Costs 160
ii. Setting the Bond 164
3. The Heightened Standard.....164
VI. CONCLUSION.....168

I. INTRODUCTION

The ongoing debate over civil pleading standards is ultimately the product of two competing asymmetries. An informational asymmetry favoring defendants over plaintiffs drives a preference for liberal pleading standards. That is, the typical defendant often has sole possession of relevant information, and plaintiffs often cannot know critical details of their claims before discovery.[1] Thus, the risk of wrongful dismissals militates against adoption of a stricter standard. This is a "false negative" or "Type II" error risk.[2]

Asymmetry in pretrial costs favoring plaintiffs over defendants instead supports stricter pleading standards in some circumstances.[3] Several traditional characteristics of U.S. civil litigation—the adversary system, relative judicial disengagement, the "American Rule" governing attorneys fees, and prevailing summary judgment standards—combine in a small but significant subset of cases to increase defendants' expected pretrial costs relative to plaintiffs'. Plaintiffs may therefore file and defendants may sometimes settle substantively frivolous claims, simply to avoid the even greater cost of defending the claims to disposition. This risk of cost arbitrage militates against the adoption of more liberal standards. This is a false positive or Type I error risk.

---

*Assistant Professor of Law, University of Illinois College of Law. Thanks to Larry Solum, Tom Ulen, Charles Tabb, and the law faculties of the University of Illinois and the University of Indiana-Indianapolis, and participants in the 2007 Big Ten Untenured Conference for their valuable comments on previous versions of this Article. Thanks also to John Leubsdorf, Michael Solimine, and Max Huffman for their helpful comments.

[1] *See, e.g.*, Christopher Fairman, *Heightened Pleading*, 81 TEXAS L. REV. 551, 552, 561–62, 565, 625 (2002).

[2] "Type I error" and "Type II error" are terms commonly used in the social sciences literature. Formally, a Type I error is an error in which the null hypothesis is rejected when in fact it is true; Type II error occurs whenever the null hypothesis is not rejected when in fact it is false. JAMES H. STOCK & MARK W. WATSON, INTRODUCTION TO ECONOMETRICS 79 (2d ed. 2007). For purposes of this Article, the null hypothesis is that a particular U.S. civil case has substantive merit. Here a Type I error occurs when a case resolves in plaintiffs' favor despite the fact that it was objectively frivolous. A Type II error occurs when a case is dismissed but was objectively meritorious.

[3] *See, e.g.*, RICHARD A. EPSTEIN, AEI-BROOKINGS JOINT CTR. FOR REGULATORY STUDIES, MOTIONS TO DISMISS ANTITRUST CASES: SEPARATING FACT FROM FANTASY 3–5, 12, 20–21 (Mar. 2006).

Because both asymmetries often exist as to a single claim type,[4] pleading standards present an apparently insoluble problem. A stricter pleading standard might reduce the rate at which opportunistic plaintiffs file frivolous claims, but it might simultaneously reject claims that would have proved valid had they been allowed to progress through discovery. Overly liberal pleading standards present the opposite concern.

Further complicating matters, modern U.S. pleading standards are generally transsubstantive; with few exceptions,[5] U.S. courts apply the same pleading standard regardless of the type of claim or the error risks that claim presents.[6] If transsubstantivity is a nonnegotiable prerequisite,[7] any wholesale modification of the pleading standard must not reduce net social welfare as a side effect. If transsubstantivity is negotiable, any new prescriptive framework must reduce error risk without adding undue complexity and expense to an already expensive system. Given the apparently inevitable circularity of the error debate and the practical challenges presented by the transsubstantivity norm, the pleading problem appears intractable at first blush.

The economics of pleading are critical to finding a solution.[8] U.S. civil litigation is in many ways an inherently economic enterprise,[9] and pleading is a critical nexus in that enterprise—a "point of no return" beyond which the parties are bound to expend substantial sums in the development of their claims and defenses. Because little can be done to alter the parties' incentives after the pleading stage, it is important to understand the economic effects of pleading rules on potential litigants' decisions and the Type I and Type II risks they engender.

---

[4] The incidence of both risks is likely to be highly correlated. *See infra* discussion at Part V.D.

[5] *See, e.g.*, FED. R. CIV. P. 9(b) (requiring heightened pleading in claims alleging fraud or mistake); *see also* Private Securities Litigation Reform Act (PSLRA) of 1995, 15 U.S.C. § 78u-4(b)(2) (requiring that the plaintiff "state with particularity facts giving rise to a strong inference that the defendant acted with the required state of mind").

[6] *See* FED. R. CIV. P. 8(a).

[7] *But see* Christopher M. Fairman, *The Myth of Notice Pleading*, 45 ARIZ. L. REV. 987, 1028 (2003) [hereinafter Fairman, *Notice Pleading*] (arguing that the actual landscape of pleading standards is fractured).

[8] This is true even if one does not accept economic theories of procedural justice. *See infra* notes 139–152 and accompanying text.

[9] *See generally* ROBERT G. BONE, CIVIL PROCEDURE: THE ECONOMICS OF CIVIL PROCEDURE (2003).

This Article analyzes the economics of the civil litigation process with an eye toward solving the pleading problem. Part II describes critical features of U.S. civil litigation. Specifically, it focuses on (1) adversarial interaction between parties with limited government enforcement and limited judicial oversight, (2) liberal discovery standards, (3) the American Rule under which parties typically pay their own costs of suit regardless of outcome, and (4) prevailing summary judgment standards under *Celotex Corp. v. Catrett*.[10] Taken together, these characteristics can create an environment conducive to economic gamesmanship. These characteristics also suggest that adjusting the pleading standard may be the best available mechanism for curtailing that gamesmanship.

Part III briefly examines the history of U.S. pleading; concluding that notice pleading is often superior to other possible forms of case initiation. Compared to its historical predecessors, notice pleading helps solve three potentially pernicious problems: (1) informational asymmetry favoring defendants in many cases, (2) potential agency issues between plaintiffs and their attorneys, and (3) disparity between the likely quality of plaintiff's representation relative to the defendant's.

Part IV offers a simplified plain-language economic model of the pretrial process.[11] Holding other features of the U.S. civil litigation system constant, this model identifies the economic circumstances under which cost arbitrage by plaintiffs is likely. The model predicts that defendants will be indifferent between settlement and litigation to conclusion when the full economic costs of each decision are equivalent. Because defendants cannot avoid most pretrial costs once the plaintiff has successfully cleared the pleading gate, plaintiffs thus will file even objectively frivolous lawsuits when they can impose higher net pretrial costs upon defendants than defendants can impose upon plaintiffs.

But the model and the available data together suggest that many common types of litigation do not actually raise substantial risk of cost arbitrage because rough parity in pretrial costs (or disparity favoring defendants rather than plaintiffs) will limit plaintiffs' incentives to file without genuine expectation of a merits recovery in excess of their costs. The model conversely predicts that frivolous filing is likely in the subset of case types where (1) the plaintiff can impose substantial pretrial costs on

[10] 477 U.S. 317 (1986).

[11] The Appendix provides a more formal economic model corresponding to the plain-language summary presented in Part IV.

defendant, (2) the defendant cannot impose substantial reciprocal pretrial costs on plaintiff, and (3) the defendant will not incur substantial external costs through early settlement.

After briefly addressing theories of procedural justice and the state of the existing pleading literature, Part V proposes a partial solution to the asymmetry conundrum. Rejecting a rigidly transsubstantive approach,[12] this part ultimately recommends bifurcating pleading standards along cost disparity lines. For the great majority of cases, liberal notice pleading will be appropriate; there is a low risk of cost arbitrage, and thus, liberal pleading rules are justified on information asymmetry grounds. But in the small subset of claims in which a defendant may be a victim of cost arbitrage, the Article recommends a stricter default standard that would: (1) lower the risk of Type I error, (2) lower the internal costs of litigation by focusing and filtering discovery, and (3) lower the administrative costs of lawsuits by reducing both the number and likely duration of certain suits. The Article recommends mitigating any concomitant increase in Type II wrongful dismissal risk through adoption of an "opt-out" pleading procedure in which plaintiffs could obtain the right to proceed under relaxed notice pleading standards by paying a bond designed to reduce cost arbitrage incentives.

II. SALIENT FEATURES OF U.S. PRETRIAL PROCEDURE

The U.S. civil litigation system has a number of unique and important characteristics that must factor into any attempt to model pleading standards.[13] In combination, these characteristics fix U.S. civil litigation as a game in which interim economic calculations can be as important as the ultimate substantive merits of any particular claim. Taken together, these characteristics also suggest that the pleading standard may be the only genuine "moving part" subject to adjustment without a massive and thus unlikely reordering of the entire U.S. litigation system. Thus, these features of U.S. civil litigation have enormous implications for the proper design of the pleading standard.

---

[12] *But see* Fairman, *Notice Pleading*, *supra* note 7, at 1037–38 (discussing the difference between the unitary transsubstantive ideal and fractured reality).

[13] For simplicity's sake, this Article is based exclusively on the federal system; all rules-based descriptions come from the Federal Rules of Civil Procedure. With respect to the features highlighted in this Article, most U.S. jurisdictions employ functionally identical mechanisms and standards. The model is thus essentially generally applicable within most U.S. jurisdictions.

A. Judicial Indifference and the Adversary System

One of the most striking features of the U.S. civil litigation landscape is the relative lack of judicial involvement between the filing of a claim and determination of the case on the merits.[14] U.S. civil litigation depends enormously upon competition between private parties—rather than judicial micromanagement or government enforcement—to guarantee just outcomes.[15]

Compared to most other legal systems, U.S. civil enforcement relies heavily upon private attorneys general to secure compliance with the substantive law.[16] A vast number of statutes, both state and federal, explicitly contemplate private enforcement concurrent with or even in place of government enforcement; the Supreme Court has found implied private rights of action in many others.[17] Though private enforcement is an important cog in the civil justice wheel, private parties' incentives do not always align with government interests. Instead, private parties generally act in their perceived self-interest during litigation. Given that, one might anticipate a judiciary actively dedicated to reigning in private parties' litigation excesses.

This is not the case. Under the Federal Rules it is possible for a judge to have no face-to-face interaction with the parties until she is presiding over a trial or ruling on dispositive summary judgment motions.[18] The Rules permit, but do not require that judges take an active role in case

[14] Compare the U.S. system to the European inquisitional model, in which a disinterested judge actively manages most of the inquiry. *See, e.g.*, GORDON TULLOCK, TRIALS ON TRIAL: THE PURE THEORY OF LEGAL PROCEDURE 119–34 (1980).

[15] Alexia Garamfalvi, *U.S. Firms Prepare for European Class Actions*, LEGAL TIMES, June 25, 2007, http://www.law.com/jsp/llf/PubArticleLLF.jsp?id=1182762353567.

[16] There is some indication that Europe is on the cusp of a private litigation revolution. In April 2008, the European Commission's (EC) competition authority issued a white paper suggesting the need for private enforcement of the EC's antitrust laws. COMM'N OF THE EUROPEAN COMMUNITIES, WHITE PAPER ON DAMAGES ACTIONS FOR BREACH OF THE EC ANTITRUST RULES 3 (2008), *available at* http://ec.europa.eu/comm/competition/antitrust/actionsdamages/files_white_paper/whitepaper_en.pdf; *see also id.* (discussing projected increase in European class action litigation).

[17] 15 U.S.C. § 15(a) (authorizing private antitrust suits); Superintendent of Ins. of N.Y. v. Bankers Life & Cas. Co., 404 U.S. 6, 13 & n.9 (1971) (finding implied right of action for securities fraud).

[18] *See, e.g.*, FED. R. CIV. P. 16(a) (stating the court may order pretrial conferences); *id.* 16(b) (stating court scheduling orders may be issued after consultations in person, by mail, or by other means); *id.* 16(c) (stating the court may hold final pretrial conference).

management, and judges and litigants have economic and social incentives to minimize judicial participation.[19] As a result, courts tend to involve themselves only infrequently in the day-to-day administration of cases.

Courts admittedly have substantial power to control cases. Federal Rules 11, 16, 26, and 37, among others, give courts the basic authority to manage litigation. But courts do not exercise this power often. In fact, that power sometimes is at most a Sword of Damocles hanging over the litigating parties' heads, with the mere threat of judicial intervention generating incentives for the parties to avoid interaction with the court in the first place. In the typical case, judges are reluctant to inject themselves into the proceedings, and the parties are reluctant to seek judicial involvement.[20]

Because they work for a fixed salary with effective lifetime tenure, federal judges face an incentive to minimize their overall workload.[21] Elected and other limited-term judges may face some countervailing pressure, but their overall incentives are similar. Judges cannot typically control the number of cases assigned to them, of course, nor can they ultimately exercise more than limited control over the effort they must put forth at case's end.[22] But judges can and do limit their efforts in between filing and disposition. The simplest and most common means by which judges reduce their workload is by insisting that the parties resolve their own pretrial disputes, especially those relating to discovery. Judges often are able to create and sustain a credible threat of retribution toward litigants—a threat that in turn limits the number of discovery disputes over which the judge must preside.[23]

---

[19] *See, e.g.*, Richard A. Posner, *What Do Judges and Justices Maximize?(The Same Thing Everybody Else Does)*, 3 SUP. CT. ECON. REV. 1, 2, 4–7, 12, 20–22, 31 (1993).

[20] *See* Frank H. Easterbrook, *Discovery as Abuse*, 69 B.U. L. REV. 635, 636–37, 639 (1989).

[21] The dynamics of U.S. legal culture and its expectations regarding the judiciary suggest that even elected judges with relatively short terms will not be penalized substantially for attempts to limit their intra-case workloads. *See* Posner, *supra* note 19, at 4–5, 7 (noting in a slightly different context that there is little "output" on which the public can judge appellate performance).

[22] Courts may attempt to limit even end-of-case workload, most frequently by denying summary judgment motions without opinion, in the hope that the denial will drive settlement. But the prospect of having to preside over a trial and other norms and incentives may limit the extent to which courts attempt this strategy. *Id.* at 20–22.

[23] Calls for increased judicial supervision without structural changes that improve judicial incentives to do so are likely to be unsuccessful. *See* Easterbrook, *supra* note 20, at 639. For every judicial Boxer, there is likely to be at least one Old Benjamin, Mollie, or cat. *See* GEORGE ORWELL, ANIMAL FARM 46–47 (Signet Classics 1996) (1946).

Litigants themselves are understandably reluctant to invoke the court's power as to interim matters and especially as to discovery disputes. In addition to the risk of formal judicial retribution, litigants understand that disturbing the court also costs money and can lead to tit-for-tat exchanges that enrich only the lawyers. The cycle of tit-for-tat interaction thus increases both the defendant's and the plaintiff's overall expenditures. A party, therefore, will tend to seek the judge's help only when the benefits are likely to exceed the costs.[24]

Thus, simple cases often resolve after few or even no direct face-to-face interactions between the court and the litigants. And even complex litigation is subject to a norm against judicial micromanagement.[25] For cases in which the parties do seek judicial intervention, the calculus is explicitly economic and is built in the shadow of a noninterventionist norm.

Liberal discovery rules[26] and rational, inescapable judicial ignorance also can limit substantially the courts' ability to intervene between filing and trial; even when a party might want to seek judicial help, there is often little or no legal basis for intervention. Rule 26 authorizes discovery "regarding any matter, not privileged, that is relevant to the claim or defense of any party . . . ."[27] And "relevance" is itself defined liberally to include all material "reasonably calculated to lead to the discovery of admissible evidence."[28] Thus, in many circumstances, even a judge inclined to intervene on behalf of a party may have little ability to do so.

The broader and less defined the plaintiff's claim (i.e., the less specific her pleadings), the less able the judge will be to limit or focus pretrial inquiry. This is partially because the judge's hands may be tied by the discovery standard. But it is partially because judges do not, and cannot, educate themselves completely regarding the issues facing them early in litigation. Unlike the parties, the judge's knowledge of most cases is limited to the pleadings and the briefing submitted for her review. She cannot (and indeed ethically should not) educate herself further, nor is it

[24] This can occur either or both by changing the net expected value of the claim at final disposition, or less commonly, by shifting the pretrial cost expectations of the parties. But seeking judicial intervention involves risks of its own. For simplicity's value, this Article models litigation without judicial intervention.

[25] *See* Easterbrook, *supra* note 20, at 640.

[26] *See infra* notes 32–42 and accompanying text.

[27] FED. R. CIV. P. 26(b)(1). Most states employ a substantively identical discovery standard.

[28] *Id.*

rational for her to do so in light of her other workload, limited judicial resources, and the high rate at which civil cases settle.

Thus, a judge confronted by broad pleadings without obvious plausible factual anchors to which discovery can be tethered may not be able to limit cost disparity—both because of the breadth of prevailing discovery standards and because she is necessarily ignorant of the details of the dispute.[29]

In such cases, absent meaningful and obvious pleading filters, judges may be unable to limit in any significant way the breadth or depth of plaintiff's discovery until it reaches some aggregate critical mass justifying summary judgment evaluation. Thus, in addition to the independent challenges presented by liberal discovery rules,[30] the discovery standard itself can hamstring judicial efforts to minimize the social costs of litigation.

The adversary system also understandably magnifies the impact of systemic distrust between the parties. Since courts typically prefer not to interfere in disputes until the summary judgment stage or later, the parties are left with a vexing agency problem: each party is expected to act as the other's agent when it comes to searching for and producing relevant information. Thus, the party obligated to provide much of the information necessary to litigation has substantial incentives to not provide that information. Ethics rules, social constraints, and the risk of sanctions (however remote) do constrain these incentives to a limited extent. However, adversaries nonetheless face incentives (and their attorneys arguably have a legal obligation) to parse their opponents' discovery requests restrictively. It is virtually always in the requesting party's interest to counteract that incentive by casting as wide a net as possible.

Adversarial incentives create a sort of feedback loop as well. Courts are aware of the parties' incentives and are therefore understandably reluctant to circumscribe discovery, especially in certain categorical ways, because they are understandably concerned that the parties are behaving strategically in asking for specific limitations on discovery.[31] Is the defendant asking that searches of his email archives be limited because the

[29] Courts' inability to manage pretrial effectively in some cases suggests that exhortations to greater judicial effort may ultimately be unavailing. *See* Easterbrook, *supra* note 20, at 638–39.

[30] *See infra* notes 32–42 and accompanying text.

[31] *See* Easterbrook, *supra* note 20, at 641 ("Lawyers practicing in good faith, therefore, engage in extensive discovery; anything less is foolish.").

social cost-benefit calculus for the search is genuinely skewed or because he suspects that his email archives contain information harmful to his defense?

B. Liberal Discovery

According to one recent study, discovery consumes approximately 50% of all federal litigation expenditures; moreover, that study noted that discovery can account for "as much as 90% of the litigation costs in the cases where discovery is actively employed."[32] Since trial is itself an expensive proposition, discovery costs are likely an even greater percentage of typical pretrial costs. Viewed ex ante, expected discovery costs overwhelmingly dominate the parties' assessments of their pretrial costs—liberal discovery comes at a steep price.

U.S. discovery is so costly in large part because the scope of permissible inquiry is remarkably broad. A standard making discoverable "any matter, not privileged, that is relevant to the claim or defense" allows inquiry into a stunning range of topics, and the definition of "relevant"—anything that "appears reasonably calculated to lead to the discovery of admissible evidence"[33]—tends to expand rather than circumscribe the universe of discoverable material. Nor are the forms of permissible discovery inherently self-limiting. In addition to traditional pretrial simulacra of trial cross-examination—interrogatories,[34] depositions,[35] requests for admission[36] —U.S. discovery rules also explicitly allow the parties to seek discovery of "documents, electronically stored information, and tangible things."[37] In other words, U.S. civil discovery rules require litigating

[32] Memorandum from Paul V. Niemeyer, Chair, Advisory Comm. on Civil Rules, to Hon. Anthony J. Scirica, Chair, Comm. on Rules of Practice and Procedure (May 11, 1999), 192 F.R.D. 340, 357. Though the memorandum does not directly define "actively employed," it does note that in almost forty percent of federal cases, no discovery occurs, and that "in an additional substantial percentage of cases, only about three hours of discovery occurs." *Id.* Presumably the remainder of the cases involve "actively employed" discovery.

[33] FED. R. CIV. P. 26(b)(1).

[34] *See id.* 33.

[35] *See id.* 30–31.

[36] *See id.* 36.

[37] *Id.* 34.

parties to search for and produce to their opponents all nonprivileged discoverable materials from their files.[38]

In document-intensive cases, this requirement is often enormously expensive. The producing party must first search for responsive documents, frequently no easy task in itself. In typical complex litigation, the producing party might identify and then interview or search every individual and institution likely to have relevant information, erring toward over inclusion to avoid any allegation that the search has been unreasonably narrow. Next, the producing party must put the documents into some sort of reviewable form—either electronic or hard copy—often at incredible expense. Then the producing party must review all materials collected, both for "responsiveness"—whether the document is relevant and fairly responds to the opposing party's request—and for attorney-client or other forms of privilege.[39] In practice, this can mean thousands of person-hours spent reviewing thousands or even millions of documents. Finally, the producing party produces the documents for inspection and prepares a "privilege log" detailing the documents withheld under a claim of privilege.[40]

Because the failure to withhold privileged information can in some circumstances result in the waiver of a privilege, and because the producing party has an overwhelming interest in identifying damaging or helpful documents and limiting production where ethically possible, it is typically incumbent upon producing parties to review every page of every document for both privileged content, and for its potential impact on the case. Thus, a request for production under ordinary circumstances puts into motion a series of extraordinarily expensive events that the producing party can bypass only at its extreme peril.

In the aggregate, discovery costs can sometimes run to the millions of dollars or more for a single party in a single complex case.[41] Thus, in

[38] *Id.* 26(b)(1). Subject to relatively recent limitations on discovery of electronically stored information that would be too expensive to produce. *See id.* 26(b)(2).

[39] *See* Kenneth J. Withers, *Electronically Stored Information: The December 2006 Amendments to the Federal Rules of Civil Procedure*, 4 NW J. OF TECH. & INTELL. PROP. 171, 181–82 (2006).

[40] *See* FED. R. CIV. P. 26(b)(5).

[41] For a real-world primer on the costs of complex case discovery and judges' limited ability to manage those costs, see generally Oracle Corp. v. SAP AG, No. 3:07-cv-01658-PJH (N.D. Cal. July 3, 2008), *available at* http://www.tnlawsuit.com/uploads/Order%20Re%20Scope%20of%20Discovery%20of%20Electr

certain situations, concerns about discovery costs can overwhelm concerns about the merits of the underlying claim; a defendant facing $3 million in discovery costs will not be much heartened by his belief that victory is ultimately inevitable. If discovery costs cannot be limited or avoided altogether, then the promise of certain eventual success on the merits may be hollow indeed.[42]

Taken together with characteristic judicial disengagement from pretrial procedure, the liberal discovery framework is remarkable. It contemplates that the parties will develop the facts necessary to their claims or defenses with minimal influence or interference from the court. The default assumption is that the parties will conduct the entire discovery process—no matter how expensive—with limited or no judicial supervision, subject only to certain Rule-imposed limits. Thus, with potentially enormous cost, and with little realistic prospect of judicial management, liberal discovery rules can create economic incentives largely independent of the merits of any particular claim.

C. The American Rule

1. The General Standard

In the vast majority of U.S. litigated cases, each party bears its own litigation costs, including attorney's fees, regardless of who prevails. This arrangement has come to be known as the "American Rule," in contrast to the "English Rule" or "Continental Rule" under which the losing party pays.[43] Though several substantive U.S. regulatory regimes provide for

onically%20Stored%20Information.pdf (Order Regarding Scope of Discovery Of Electronically Stored Information).

[42] *See infra* note 127 and accompanying text (discussing how cost disparity can effectively deny defendants' participation rights).

[43] Virtually every other western legal system employs some form of the English Rule. Adoption of the English Rule as a solution to cost arbitrage is beyond the scope of this Article, but it should be noted that the evidence and theory on that solution are decidedly mixed. *See generally* Avery Katz, *Measuring the Demand for Litigation: Is the English Rule Really Cheaper?*, 3 J. L. ECON. & ORG. 143 (1987); *see also* Jiong Gong & R. Preston McAfee, *Pretrial Negotiation, Litigation, and Procedural Rules*, 38 ECON. INQUIRY 218, 218 (2000); *see also* Roger Bowles, *Settlement Range and Cost Allocation Rules: A Comment on Avery Katz's* Measuring the Demand for Litigation: Is the English Rule Really Cheaper?, 3 J. L. ECON. & ORG. 177 (1987). Moreover, English Rule fee shifting is rarely complete in practice. *See* TULLOCK, *supra* note 14, at 110–11.

one-sided "loser pays" fee-shifting[44] such situations are the exception rather than the rule.[45]

In American Rule contexts, parties must operate on the assumption that they will bear their own costs, regardless of the outcome. As important, under most circumstances, a party knows that its opponents will bear whatever costs the party can impose on it, regardless of the merits of the underlying claim. The implications are obvious: in most litigated cases, the parties' expected pretrial payoffs are driven in large part by their expected internal litigation costs. In the absence of "loser pays" fee-shifting, success (or failure) on the merits becomes irrelevant to the parties pretrial expectations.

2. Courts' Authority to Cost-Shift

a. Post-Resolution Cost-Shifting

In theory, at least, courts have the authority under Rule 11 to deviate from the American Rule if they determine that a plaintiff's initial pleading was frivolous.[46] But this authority is rarely exercised,[47] and defendants faced with strike suit risk cannot generally rely upon Rule 11 sanctions to make them whole.[48] Rule 11 requires plaintiffs to certify that they have a good faith factual basis for their claims[49] or that they "will likely have evidentiary support after a reasonable opportunity for further investigation or discovery."[50] If a court ultimately finds that the plaintiffs' contentions

[44] *See, e.g.*, 15 U.S.C. § 15 (2000) (authorizing fee award for successful plaintiff; Congress did not authorize a corresponding fee award for successful defendants).

[45] *See, e.g.*, FED. R. CIV. P. 54(d)(2)(ii) (requiring party seeking attorney's fees to specify "the statute, rule, or other grounds entitling the movant to the award").

[46] *See id.* 11(c).

[47] Gerald F. Hess, *Rule 11 Practice in Federal and State Court: An Empirical, Comparative Study*, 75 MARQ. L. REV. 313, 326–27 (1992) (noting that sanctions are rarely sought or imposed in state courts).

[48] *See, e.g.*, Cary Coglianese, *Insuring Rule 11 Sanctions*, 88 MICH. L. REV. 344, 344 (1989) (noting estimates of 600 to 1000 nationwide Rule 11 decisions in the past six years, or between 100 and 167 annually, with awards "as high as $200,000 to $300,000"). While arguably impressive in the abstract, these figures (and especially the relatively low amounts for high-end monetary sanctions) are unlikely to prove a sufficient deterrent to plaintiffs seeking cost arbitrage in complex cases.

[49] *See* FED. R. CIV. P. 11(b)(1).

[50] *Id.* 11(b)(3).

were unreasonable, it "may impose an appropriate sanction on any attorney, law firm, or party that violated the rule."[51] Read broadly, these provisions would seem to provide courts with the ability to correct opportunistic frivolous pleading; though the sanction is to be "limited to what suffices to deter repetition of the conduct," it can include "all of the reasonable attorney's fees and other expenses directly resulting from the violation."[52] Rule 11 even seems to contemplate systemic deterrence, authorizing sanctions sufficient to deter "comparable conduct by others similarly situated."[53]

But there are several problems with relying on Rule 11. First and foremost, it just doesn't work. Reported cases involving Rule 11 sanctions of any type are surprisingly rare and reported cases in which the sanction for frivolous pleading includes the defendant's discovery costs are virtually nonexistent. There are several possible explanations for this phenomenon. At one end of the spectrum, it may be that there are no frivolous filings, and thus there is no need for such sanctions. Perhaps, but as discussed below, there exists in certain limited circumstances a strong economic incentive for cost arbitrage—an incentive arising out of the defendant's corresponding incentive to settle rather than litigate.[54] It is possible that attorney's fees sanctions are rare because misbehavior is rare. But the absence of reported cases just as likely reflects defendants' collective discounting of any "post-dismissal sanctions payoff." The more steeply the payoff is discounted, the more likely settlement becomes.

Defendants may discount their expected back-end sanctions payoffs for at least three reasons. First, hindsight is 20/20. That is, defendants may rationally discount expected sanctions payouts because they perceive a low probability that a court will authorize sanctions in an amount sufficient to offset pretrial costs. Assuming a pleading satisfies the pleading standard (begging a question we will take up in Part V), it may be difficult for a court to look backward to the plaintiff's filing with the clarity necessary to impose meaningful sanctions that will actually deter future misbehavior by others similarly situated. Somewhat ironically, hindsight may get blurrier as pretrial costs increase because, having acquiesced to expensive injury on

[51] *Id.* 11(c)(1).

[52] *Id.* 11(c)(4).

[53] *Id.*

[54] *See* Tobias Holdings, Inc. v. Bank United Corp., 177 F. Supp. 2d 162, 166 (S.D.N.Y. 2001).

the front end, courts will be understandably reluctant to impose economically meaningful sanctions.

Defendants may also discount their expected back end sanctions payoffs because they do not believe the plaintiff/or her attorney will have the ability to pay. Plaintiffs generally have fewer resources than defendants, and even a well-financed plaintiff's attorney may not be able to make defendant whole. In the Supreme Court's recent *Twombly* opinion,[55] for example, the defendants' estimated discovery costs were in the tens of millions of dollars while plaintiffs' costs were likely an order of magnitude (or more) lower. Few plaintiffs' attorneys could satisfy a $10 million post-dismissal sanction.

Finally, there is some prospect of a "deterrence disconnect"—that is, the court may decide that a given sanction is sufficient "to defer repetition of the conduct or comparable conduct by others similarly situated" without in fact making defendant whole.[56] In economic terms, this reasoning is inherently fallacious: deterrence is economically insufficient until the expected value of the defendants' sanctions payout fully offsets cost disproportionately, which means it must also account for the risk of nonpayment.[57] But it would nonetheless be tempting for courts to undersanction.[58]

In any event, to the extent the defendant discounts the likelihood of being made whole through imposition of postjudgment sanctions, those sanctions no longer provide a meaningful disincentive to settlement. The economically rational defendant who expects, for example, even a nominally large $500,000 fee sanction in "compensation" for a $3 million disparity in pretrial costs may not pursue the matter to disposition,[59] although the settlement amount will decrease by the expected value of the

[55] *See* Bell Atl. Corp. v. Twombly, 127 S. Ct. 1955, 1967 n.6 (2007).

[56] FED. R. CIV. P. 11(c)(4).

[57] Thus, the solution offered in Part V relies in part on precommitment of the optimal sanction amount as precondition to pursuing high risk claims under a notice pleading standard.

[58] An economically effective sanction would have to fully account for the disparity in pretrial costs, discounted for likelihood of imposition and likelihood of collection. If a $1,000,000 payment would be necessary to render the defendant indifferent between settlement and litigation of a frivolous claim, but the defendant regards the likelihood of imposition at twenty-five percent and the likelihood of payment at twenty-five percent, the expected sanction amount would need to be much higher than $1,000,000.00 to achieve indifference.

[59] See *infra* Part IV.

sanctions amount.[60] And the rational plaintiff will not be deterred by that nominally large expected sanction, because she knows the defendant will never get that far. Moreover, even taking Rule 11 at face value, a plaintiff ostensibly could file any claim consistent with Rule 8(a)(2), conduct discovery, and then withdraw the claim before a sanctions motion was filed.[61] A court is unlikely to entertain a sanctions motion before adequate time for discovery has passed, so the risk of Rule 11 sanctions would not seem to offer much deterrence value.

b. Interim Cost-Shifting

For similar reasons, the courts' ability to shift costs while the case is ongoing may be more apparent than real. Courts theoretically have the power to shift pretrial costs on an ongoing basis under a variety of Rules provisions.[62] But the deterrent value of intra-case cost-shifting again depends upon the defendant's ex ante assessment of his likely payoffs. Thus, the same sorts of concerns—judicial myopia, plaintiff insolvency, suboptimal judicial economic analysis—may lead the defendant to discount the value of that solution. More important, the court's apparent authority to cost shift is triggered only in the event of either specific circumstances involving electronically stored information or if a party violates a pretrial order.[63] The Federal Rules do not appear to contemplate or to authorize intra-case cost shifting as a matter of general equity.

The prospect of intra-case fee-shifting may also create incentives for additional strategic behavior on the part of litigants. If the responding party regards cost shifting as likely ex ante, it may face an incentive to maximize its own costs, especially if there are significant differences in the marginal

[60] Here and elsewhere, the term "expected value" refers to the probability of a given event multiplied by the magnitude of that event. For events denominated in dollars, the expected value of the event is the expected payoff multiplied by the probability of that payoff. *See, e.g.*, DAVID BESANKO & RONALD R. BRAEUTIGAM, MICROECONOMICS 571–72 (3d ed. 2007). Thus, the expected value of a million-dollar lottery that one has a ten percent chance of winning is $100,000. If a defendant believes that the Rule 11 sanction for plaintiff's spurious filing will be $100,000 and that he has a twenty-five percent chance of obtaining that sanction from the court after dismissal of the claim, the expected value of the sanction is $25,000, and defendant may rationally discount his strike settlement value by that amount.

[61] Rule 11's structure also militates against its potential effectiveness. Rule 11(c)(2) allows parties time to withdraw frivolous claims before a motion for sanctions can be filed.

[62] *See, e.g.*, FED. R. CIV. P. 16(f); *id.* 26(b)(4)(C); *id.* 37 (a)(5)(A).

[63] *See id.* 26(b)(4)(C); *id.* 16(f)(2).

value of money to the two parties. Similarly, there is no obvious and coherent way to shift the bulk of privilege and responsiveness review costs from one party to another—though it may be possible to cost shift as to electronic searching or database creation, it will be functionally impossible for courts to cost shift as to the producing party's own review. Between gaming concerns and endless arguments over what is a "legitimate" review cost, courts likely would be unwilling to allow a defendant to finance a "Cadillac" document review partially at the plaintiff's expense, and would be ill-suited to determine what a "Chevy" review would look like.[64] Even in English Rule jurisdictions, fee shifting is rarely complete, and there is little reason to believe it could work in the United States.[65]

The American Rule reality—that parties generally must anticipate bearing their own litigation costs for most claims—dramatically affects the parties' pretrial analyses of their claims and defenses. The primary effect of the American Rule is to elevate cost parity to an importance it would not enjoy under the English Rule or under a mandatory intra-case fee-shifting regime. Because the parties will pay their own way regardless of the outcome they face incentives and have some ability to maximize their opponents' costs and minimize their own.

D. Summary Judgment Standards

The final relevant component of U.S. pretrial procedure is the prevailing summary judgment standard. Under *Celotex Corp. v. Catrett*, judges may grant summary judgment to the moving party when the non-moving party has failed "after adequate time for discovery" to produce summary judgment evidence in support of an essential element of the non-moving party's claim or defense.[66] Consider a negligence case requiring proof of negligence, causation, and damages. If after adequate time for discovery a plaintiff has produced no evidence that the defendant's negligent act caused plaintiff's injuries, summary judgment in favor of the defendant would be appropriate even if the plaintiff had produced evidence of both defendant's negligence and of the plaintiff's injuries themselves.

---

[64] *See supra* notes 14–44 and accompanying text.

[65] *See* TULLOCK, *supra* note 14, at 110–11.

[66] 477 U.S. 317, 322 (1986). There are literally thousands of lower court cases fleshing out the full contours of post-*Celotex* summary judgment procedure, but the root of modern no-evidence summary judgment procedure is *Celotex* itself.

Celotex reaffirms the uniquely American commitment to substantial party-driven pretrial discovery. By so doing, it leaves the door open to the possibility of opportunistic pleading by plaintiffs.

Most of the economic impact of the *Celotex* standard lies in the phrase "after adequate time for discovery." A defendant faced with even the most ridiculous imaginable substantive claim—assuming it is properly pleaded—cannot typically move for summary judgment until there has been "adequate time for discovery." So, if for example the defendant knows with certainty that he is not liable for plaintiff's injuries because he was never involved in an accident with plaintiff, he cannot generally move for summary judgment or dismissal until after the plaintiff has had adequate time to conduct discovery. And "adequate time" usually means "adequate discovery," which in turn means that defendants cannot escape even completely frivolous suits until after they have expended substantial sums of money responding to plaintiff's discovery requests.

In the hypothetical introduced above, of course, cost imposition would likely be minimal—a defendant denying the fact of an accident pled with typical specificity may well be able to convince a court to limit discovery to that fact, and the plaintiff's costs of bringing such a claim would likely equal or exceed defendant's.[67] In addition, in cases of that sort, the prospect of sanctions may well prove a legitimate deterrent.

But consider instead a complex claim alleging the existence of a years-long conspiracy that defendant ultimately can prove never occurred. Under *Celotex*, the plaintiff has a right to substantial discovery; she is in effect entitled to search the haystack for needles. Assuming plaintiff's complaint passes the pleading bar, the defendant's next opportunity to dispose of the case will come only after "adequate time for discovery" has passed as to an essential element of the plaintiff's claim. Depending upon the nature of the claim and the nature of the pleading standard, "adequate time for discovery" may imply thousands or even millions of dollars of unavoidable expense for the defendant, regardless of the merits of plaintiff's underlying claim. These expenses will necessarily increase in the absence of sequential filters that provide a court with the means to manage discovery under Rule 16.[68]

[67] *See infra* notes 123–25 and accompanying text.

[68] *See, e.g.*, FED. R. CIV. P. 16(b)(3)(B)(ii).

E. Summary

U.S. civil litigation is a world in which certain litigants may find it tempting to engage in strategic behavior unrelated to the merits of their claims or defenses. The front-loading of litigation expenses and the laissez faire approach of the judiciary create particular risk. Because the pleading standard serves as the gatekeeper to the heart of the U.S. pretrial system, it is worth exploring the current approach to pleading under federal law.

III. NOTICE PLEADING

A. What is Pleading?

For the purposes of this Article "pleading" refers exclusively to the initial written filing submitted to a court by plaintiff to initiate a civil case. As discussed above the salient features of post-pleading U.S. pretrial procedure tend to make the pleading stage critically important; once the plaintiff's complaint passes muster under Rule 8, the train has effectively left the station. The "pleading standard" therefore is the set of threshold requirements a plaintiff's case-initiating written filing must satisfy to obtain access to discovery and the other case development tools and procedures characteristic of U.S. litigation.

B. A Brief History of U.S. Pleading

There have been three major periods in the history of U.S. civil pleading. The common law writ era, imported from England, lasted from the Founding through roughly the middle of the nineteenth century. During this period, pleading followed traditional English common law practice: parties shoehorned legal claims into one of a series of preexisting writs and responses, sequentially narrowing the issues subject to trial.[69] Common law writ pleading was a highly specialized and technical exercise, and selection of the wrong writ at the wrong time could doom an objectively meritorious claim to dismissal without a hearing on the merits.[70]

---

[69] *See* William H. Lloyd, *Pleading*, 71 U. PA. L. REV. 26, 26–36 (1922) (providing summary of common law pleading practice).

[70] *See id.* at 33 (By the early nineteenth century, "the complaints against the prolixity, obscurity and triviality of common law pleading became louder; the demand for reform more and more insistent").

Beginning in the late 1840s with New York's adoption of the Field Code, U.S. procedure began a gradual move away from common law pleading toward "code pleading" or "fact pleading."[71] Designed to be simpler and more just than common law pleading was believed to be, fact pleading required that the plaintiff plead facts in support of each of the elements of her claim. That is, fact pleading required the plaintiff to plead facts that tended to make each element of her claim plausible.[72] By the 1920s, over half of the states had adopted some form of code pleading.[73]

The modern era of pleading procedure begins with the adoption of the Federal Rules of Civil Procedure in 1938. Responding to a belief that common law pleading and code pleading were both fatally flawed,[74] the Federal Rules explicitly adopted a radical new pleading standard. Specifically, Rule 8(a) required only that plaintiff provide "a short and plain statement of the claim showing that the pleader is entitled to relief."[75] The new standard purported to eliminate arcane factual hierarchies and other complexities that had come to dominate predecessor pleading forms,[76] replacing them with a less formal standard less susceptible to confusion or gamesmanship.

But the "short and plain statement" requirement of Rule 8(a) is itself inherently ambiguous,[77] and it remained a source of confusion to attorneys

---

[71] *See id.* at 34.

[72] There was always substantial debate regarding the precise factual pleading requirements under the various codes. *See* Charles E. Clark, *The Complaint in Code Pleading*, 35 YALE L.J. 259, 260–68 (1926) [hereinafter Clark, *Code Pleading*] (arguing that distinctions between law, facts, and evidence were often meaningless; describing then-dominant fact pleading standard as requiring pleading of "material" facts).

[73] *See* Lloyd, *supra* note 69, at 35.

[74] *See, e.g.*, David M. Roberts, *Fact Pleading, Notice Pleading and Standing*, 65 CORNELL L. REV. 390, 391–92 (1979) (collecting criticisms and discussing history of notice pleading and its predecessors); *see also* Charles E. Clark & James W. Moore, *A New Federal Civil Procedure II: Pleadings and Parties*, 44 YALE L.J. 1291, 1299 (1935).

[75] FED. R. CIV. P. 8(a)(2) (1938) (original version). In December 2007, the text of the rule was revised slightly to clarify pleading requirements as to damages and jurisdiction; the 1938 language governing statement of claims was left undisturbed. *See* FED. R. CIV. P. 8(a)(2) (as amended in 2007). For purposes of this Article, the term "plaintiff" is interchangeable with the term "pleader" in Rule 8(a)(2) because the model developed in this Article focuses only upon a unidirectional dispute between a unitary plaintiff interest and a unitary defendant interest.

[76] *See* Lloyd, *supra* note 69, at 33.

[77] *See* Roberts, *supra* note 74, at 419–20 (describing the "semantic slipperiness" of Rule 8(a)(2) as "almost as fuzzy as the older code standard").

and courts alike both before and after the Supreme Court's landmark 1957 decision in *Conley v. Gibson*.[78] In that case, the Fifth Circuit had dismissed a civil rights claim brought on behalf of African-American union members who claimed their union was discriminating against them in favor of white union members. Though the courts below had found against the plaintiffs on different grounds, the Supreme Court nonetheless addressed the sufficiency of plaintiffs' pleading, holding that a claim need only provide the defendant "fair notice of what the plaintiff's claim is and the grounds upon which it rests."[79] *Conley* also permanently introduced the term "notice pleading" into the judicial lexicon.[80] In so doing, it is clear that the Court was primarily interested in contrasting the simplified and liberalized pleading standards under the Federal Rules with their arcane, demanding, and complex predecessors: "The Federal Rules reject the approach that pleading is a game of skill in which one misstep by counsel may be decisive to the outcome . . . ."[81] The Court concluded by noting that "the purpose of pleading is to facilitate a proper decision on the merits."[82]

If *Conley* had gone no further, perhaps courts ultimately would have coalesced around a unitary notice pleading standard. But *Conley* instead offered a peculiar amplification of the notice pleading standard, stating that a plaintiff's complaint "should not be dismissed for failure to state a claim unless it appears beyond doubt that the plaintiff can prove *no set of facts* in support of his claim which would entitle him to relief."[83] Taken to its literal extreme, *Conley* thus seems to say that the mere pleading of a viable theory of recovery is sufficient to state a claim, so long as there is some possible set of facts that could be proved in support of that claim.

Conley quickly became the dominant case interpreting modern pleading doctrine. And though it was cited extensively for its general approach to notice pleading, tens of thousands of briefs and lower court opinions also

[78] 355 U.S. 41, 45–46 (1957).

[79] *Id.* at 47.

[80] *Id.* Judge and Professor Charles Clark, the architect of the Federal Rules, had been arguing explicitly for "notice pleading" since at least the mid-1920s. *See* Clark, *Code Pleading*, *supra* note 72, at 265; Charles E. Clark, *History, Systems and Functions of Pleading*, 11 VA. L. REV. 517 (1925) [hereinafter Clark, *Functions of Pleading*].

[81] *Conley*, 355 U.S. at 48.

[82] *Id*. The irony, of course, is that in the rare but real cost disparity conditions outlined in the model, liberal pleading standards may actually *impede* a proper decision on the merits by removing merits considerations from the equation entirely.

[83] *See id.* at 45–46 (emphasis added).

expressly cite the "no set of facts" proposition.[84] By the turn of the twentieth century *Conley* had become a cornerstone of civil procedure casebooks; before 2007, *Conley* had evolved into procedural holy writ or something very like it.

As the Supreme Court continued to grapple with application of Rule 8(a)(2),[85] lower courts eventually diverged substantially in their understanding of the Rule and of the import and meaning of *Conley's* "no set of facts" language. By the end of the twentieth century, some courts were flatly rejecting a literal and liberal interpretation of *Conley*[86] while others read *Conley* almost to the outer limits of its "no set of facts" dictum.[87]

In 2007, the Supreme Court once again took up the issue of pleading standards, this time in the context of an antitrust claim brought against incumbent local telephone companies. In *Bell Atlantic Corp. v. Twombly*,[88] the Supreme Court seemingly endorsed a transsubstantive "plausibility" pleading standard, holding that the "no set of facts" language in *Conley* had "earned its retirement."[89] Instead, the Supreme Court held that Rule 8 requires factual allegations sufficient to "raise a right to relief above the speculative level."[90] The Court reaffirmed its earlier holding that courts need not "accept as true a legal conclusion couched as a factual allegation."[91] In the antitrust context, the Supreme Court held that a

[84] A Westlaw KeyCite search conducted on January 23, 2008, revealed over 87,000 positive citations of *Conley* in the database, the vast majority of which appear to relate to its "no set of facts" dictum regarding pleading standards.

[85] *See, e.g.*, Dura Pharm., Inc. v. Broudo, 544 U.S. 336, 346–48 (2005); Swierkiewicz v. Sorema N.A., 534 U.S. 506, 513–14 (2002); Leatherman v. Tarrant County Narcotics Intelligence & Coordination Unit, 507 U.S. 163, 168 (1993); Neitzke v. Williams, 490 U.S. 319, 326–27 (1989); Scheuer v. Rhodes, 416 U.S. 232, 236–37 (1974).

[86] *See, e.g.*, Car Carriers, Inc. v. Ford Motor Co., 745 F.2d 1101, 1106 (7th Cir. 1984) (rejecting literal interpretation of *Conley*, requiring instead "direct or inferential allegations respecting all the material elements necessary to sustain recovery under *some* viable legal theory" (citing Sutliff, Inc. v. Donovan Cos., 727 F.2d 648, 654 (7th Cir. 1984))).

[87] In *Bell Atlantic Corp. v. Twombly*, the Supreme Court characterized this extreme interpretation as holding that "any statement revealing the theory of the claim will suffice unless its factual impossibility may be shown from the face of the pleadings." 127 S. Ct. 1955, 1968 (2007).

[88] 127 S. Ct. 1955 (2007).

[89] *Id.* at 1969.

[90] *Id.* at 1959.

[91] *Id.* at 1965 (quoting Papasan v. Allain, 478 U.S. 265, 286 (1986)).

complaint must provide "enough factual matter (taken as true) to suggest that an agreement was made."[92]

In defense of *Conley's* "retirement," Justice Souter flirted with but did not fully endorse an economic theory of pleading. His majority opinion acknowledged "*in terrorem*" settlement risks,[93] and acknowledged that "proceeding to antitrust discovery can be expensive."[94] In support of its conclusion, the Court cites a 1999 memorandum to the federal judiciary's Chair of the Committee on Rules of Practice and Procedure finding that discovery "accounts for as much as 90 percent of litigation costs when discovery is actively employed."[95] Souter completes his analysis by noting that effective judicial supervision may be hard to come by, and that neither summary judgment scrutiny nor "lucid instructions to juries" come in time to mitigate the risks.[96]

Almost immediately after it was handed down, lower courts and scholars began struggling to identify the true reach and meaning of the *Twombly* opinion.[97] Perhaps the most plausible interpretation of the case is that it substitutes a somewhat relaxed and informal version of fact pleading for certain antitrust claims at the very least and perhaps for a somewhat broader swath of civil litigation.[98] In any event, the furor regarding

[92] *Id.*

[93] *Id.* at 1959 (quoting Dura Pharm., Inc. v. Broudo, 544 U.S. 336, 347 (2005)).

[94] *Id.*

[95] *Id.* at 1967 (citing Niemeyer, *supra* note 32, at 357).

[96] *Id.*

[97] A November 14, 2008 Westlaw Keycite search yields 24,495 cases citing *Twombly*. *See also*, Scott Dodson, *Pleading Standards After* Bell Atlantic Corp. v. Twombly, 93 VA. L. REV. *IN BRIEF* 121, 127–28 (2007); Allen Ides, Bell Atlantic *and the Principle of Substantive Sufficiency Under Federal Rule of Civil Procedure 8(a)(2): Toward a Structured Approach to Federal Pleading Practice*, 243 F.R.D. 604, 604–605 (2007); Suja A. Thomas, *Why the Motion To Dismiss Is Now Unconstitutional*, 92 MINN. L. REV. 1851, 1853, 1860–63, 1867–71, 1878–80 (2008). *See generally* A. Benjamin Spencer, *Plausibility Pleading*, 49 B.C. L. REV. 431 (2008); Ettie Ward, *The After-Shocks of* Twombly*: Will We "Notice" Pleading Changes*, 82. ST. JOHN'S L. REV. 893 (2008).

[98] Just two weeks after *Twombly*, the Supreme Court issued a per curiam reversal in *Erickson v. Pardus*, 127 S. Ct. 2197 (2007), where a pro se prisoner lawsuit had been dismissed by the Tenth Circuit on pleading grounds. The Court held that "[s]pecific facts are not necessary; the statement need only give the defendant fair notice" of the claim and grounds for relief." *Erickson*, 127 S. Ct. at 2200 (quoting Conley v. Gibson, 355 U.S. 41, 47 (1957)). To the extent *Twombly* purported to undo part of *Conley*, the per curiam reversal in *Erickson* clarified that the Court was retiring only the "no set of facts" language, which had been, in the Court's view, extended well beyond its intended meaning over time. Taken together, these two cases apparently indicate a

Twombly continues to rage, and it is likely to be many years before courts regain their equilibrium.[99]

C. A Partial Economic Defense of Liberal Pleading

In light of this short history, a threshold economic question arises: Does notice pleading in general make economic sense? Though the economic costs and benefits of various pleading regimes are beyond precise quantification, notice pleading is a priori defensible on economic grounds. This defense has been articulated elsewhere in less explicitly economic terms,[100] but it consists of two essentially economic arguments: (1) Notice pleading helps address potential informational asymmetry that otherwise could affect the function of the dispute resolution market, and (2) notice pleading helps mitigate potential agency costs associated with recourse to the dispute resolution system.

But the benefits are only part of the story; there are also economic costs associated with notice pleading. A proper pleading standard must also account for those costs.

1. Informational Asymmetry

Defendants often have sole custody of relevant information critical to the plaintiff's claim. If the operative pleading standard required plaintiff to allege facts that she cannot reasonably be expected to know at the case's inception, this informational asymmetry would in turn prevent proper functioning of the litigation market. The plaintiff's failure to plead unknowable facts could in some cases result in dismissal of a claim that should have been successful, and net social welfare decreases as the defendant unjustly retains wealth that should have compensated plaintiff for her injury.

---

limited retreat from the most extreme lower court extensions of *Conley*, not a wholesale rejection of notice pleading.

[99] *See, e.g.*, Michael C. Dorf, *The Supreme Court Wreaks Havoc in the Lower Federal Courts—Again*, FINDLAW'S WRIT, Aug. 13, 2007, http://writ.news.findlaw.com/dorf/20070813.html (describing confusion created by *Twombly* and noting that "the hundreds of lower court opinions citing *Twombly* take a variety of positions on the meaning of the case").

[100] *See, e.g.*, Clark, *Functions of Pleading*, *supra* note 80, at 543–44 (arguing in favor of broad adoption of notice pleading approach); *see also* Fairman, *Heightened Pleading*, *supra* note 1, at 556.

Though there are certainly claim categories for which this informational asymmetry is less of a problem, defendants in many circumstances do know more than plaintiffs about the facts relevant to the claim. It is the defendant, after all, who is accused of wrongdoing. It may be both highly relevant and (from the plaintiff's perspective) unknowable ex ante whether the defendant had been drinking just before an accident or whether the defendant intentionally interfered with the plaintiff's business opportunities by calling the plaintiff's customers and providing them with false information. Moreover, the defendant has little independent incentive to fill in the blanks. To address this problem, notice pleading gives the plaintiff the benefit of the doubt early in the case.

2. Agency Costs

Notice pleading also minimizes agency costs associated with dispute resolution. Judge Clark recognized this justification when he defended liberal pleading standards as protecting the plaintiff from "deciding at his peril on the correct legal theory applicable to his case."[101]

Agency costs are inherent in litigation, with or without the formal participation of attorneys. Each dispute that finds its way into the court system is immediately packaged with the complex and sometimes Byzantine trappings of legal procedure, and thus even (or perhaps especially) the litigant who insists upon representing himself finds himself in a metaphorical agency relationship with the underlying legal claim. That is, the plaintiff-as-navigator of the legal system serves a metaphorical agent for the plaintiff-as-injured-party for whom redress is sought.

U.S. civil courts do not require parties to hire attorneys, nor do they provide attorneys as a matter of right. Accordingly, one can defend notice pleading standards as protective of the pro se litigant's imperfections as agent for his claims or defenses. In practice, of course, pro se litigants are given additional leeway in pleading and other matters *vis-à-vis* represented parties.[102] But similar agency issues arise even when the plaintiff is represented by counsel.

---

[101] *See* Clark & Moore, *supra* note 74, at 1301.

[102] *See, e.g.*, Young v. Corbin, 889 F. Supp. 582, 586 (N.D.N.Y. 1995) (citing Harris v. Heinrich, 919 F.2d 1515, 1516 (11th Cir. 1990); Kurkowski v. Volcker, 819 F.2d 201, 204 (8th Cir. 1987)); Brown v. Consol. Freightway, 152 F.R.D. 656, 660 (N.D. Ga. 1993); Harmon v. O'Keefe, 149 F.R.D. 114, 116 (E.D. Va. 1993).

The broader economics of litigation suggest that defendants will routinely enjoy agency advantages over plaintiffs, even when both sides are represented by counsel. Plaintiffs as a class are not interested in moral victories—they rationally file suit when they believe that the defendant has the ability to compensate them for their injuries. But plaintiff's own ability to finance litigation at an equivalent level of skill and experience is not always or even often relevant to her decision to file a claim. If plaintiffs have fewer resources than defendants (in the aggregate, this will be the case almost by definition, because judgment-proof defendants will not face suit and plaintiff's resources are often irrelevant), they may be forced to hire lower-quality agents to pursue their claims. The notice pleading standard thus also arguably stands as a partial bulwark against systemic differences in the quality of advocacy available to plaintiffs and defendants.

3. Countervailing Costs of Notice Pleading.

Despite its potential societal benefits, notice pleading is not costless. U.S. civil procedure is often expensive business, and the economic justifications for notice pleading cannot always justify the economic costs. In particular, the lower the pleading standard, the greater the potential disparity between defendant's and plaintiff's costs for several claim types. This is because the range of permissible inquiry into defendant's affairs increases as pleading specificity requirements decrease, especially for claims in which the plaintiff's own conduct is of little moment. Fishing expeditions are sometimes so expensive that the defendant will pay the plaintiff to leave even a lake the defendant knows to be empty. The following section explores the economic costs and benefits of pleading.

IV. A NARRATIVE DESCRIPTION OF PRETRIAL ECONOMICS[103]

U.S. pretrial proceedings are facially complex. Every lawsuit consists of a myriad of actual and potential interactions between parties and courts, and no two litigation stories are ever exactly the same. But the apparent complexity of the pretrial process is somewhat misleading, at least as it relates to the interaction between economic incentives and pleading standards. This Article derives a powerful, game theoretic model of pretrial litigation incentives from several key variables and a few conservative assumptions. The model is not infallibly predictive. But it nonetheless provides a worthwhile starting point for gauging both the incentives facing civil litigants and the ways in which pleading standards affect those incentives.

A more technical summary of the model appears in the attached Appendix. This section of the Article focuses instead upon a plain language explanation of the model and its implications.

---

[103] Numerous scholars have modeled various aspects of the pretrial litigation process, including models attempting to address the frivolous suit problem. None of these articles attempt to address the role of pleading standards. For a representative sample, *see generally* Lucian Arye Bebchuk, *Suing Solely to Extract a Settlement Offer*, 17 J. LEGAL STUD. 437 (1988); Lucian Arye Bebchuk, *A New Theory Concerning the Credibility and Success of Threats to Sue*, 25 J. LEGAL STUD. 1 (1996); Robert D. Cooter & Daniel L. Rubinfeld, *Economic Analysis of Legal Disputes and Their Resolution*, 27 J. ECON. LIT. 1067 (1989); Andrew F. Daugherty & Jennifer F. Reinganum, *Hush Money*, 30 RAND J. ECON. 661 (1999); Philip L. Hersch, *Indemnity, Settlement, and Litigation: Comment and Extension*, 19 J. LEGAL STUD. 235 (1990); Keith N. Hylton, *Asymmetric Information and the Selection of Disputes for Litigation*, 22 J. LEGAL STUD. 187 (1993); Avery Katz, *The Effect of Frivolous Lawsuits on the Settlement of Litigation*, 10 INT'L REV. L. & ECON. 3 (1990); Randy J. Kozel & David Rosenberg, *Solving the Nuisance-Value Settlement Problem: Mandatory Summary Judgment*, 90 VA. L. REV. 1849 (2004); Geoffrey P. Miller, *An Economic Analysis of Rule 68*, 15 J. LEGAL STUD. 93 (1986); Barry Nalebuff, *Credible Pretrial Negotiation*, 18 RAND J. ECON. 198 (1987); I.P.L. P'ng, *Strategic Behavior in Suit, Settlement, and Trial*, 14 BELL J. ECON. 539 (1983); A. Mitchell Polinsky & Daniel L. Rubinfeld, *Sanctioning Frivolous Suits: An Economic Analysis*, 82 GEO. L.J. 397 (1993–1994); George L. Priest, *Regulating the Content and Volume of Litigation: An Economic Analysis*, 1 SUP. CT. ECON. REV. 163 (1982); Jennifer F. Reinganum & Louis L. Wilde, *Settlement, Litigation, and the Allocation of Litigation Costs*, 17 RAND J. ECON. 557 (1986); D. Rosenberg & S. Shavell, *A Model In Which Suits Are Brought For Their Nuisance Value*, 5 INT'L REV. L. & ECON. 3 (1985); Steven Shavell, *Suit, Settlement, and Trial: A Theoretical Analysis Under Alternative Methods for the Allocation of Legal Costs*, 11 J. LEGAL STUD. 55 (1982); William H. Wagener, *Modeling the Effect of One-Way Fee Shifting on Discovery Abuse In Private Antitrust Litigation*, 78 N.Y.U. L. REV. 1887 (1987).

A. Key Assumptions

Like most economic analyses of the litigation process, this model makes several important simplifying assumptions. First and foremost, the model assumes one rational, utility-maximizing litigating interest on each side of a single dispute.[104] The model further assumes that each litigant attempts to maximize its utility by maximizing its expected economic returns. This is not to deny the existence of irrational litigants or of litigants who litigate on principle rather than to maximize their economic well-being. But the great majority of civil litigants desire only to litigate as efficiently as possible; in the aggregate, economic rationality assumptions accurately describe most of the actors participating in the civil litigation system.[105] To the extent that the model is designed to yield prescriptive insights as to the optimal design of the civil litigation system as a whole, economic rationality is an appropriate assumption.

Second, the model treats each discrete litigated case as a separate game between plaintiff and defendant. Thus, to the extent either litigating party is concerned about consequences external to the modeled case, these concerns are expressed as fully realized economic terms within the initial game, rather than as additional iterations. For example, the concern that early settlement of an objectively frivolous claim will yield additional costs in the form of follow-on lawsuits is reflected in the model through a term aggregating all of the expected external costs of settlement rather than by modeling the defendant's decisions as a series of sequential games. Similarly, if the plaintiff is concerned that filing and subsequently losing a frivolous suit will affect her standing with a court in which she is likely to be a repeat player, that expected cost is expressed as an external cost term attendant with losing the case at summary judgment.

The model also assumes that the parties move sequentially within the game, each responding to the opposing party's previous move. Thus, the plaintiff's decision to file or not file the lawsuit is followed by the defendant's decision to fight or settle, etc. The model further assumes

[104] The model thus captures class actions and mass actions where the plaintiffs share rough unity of interest. It does not address scenarios in which parties sitting on the same side of the courtroom nonetheless have potentially divergent interests.

[105] *See* NEIL K. KOMESAR, IMPERFECT ALTERNATIVES: CHOOSING INSTITUTIONS IN LAW, ECONOMICS, AND PUBLIC POLICY 123–137 (1994).

"perfect information"; that is, that each party knows how his opponent moved at each previous node in the game.[106]

Similarly, for computational convenience only, the model assumes that each party shares the same ex ante perceptions regarding both the merit and the magnitude of the claim; that is, both plaintiff and defendant agree on the probability of plaintiff's success at trial and on the size of the award in the event of success.[107] For the same reason, the model assumes that the parties also can predict their own and their opponent's litigation costs accurately,[108] and that pretrial activities will not affect either the expected value of the claim at disposition or on opponents' litigation costs.

None of these assumptions substantially diminishes the real-world predictive value of the model. Litigants tend to treat each actual or potential suit as a single event, and even cases involving potential externalities are often more easily modeled as discrete events with external implications, rather than in series. Moreover, litigants typically move sequentially in litigation and are aware of their adversaries' moves. And in broad terms, at least, parties also can reasonably estimate their own costs and their opponents' costs. Thus, the assumption that the pretrial process can be modeled as a sequential game of perfect and complete information is reasonable.

Because the model focuses upon the pleading nexus, it collapses the pretrial process down to the minimum game necessary to replicate the general sense of actual litigation. In concrete terms, this means that only the major landmarks of the pretrial process merit decision nodes. Thus, the analysis does not include Bayesian updating, instead assuming that the parties retain the same assessments of their litigation prospects throughout the pretrial process. The four critical nodes in the model are (1) the plaintiff's "sue/do nothing" decision point, (2) the defendant's decision to resist or offer to settle, (3) if the defendant makes an offer, the plaintiff's decision to accept or refuse, and (4) if defendant resists, the plaintiff's decision to try or drop the case.[109] Relevant data include each party's

[106] "Perfect information" is not to be confused with "complete information," a different state of the world in which each party is also fully aware of all parties' possible strategies and payoffs.

[107] Changing this assumption would complicate any given analysis but would not affect any of the relevant insights derived from the model.

[108] Differing estimates of cost create additional computational complexity and can in certain circumstances yield different results, but they do not affect the validity of the model.

[109] *See, e.g.*, ERIC RASMUSEN, GAMES AND INFORMATION: AN INTRODUCTION TO GAME THEORY 59–62 (3d ed. 2001).

expected value of trial, expected direct litigation costs, and the expected external costs associated with various possible decisions. Because the game as modeled is one of perfect and complete information, it is fully determined and thus amenable to backwards induction analysis.[110]

B. Building the Model

A rational, risk-neutral, profit-maximizing plaintiff will file a lawsuit only if she expects to benefit financially from its filing. Thus, she will file only if the expected value of the suit—either from trial or from settlement—exceeds the costs of filing and prosecuting the claim. In the simplest game theoretical models of the litigation process, the decision to file a lawsuit involves only an assessment of whether the costs of filing and litigating the case are less than the probability-adjusted expected verdict. If the expected net economic costs of filing and prosecuting the lawsuit are less than the expected value of the claim at trial, then the plaintiff has an economic incentive to file. Conversely, if the plaintiff's expected costs exceed her expected gain; traditional models suggest that she will have no incentive to bring the suit.

But a more realistic model must account for the realities of the U.S. civil litigation system, and especially for the fact that much of the cost of litigation is unavoidable once the pleading bar is passed. The Rules contemplate a world with extensive, costly pretrial discovery and minimal judicial oversight. And it is a world where each party expects to bear its own pretrial costs. Thus, the court's decision at the pleading nexus to dismiss the case or to allow it to proceed is the primary driver of the pretrial cost function. Once a case survives pleading challenges, it is difficult if not impossible to mitigate litigation costs under prevailing legal standards absent some effective means of filtering pretrial discovery.[111]

Several economic variables are critical to any realistic model. First and foremost, the model must reflect each party's expected pretrial litigation costs—filing fees, discovery costs, expert witness fees, etc.[112]

---

[110] *See, e.g.*, DOUGLAS G. BAIRD ET AL., GAME THEORY AND THE LAW 50–55 (1994).

[111] *See, e.g.*, Celotex Corp. v. Catrett, 477 U.S. 317, 322–23 (1986) (authorizing "no evidence" summary judgment only after there has been adequate time for discovery).

[112] As discussed in Part II.B., *supra*, discovery costs are likely to be the single largest component of pretrial costs, often by a substantial majority.

Second, the model must incorporate an estimate of the outcome at trial. This is typically expressed as a verdict amount adjusted by the probability of obtaining that verdict—in economic terms, an expected value of trial.

The model also must account for the external costs associated with specific litigation decisions. For example, the model should incorporate the external costs to the plaintiff of trying or dropping a case, and the external costs to the defendant of settling a case without trial. Thus, if a plaintiff will damage its reputation by dropping a frivolous claim after filing or by pursuing a frivolous claim to disposition, its disposition/dismissal payoffs should incorporate negative external cost terms.[113]

Similarly, a defendant's decision to settle may carry with it external costs, in the form of likely follow-on litigation or reputational harm, for example. If settling a specific claim will encourage other similarly situated parties to file suit, the defendant's "settlement" payoff must incorporate a negative external cost term as well.

Finally, the model must also include the parties' expected trial costs in the event that the defendant chooses not to settle the case. If an equilibrium settlement amount exists, it can be derived from those terms.

Consider a hypothetical suit involving Donald Defendant and Peggy Plaintiff, one in which Donald (and Peggy's) objective assessment of defense liability risk is zero—the paradigmatically "frivolous" suit.[114] If Peggy files the claim, Donald does not expect to have any trial costs at all (the case is destined for resolution in Donald's favor on Donald's summary judgment motion), and the expected verdict is zero as well. Nonetheless, Donald may still face an overwhelming incentive to settle. If Donald believes that his pretrial costs will be $12, and that he will suffer $3 in external costs (reputation damages, copycat suits, etc.) from a pretrial surrender, he should be willing to settle for up to $9 in lieu of pursuing the claim all the way to a successful summary judgment. Once Donald accounts for his external costs, he is truly indifferent as to who receives his money.

But whether Peggy will file the negative-expected value suit in the first place requires putting the two pieces of the puzzle together. In a game of perfect and symmetrical information in which each party shares the same

[113] This is most likely when the plaintiff or its agents are likely repeat players in front of the same tribunal.

[114] *See* Robert G. Bone, *Modeling Frivolous Suits*, 145 U. PA. L. REV. 519, 524–33 (1997) (exploring the difficulty inherent in defining "frivolous" litigation).

beliefs regarding the relevant variables, the outcome is fully determined. Thus, Peggy's "file/don't file" decision can be divined through backwards induction, informed by the interaction of Peggy's own cost/benefit calculus and Donald's.[115] If Donald's expected costs of suit, minus the external costs of settlement, exceed Peggy's expected costs of filing, she will have an incentive to sue regardless of the substantive merits of her claim.

So, if Peggy anticipates that an objectively frivolous claim (expected verdict of $0) will cost her $5 to litigate through pretrial,[116] she will nonetheless have an incentive to file suit if Donald's expected pretrial costs, less his external costs of settlement, are over $5. For example, again assume Donald expects to win at summary judgment with 100% certainty, but expects to pay $12 to get to the summary judgment node and would suffer only $3 in external costs from a pretrial settlement.[117] Donald would thus be indifferent as to paying $12 to litigate the case to dismissal, suffering no external costs along the way, or paying Peggy $9 in settlement and suffering $3 in external effects from that settlement. Because Peggy's total economic cost of litigating the suit through pretrial is only $5, she will file her claim, anticipating that Donald will have an incentive to settle before summary judgment for more than Peggy expects to pay to litigate the claim.

Once Peggy has filed her suit, Donald has an incentive to settle, and should in fact be willing to settle for any amount up to his pretrial-costs-less-external-costs indifference value. But just as Peggy was able to execute a bit of litigation judo through her knowledge of Donald's pretrial

---

[115] *See* BAIRD ET AL., *supra* note 110, at 50–55.

[116] One avenue for further research would involve explicit modeling of the agent-principal difficulties inherent in cases handled by attorneys. As a general rule, the model predicts that the agent-principal problem would tend to *reduce* the risks of "frivolous" suits because an additional layer of agency increases potential external costs. In this model, external costs tend to reduce the incentive for strategic behavior on both sides, by (1) increasing the plaintiff's total expected pretrial costs and (2) decreasing the amount for which the defendant would be willing to settle. It is possible to imagine scenarios in which the agent-principal relationship pushes incentives in the opposite direction (for example if a lawyer seeks a specific reputation in the marketplace), but they are less likely.

[117] Another area for further refinement: accounting for the likely inverse relationship between the amount of settlement and the magnitude of expected external costs. This model assumes that the external costs of settlement are fixed; a slightly more complex model might adjust the external costs as a function of settlement amount. The core insights of the model do not change, but it would impact both the outcomes of certain close cases and the likely settlement within the range (tending to push the settlement down to the plaintiff's minimum).

costs, so too can Donald use Peggy's expected costs against her. At the end of the day, Peggy is indifferent to any combination of costs and settlement that will yield her the highest net return. In the hypothetical above, Peggy can be made to spend $5 and get nothing in return. Knowing this, Donald will make Peggy an offer she can't refuse: Before the parties incur pretrial costs, Donald will settle the claim for $4. Peggy's net payoff is the $4 settlement, and Donald's net losses are $7, including external costs.

The case ultimately settles for $4 because Donald's threat to force Peggy to expend pretrial resources is credible. By insisting on his legal rights before the court, Donald can require Peggy to comply with discovery requests and otherwise perform her pretrial obligations, at minimal additional cost to him. [118] If Peggy were to demand more than $4 in settlement, Donald could simply enforce Peggy's pretrial obligations, thus cutting into her net proceeds.[119]

Peggy does not enjoy a reciprocal ability to force Donald to settle for a greater percentage of his indifference value, because her threat to impose costs on Donald is not credible. Suppose Peggy seeks more than Donald's initial $4 offer, claiming that she, too, will impose costs on Donald such that his net losses increase. From Donald's perspective, any amount Peggy forces him to spend on pretrial litigation costs decreases Donald's indifference settlement value proportionally. For example, if in the hypothetical above, Donald expends $4 of his expected $12 in pretrial costs, his indifference value is now considerably lower. It would cost him only $8 more to obtain summary judgment, thus he would now only be willing to settle for $5, since his total payout at settlement would also include $3 in external costs. Peggy's threat to impose costs on Donald is credible ex ante, but she cannot use that threat to increase her settlement amount above the difference between her expected costs and Donald's indifference value.

[118] In this model, enforcement (filing of the claim, promulgation of discovery requests, motions to compel, etc.) is assumed to be effectively costless to the parties; these costs are generally nominal relative to other pretrial costs. A subsequent iteration of the model might incorporate enforcement costs, but it is unlikely to add much to the analysis.

[119] Peggy's external costs of suit only enter into the equation if Donald's impositional rights extend that far. Though his threat to force Peggy to expend $5 in pretrial litigation costs is credible, his threat to force Peggy to incur external costs may not be, because it would require that Donald expend his full pretrial costs. Since Donald ultimately has an incentive to settle, regardless, he may only be able to obtain a discount off of his full indifference amount proportional to plaintiff's internal costs of litigation.

Thus, at the end of the day, on these facts Peggy will file suit without any expectation of a merits recovery, and the case will ultimately settle for $4.

C. General Implications of the Model

If this model captures the essence of the real-world litigation process, there is a real risk of cost arbitrage only when substantial pretrial cost disparity favors the plaintiff. But empirical evidence regarding the proportionality or disproportionality of pretrial economic costs in litigated claims is hard to come by for two reasons. First, a huge percentage of litigated cases settle—70% or more according to one widely cited study.[120] Thus, although researchers have explored average litigation costs in a variety of case types, they cannot reliably estimate the parties' full expected costs of litigation. Rather, the existing estimates tend to show how much money the parties actually spent to reach resolution, whenever it occurred. Second, it is difficult if not impossible to estimate the various external costs that may arise in connection with pursuit or settlement of claims. Existing research thus tends to focus only upon the accounting costs of litigation; there is no obvious source of empirical data on the external costs associated with various litigation behaviors.

That said, there is reason to believe that most litigated cases do not involve substantial cost disproportionality favoring plaintiffs. Intuition, common sense, and the available data suggest that most pretrial costs are either roughly equal as to plaintiffs and defendants, or that they favor defendants rather than plaintiffs. In other words, the litigation market is often self-correcting. But several types of real-world claims do seem to present a risk of systemic cost disparity favoring plaintiffs. These types of claims, while no more than a small fraction of litigated cases by number, are problematic. And as discussed in Part V, the risks presented by these claims can be limited by adoption of a stricter default pleading standard applicable only to such claims.[121]

---

[120] *See* Herbert M. Kritzer, *Adjudication to Settlement: Shading in the Gray*, 70 JUDICATURE 161, 162–64 (1986); *see also* Marc Galanter & Mia Cahill, *"Most Cases Settle": Judicial Promotion and Regulation of Settlements*, 46 STAN. L. REV. 1339, 1339–40 (1994).

[121] *See infra* Part V.D.1–3.

1. A Note About Agency

Civil litigation and agency relationships go hand-in-hand. Most important among these are the attorney-client and client-insurer relationships present in much civil litigation. But there has been no significant research on the effects that agency relationships have as to the appropriate pleading standards. Perhaps somewhat surprisingly, in the pleading context, these relationships on balance tend to mitigate the risk of opportunism rather than increase it.

For example, for each of the primary agency relationships—plaintiff/plaintiff's attorney, defendant/defendant's attorney, defendant/insurer—the presence of the agent tends to increase the external costs associated with filing or settlement of spurious claims. Though the plaintiff herself may suffer few external consequences for the filing of a frivolous claim, her attorney has a reputation to protect; the presence of a plaintiff's attorney—especially one likely to be a repeat player either before a particular court or against particular defendants or their insurers—will decrease the economic attractiveness of opportunistic pleading by increasing the plaintiff's external costs of filing. The defense attorney cares about her reputation as well, and a reputation for settling nuisance suits is not likely to attract clients.[122]

The story is the same for insurers when they are present in a case. In fact, the presence of insurance can dramatically increase external costs of settlement, because an insurer's reputation is effectively its business. A single defendant may worry about follow-on litigation to some degree, but an insurer whose business is built on maximizing premiums and minimizing claim payments may be far more concerned about the reputational effects of settlement. And because insurers typically cover similar risks for many clients, the risk of follow-on litigation is not limited to the specific defendant in the case; insurers are concerned about follow-on litigation against other insured entities as well. Thus, the agency relationship between insurers and their insured, while it may have perverse consequences elsewhere, tends to mitigate the risk of cost arbitrage by increasing the defendant's external costs of settlement for frivolous claims.

---

[122] The traditional agency tensions between attorney and client—maximizing fees, minimizing effort, etc.—usually have little effect upon the pleading calculus. Contingent fee plaintiff's attorneys' "file/don't file" incentives are roughly aligned with their clients' incentives. Hourly attorneys on both sides face obvious incentives to maximize revenue in each case, but those are likely held in check by countervailing longitudinal incentives to maximize revenue over time.

2. Plaintiff's Costs

a. Internal Costs

All else equal, the risk of cost arbitrage increases as the plaintiff's internal pretrial costs of litigation decrease relative to the defendant's. Thus, holding defendant's costs constant, we are less concerned when the plaintiff's anticipated pretrial costs are high. In abstract terms, the plaintiff's internal pretrial costs are likely to be high when she possesses substantial discoverable information. This occurs primarily when one or both of two conditions are satisfied: (1) the plaintiff's transaction with the defendant is in and of itself relevant to the existence or nonexistence of legal liability, or (2) the plaintiff's damages will be subject to substantial inquiry and potential dispute. In both situations, the defendant will have legitimate need to inquire deeply into plaintiff's affairs. This inquiry will increase the plaintiff's costs.

In the vast majority of litigated cases, both conditions are satisfied: the plaintiff's transactions with the defendant are directly relevant to the liability inquiry and the plaintiff's damages are genuinely contestable and subject to substantial inquiry. Take, for example, one of the most common types of civil litigation, the automobile accident tort suit. In those cases, the transaction at issue routinely requires substantial inquiry into the plaintiff's own behavior and perceptions: Did the plaintiff's own negligence cause or contribute to the accident? What does the plaintiff remember about the relevant events, and how does her recollection differ from the defendant's, the police report, and the physical evidence? As important, the typical automobile accident tort claim involves substantial inquiry into the plaintiff's claimed damages; in fact, this is often the most hotly contested and expensive component of the litigation.

When the plaintiff's involvement and activities in the relevant transaction is important to the liability determination, plaintiff will incur costs in connection with her case. The same is true for contestable damages claims; a fight over damages implies direct and opportunity costs for the plaintiff in the form of traditional and expert discovery at the very least.

Automobile accident tort suits are not the only type of litigation in which the typical plaintiff bears high internal pretrial costs. Virtually every form of personal injury tort suit, breach of contract suit, and many statutory causes of action (e.g., single-instance employment discrimination) inherently require searching inquiry into the plaintiff's interactions with the defendant and into the plaintiff's purported injuries. And as discussed

below, in most of these sorts of cases, the defendant's own cost structure is likely to mirror the plaintiff's—the defendant's possession of additional information in certain types of cases is likely to be counterbalanced to a large extent, if not outright exceeded, by the costs the plaintiff incurs proving her damages.

The available data, while admittedly incomplete and imperfect, bears this out. A 1986 Rand Corporation study found rough parity between plaintiff's nominal litigation costs and defendant's costs in tort cases generally—$8,000 for plaintiff to $10,000 for defendant.[123] According to a 1988 Rand study, the cost disparity actually favored defendants for aviation accident claims; defendants paid an average of $49,000 to the plaintiffs' $72,000.[124] And though a 1980s study of asbestos claims showed a small disparity favoring plaintiffs, by the 1990s, that disparity had been reversed.[125] Keeping in mind that these data fail to reflect the full expected economic costs of litigation (or even to differentiate between tried cases and settled cases), the cost disparities between plaintiffs and defendants are actually quite small. After factoring in expected external costs, one might well expect that disparities apparently favoring plaintiffs disappear entirely or in fact shift to favor defendants in most common types of claim.

Conversely, claims in which the plaintiff's internal pretrial costs are low tend to be claims for which there will be little inquiry into the plaintiff's activities or damages. The paradigmatic case is the "fraud on the market" claim, in which the plaintiff's sole interaction with the defendant is through the purchase of the plaintiff's goods, services, or stock.[126] Other types of similar internal cost structures may be found in, for example, "pattern or practice" employment discrimination claims or certain types of creditor claims against bankrupt estates. The key is that in these cases, the defendant can impose few costs upon the plaintiff during the pretrial phase. In the "fraud on the market" cases, for example, defendant's direct discovery options are remarkably limited: How much did you buy? When did you buy? How much did you pay? The remainder of the inquiry

[123] JAMES S. KAKALIK & NICHOLAS M. PACE, THE RAND CORPORATION, COSTS AND COMPENSATION PAID IN TORT LITIGATION, R-3391-ICJ (1986).

[124] JAMES S. KAKALIK ET AL., THE RAND CORPORATION, COSTS AND COMPENSATION PAID IN AVIATION ACCIDENT LITIGATION xiii, xvi (1988).

[125] STEPHEN J. CARROLL ET AL., THE RAND CORPORATION, ASBESTOS LITIGATION COSTS AND COMPENSATION: AN INTERIM REPORT 60 (2002), *available at* http://www.rand.org/pubs/documented_briefings/DB397/.

[126] Consumer Antitrust, securities fraud, and other similar claims come to mind.

focuses entirely upon defendant's actions: Did it conspire with its competitors? Did it conceal material information from investors?[127]

b. External Costs

The plaintiff's external costs of suit are largely dependent upon the reputational consequences the plaintiff and her attorney will suffer if they file a frivolous claim.[128] A single frivolous claim that nonetheless yields a settlement is unlikely to generate substantial external costs for the plaintiff; viewed from the outside, this settlement is arguably more likely to signal a potentially meritorious claim than a nuisance suit payment. Thus, the real issues are whether repeated such filings and settlements, in the aggregate, may generate external costs, and in particular whether unsuccessful suits, dismissed by the court as frivolous, will impose external costs on the plaintiff as well.

Though precise quantification of these costs is impossible, intuition and common sense suggest a few situations in which external costs might exist for plaintiffs. In particular, plaintiffs are likely to face higher external costs when repeat player concerns—both between and among the same litigant and agent pools or before the same courts—are the greatest. Consider the typical automobile accident claim again. In these cases, external costs to the plaintiff are likely to be relatively high for several reasons. First, the nature of the personal injury plaintiff's bar and the venue rules in most jurisdictions[129] suggest that lawyers will necessarily be appearing before the same judges repeatedly. In such circumstances, the reputational consequences (primarily to the plaintiff's attorney) of filing spurious claims may be high.[130] Second, these cases tend to yield repeat games between the

[127] This deliberately ignores certain costs plaintiffs may eventually incur, most notably expert costs and discovery review costs. But plaintiffs cannot generally be forced into these expenditures in the same way they can be forced to search for and produce discoverable information. Expert costs may be unavoidable in some cases, but even then, defendants cannot rely upon parity of cost imposition functions to prevent opportunistic pleading. These costs will likely mirror each other on either side, leaving discovery costs to determine parity or disparity.

[128] For purposes of this Article, it is assumed that a plaintiff suffers no external costs for the filing of a claim ultimately found to have merit.

[129] *See, e.g.*, 28 U.S.C. § 1391 (generally fixing venue where defendant resides or where "a substantial part of the events or omissions giving rise to the claim occurred").

[130] Though they are highest when the plaintiff's frivolous claims are ultimately revealed publicly as such, external costs can also accrue in certain environments even if the plaintiff's attorney is able to obtain settlements.

plaintiff's attorney, the local pool of defense attorneys, and the area's important insurers. Lawyers with a reputation for filing frivolous claims ultimately put both their earning potential and their social status at risk.

If a given type of claim systematically yields similar reputational consequences to the plaintiff or her attorney, that category of suit is less likely to raise opportunistic pleading concerns. In addition, claim types in which attorney-agents are equally likely to end up on either side (e.g., bilateral business disputes) are also likely to bring with them higher relative external costs, as attorneys seek to protect their reputation as to both potential clients and potential attorney adversaries. Though the "external costs" effect for plaintiffs is likely substantially less significant than the internal cost function in dissuading opportunistic pleading, it nonetheless may be an important part of the litigation calculus in a surprising percentage of cases.

Cases involving low potential external costs bring a concomitantly higher risk of frivolous claims. Again, "fraud on the market" cases provide an illustrative example. In these cases, plaintiff's attorneys are often national class action firms with very little expectation of repeat litigation before a particular court.[131] In addition, though these attorneys may sue the same companies and fight with the same opposing counsel more than once, intra-community social pressures are much less likely to constrain behavior. These cases also visit few or no external consequences upon the plaintiffs themselves; the class action litigant may not even be aware of her participation until settlement, if ever, and even class representatives are unlikely to suffer reputational or other external harm for lending their names to frivolous suits.

3. Defendant's Costs

a. Internal Costs

All else equal, the risk of cost arbitrage suits increases as the defendant's internal pretrial costs increase relative to the plaintiff's. Holding plaintiff's costs constant, we are more concerned when the defendant's pretrial costs are relatively high. Thus, the defendant's internal costs story is essentially the negative image of the plaintiff's. For example, claims in which the defendant's internal litigation costs may typically be

[131] Unless they affirmatively prefer that court, of course.

relatively low include automobile tort claims, in which the defendant is likely to have similar or lower net costs because (1) he is not generally claiming damages, and (2) inquiry into the plaintiff's own damages is substantial. In breach of contract and other business cases, the defense's internal costs are likely quite similar to plaintiff's; it takes about as much effort to tell a "he said" story as the "she said" version.

As important, defendant's pretrial costs can for certain claim types be mitigated and managed through sequential or piecemeal case development. Thus, when a defendant's aggregate expected pretrial costs seem disproportionately large *vis-à-vis* the plaintiff's, the disparity may diminish or disappear if the defendant can persuade the court to limit initial discovery to a specific element of liability or damages. In a toxic tort case, for example, the defendant may be able to reduce pretrial expenditures by persuading the court to limit initial discovery to whether the plaintiff was exposed to the defendant's product. If the plaintiff cannot establish exposure, summary judgment will be appropriate under *Celotex*, regardless of the defendant's abstract culpability.

By contrast, the cases with the largest internal defense costs tend to be those (1) in which the scope and depth of genuinely discoverable information under Rule 26(b)(2) is significant, and (2) without an obvious factual transaction around which to limit discovery. Again, the "fraud on the market" cases are illustrative. In certain antitrust cases, for example, the defendant's summary judgment liability case effectively rests upon proving that it did not conspire with its competitors to raise prices or limit output. In securities fraud cases, the defendant similarly must prove that it did not conceal material information from investors.

At trial, of course, the plaintiff bears the burden of proof as to these matters. But, summary judgment law temporarily turns the burden on its ear. The "after adequate time for discovery" requirement essentially guarantees the plaintiff a full opportunity to inquire into the existence of liability. For most other claim types, this inquiry is inherently self-limiting; the plaintiff can inquire into the details of the transaction giving rise to her claim and little else. By contrast, fraud on the market claims arguably allow plaintiffs to look (or more accurately, to force defendants to look, largely at their own expense) far and wide for evidence that the defendant violated the law. Depending upon the operative pleading standard, there may be no obvious factual transaction upon which courts can focus and thus

limit plaintiff's inquiry.[132] The plaintiff can force the defendant to search a staggering percentage of his records—email, hard copy files, archives, etc.—looking for a potentially nonexistent needle in the haystack. And summary judgment will only be available to the defendant after the court is satisfied that enough of the haystack has been searched.[133]

b. External Costs

Defendant's external costs generally arise out of the fallout from settlement.[134] If a defendant's settlement of a case will induce copycat lawsuits or harm defendant's reputation, this reduces defendant's incentive to settle by increasing its overall expected costs of settlement. By contrast, if settlement has few external consequences, defendants will, *ceteris paribus*, be more likely to pay the arbitrage.

The presence of liability insurance likely significantly increases the external costs of settlement *vis-à-vis* cases in which insurance is not involved. The risk of repeat suits against that particular defendant remains constant, and the risk of similar suits against other insurance company clients increases. Thus, the presence of insurance will tend to mitigate the risk of opportunistic pleading in most cases.

In many cases, the defendant's attorneys may also incur external costs if they encourage settlement of spurious claims. In potential repeat-player environments, defense counsel may be less likely to encourage settlement, because she is worried about obtaining and retaining clients.

As before, the classic automobile tort case and the typical "fraud on the market" claim represent different ends of the defendant's external costs spectrum. Automobile accident claims are repeat player at virtually every level; the same attorneys, defense counsel, insurance companies and courts simply play mix-and-match over time. In those cases, therefore, though the defendant himself may have little at stake externally (unless he is a serially bad driver), the defendant's attorney and insurer have strong incentives to take the long view. An insurer with a reputation for quick settlement

---

[132] *See* Easterbrook, *supra* note 20, at 640.

[133] *See* Celotex Corp. v. Catrett, 477 U.S. 317, 332 (1986) (Brennan, J., dissenting).

[134] There are sometimes positive external payoffs from trying cases, but the model subsumes those benefits in the external cost term related to settlement. In other words, instead of including a positive externality term in the defendant's trial payoff corresponding to the defendant's decision to "send a message" to other potential plaintiffs, the model accounts for that incentive as a negative externality term at settlement.

invites additional claims, and thus an attorney with a reputation for quick settlement loses insurance company clients.

On the other hand, the "fraud on the market" case may involve significantly fewer external costs for defendant if the plaintiff's side has aggregated claims. In a price-fixing case, a single consumer's claim is not an attractive candidate for settlement, because settlement of one claim may induce thousands of additional claims.[135] External costs of settlement would therefore be high. By contrast, the defendant's external costs of settlement as to an aggregated claim (a putative antitrust class action, for example) may be relatively low; price-fixing claims are largely *sui generis*, and the defendant may rationally discount the expected cost of follow-on litigation. As if not more important, many "fraud on the market" claims are not typically covered by insurance,[136] thus eliminating the insurer's multiple client repeat player incentives to deter future litigation.

4. Summary and Implications

In summary, the risk of cost arbitrage is highest when: (1) the plaintiff's internal costs of litigation are lowest, as in claims in which the plaintiff has little discoverable information in its possession, custody, or control; (2) the plaintiff's external costs of filing a frivolous suit are lowest, as in claims in which the plaintiff's attorneys are unlikely to be repeat players in the same court or against the same insurer/payer; (3) the defendant's internal costs of litigation are highest, as in claims in which the defendant's discovery costs are high and cannot be filtered or sequenced to minimize expenditures; and (4) the defendant's external costs of settlement are lowest, as in claims in which there is little risk of reputational harm or copycat litigation and there is no insurance coverage.

The model thus has two primary real-world implications. First and foremost, there is a relatively low risk of opportunistic pleading in most types of civil claims. Though full empirical analysis is beyond the scope of the current Article, there is reason to believe that many of the most common

[135] This claim is also unlikely to be filed, because the economics are unfavorable for the plaintiff.

[136] In the last several years, insurers have begun to offer so-called "side C" or "entity" coverage in their Directors and Officers (D&O) liability policies. Those policies often expressly cover securities fraud claims, but do not typically cover antitrust claims. To the extent side C policies trigger external cost concerns at the insurer level, they will tend to reduce plaintiff's incentives to file frivolous claims.

forms of civil litigation—automobile torts, breach of contract cases, intellectual property disputes, etc.—inherently involve or can be made to involve[137] relative parity in pretrial costs, or even disparity in favor of the defendant. Thus, stricter pleading standards across all claim types are unnecessary. In addition, the presence of insurance, especially as to large corporate defendants that would otherwise be capable of paying judgments on their own, actually reduces the risk of opportunistic pleading by increasing the external costs of settlement. In most cases, the cost incentives are such that plaintiffs are likely to file suit only if they believe the net expected value of their claim at trial to be positive.

Second, there is a real risk of frivolous suit in the perfect storm. When cost disparity significantly favors the plaintiff, the expected trial value of her claim becomes irrelevant to her filing decision. In the extreme, the economic model predicts that a plaintiff may file suit—and the defendant may settle the claim—even when the plaintiff's claim is wholly frivolous.[138]

D. The Type I/Type II Correlation Complication

Though the model itself does not expressly speak to this issue, there is likely a significant correlation between the types of claims likely to raise Type I risks and the types of claims likely to raise Type II risks. Claims most likely to raise Type I risks are those where the plaintiff's own activities are less relevant to the claim; they are typically claims in which most of the relevant discovery costs are borne by defendants. But when a defendant faces disproportionately higher discovery costs, it is usually because he has control over more relevant information than the plaintiff does. Thus, cases in which the model predicts a substantial risk of cost arbitrage may also be cases in which the plaintiff's ex ante ignorance of relevant facts might lead to Type II error in the form of erroneous dismissal.

This correlation presents a conundrum: the cases in which a stricter pleading standard is most justified on Type I error grounds are precisely the cases in which a more liberal pleading standard is most justified on Type II error grounds. The prescriptive goal of this Article is to locate a solution to the Type I and Type II conundrum that adequately mitigates both types of risk.

---

[137] Most notably, through the filing of counterclaims, which often substantially increase the original plaintiff's costs.

[138] As discussed below, it is possible to have a rational probabilistic belief in a suit's merit, even in the absence of objective indicia of worth. *See infra* note 183 and accompanying text.

V. A POSSIBLE PRESCRIPTIVE APPLICATION: BALANCING THE SCALES

A. Theories of Procedural Justice

Because this Article ultimately offers a prescription for procedural reform, a brief discussion of competing procedural justice theories is appropriate. After all, commentators and critics often have quite different ideas regarding not only the appropriate shape of procedure, but also regarding the very goals of the civil procedural system. The sketch that follows demonstrates the near-universal appeal of this Article's approach.

Professor Larry Solum offers an immensely useful taxonomy of various competing philosophical approaches to procedural justice.[139] In general terms, Solum offers three possible philosophical focal points for the development of procedural systems: accuracy, balancing, and participation. That is, the three potential goals of procedure are (1) maximizing the accuracy (whether on a systemic basis or in case-specific terms) of legal proceedings such that litigation outcomes comport with the substantive law,[140] (2) balancing the various relevant factors and values tied up in civil litigation in some socially optimal way,[141] or (3) ensuring that interested parties have an opportunity to participate meaningfully in legal proceedings, perhaps independent of the effect that participation may have on the substantive result.[142]

---

[139] *See generally* Lawrence B. Solum, *Procedural Justice*, 78 S. CAL. L. REV. 181 (2004); *see also* Robert G. Bone, *Agreeing to Fair Process: The Problem With Contractarian Theories of Procedural Fairness*, 83 B.U. L. REV. 485 (2003).

[140] Solum, *supra* note 139, at 244–52.

[141] Solum offers at least two competing versions of a "balancing" approach, one expressly utilitarian and consequentialist, and one based upon deontological notions of fairness and rights. *Id.* at 253, 257. In the utilitarian version, the balancing approach would seek to minimize net social costs of litigation, which, as Solum expressly notes is an approach similar to many traditional law and economics approaches. *See id.* at 253–54. In the deontological version of balancing, Solum posits instead a system designed to result in maximization of other, potentially noneconomic values. *Id.* at 257–59. Both approaches are properly characterized as seeking imperfect procedural justice. *Id.* at 253.

[142] In his article, Solum ultimately argues for a theory of procedural justice that recognizes and in fact gives primacy to participation as an independently desirable and irreducible value. *See id.* at 305–08. Whether Solum's lexical ordering is preferable is irrelevant; the asymmetry problem associated with modern U.S. pleading standards exists even if participation is the only value we seek to protect.

1. Descriptive Accuracy

None of these theories is accurate and complete as a descriptive matter.[143] Instead, to the extent a coherent "philosophy of civil procedure" can be extracted from the patchwork of existing rules and standards, it is a theoretical pastiche of all three.[144] While the system contains features intended to maximize substantive accuracy, it does not do so without regard for cost—instead, it routinely considers matters of economic cost when designing procedural features. Nor does it balance costs solely on utilitarian grounds; our procedural jurisprudence frequently emphasizes rights and fairness—including "participation" values—despite the high probability that those rights sometimes cost more to protect than any purely economic benefits derived from their protection. And even the strongest articulations of participation values found in case law recognize that there are practical limits to the access that can be afforded, both in terms of the economic costs it imposes upon the system and other parties, and in terms of the countervailing impact participation one party can have on the rights of others.

In fact, the Federal Rules themselves expressly recognize the mixed nature of "procedural justice." Rule 1 embraces the tension in a single clause, admonishing courts to construe and administer the Rules "to secure the just, speedy, and inexpensive determination of every action . . . ."[145] Further evidence of mixed theoretical heritage abounds, from rules designed to allow for summary disposal of claims that fail to meet certain evidentiary thresholds[146] to Supreme Court pronouncements that explicitly privilege constitutional values over speed and efficiency.[147] The Supreme Court has authorized an explicitly economic situational balancing of burdens in the

[143] *See id.* at 243.

[144] *Id.* at 242 ("In this section we examine three simple . . . models of procedural justice that are, at least partially, implicit in current practice.").

[145] FED. R. CIV. P. 1.

[146] *See, e.g.*, *id.* 56(c).

[147] Fuentes v. Shevin, 407 U.S. 67, 90 n.22 (1972) ("'The establishment of prompt efficacious procedures to achieve legitimate state ends is a proper state interest . . . . But the Constitution recognizes higher values than speed and efficiency. Indeed one might fairly say of the Bill of Rights in general, and the Due Process Clause in particular, that they were designed to protect the fragile values of a vulnerable citizenry from the overbearing concern for efficiency and efficacy that may characterize praiseworthy government officials no less, and perhaps more, than mediocre ones.'" (quoting Stanley v. Illinois, 405 U.S. 645, 656 (1972))).

procedural context.[148] But the same Court later speaks reverently of the "deep-rooted historic tradition that everyone should have his own day in court" in rejecting a liberal application of preclusion doctrine.[149]

When attempting to describe the theoretical underpinnings of our existing system of civil procedure, it is impossible to disentangle the various strands. U.S. civil procedure is about accuracy. But, it is also about balancing of costs and other rights and values, and it is about participation as an independent value with independent, intrinsic worth. And though commentators may order their preferences differently, all of these values generally play a part in normative conceptions of procedure as well.

2. The Irrelevance of Normative Preferences

But, to the extent proceduralists differ with respect to their normative preferences, those differences do not necessarily imply corresponding differences in pleading standard preferences. All serious theories of procedural justice ultimately aim for some sort of fairness: fair results, fair balancing of the net costs and benefits of litigation, fair weighing of important rights and values, or fair access to the courts. The economic analysis of pleading identifies a set of incentives that can in certain circumstances call into question the fairness of civil litigation under virtually any definition of "fair." Assuming a symmetrical fairness baseline in which there is no systemic ex ante preference for or against a category of litigants, the operative normative theory of procedural justice is ultimately irrelevant.

Regardless of our normative conception of procedural justice, the economic analysis set forth in Part IV is troubling. For example, the analysis suggests that under certain cost disparity conditions, the substantive accuracy of civil litigation is suspect; defendants will settle claims without regard for their merits if the price is right. For these claims, stricter pleading standards may improve accuracy of result by limiting the incidence of such settlements.

Because the model predicts inefficient filing and settlements, it also necessarily suggests excessive social costs resulting from such claims. Provided the cure is not worse than the disease, a stricter pleading standard

---

[148] *See, e.g.*, Mathews v. Eldridge, 424 U.S. 319, 335 (1976).

[149] *See* Taylor v. Sturgell, 128 S. Ct. 2161, 2171 (2008) (quoting Richards v. Jefferson County, 517 U.S. 793, 798 (1996)).

for high-risk claims will also decrease net social costs in ways that promote utilitarian balancing goals.[150]

Finally, the model also suggests that under certain conditions, defendants' participation rights may be substantially limited by the pressure to settle. To the extent participation theorists insist upon meaningful participation,[151] the defendant's apparent right to participate in a frivolous case whose economics favor early settlement is arguably no right at all. He would be irrational to expend the resources necessary to vindicate that right. If a stricter pleading standard mitigates defendants' potentially perverse but rational incentives to settle certain frivolous claims, that standard also vindicates participation as an independent value. Thus, if we are genuinely concerned about participation for both sides to a dispute, balancing economic incentives serves that goal as well.

Because this Article approaches the pleading problem from a law and economics perspective, the prescriptive analysis that follows is framed in largely consequentialist terms. That is, the economic model and the prescriptions derived therefrom embrace a standard law and economics-influenced theory of procedural justice in which preference is given to minimizing net social costs on utilitarian grounds. But the step from utilitarian balancing to a preference for substantive accuracy is a small one, since the general upshot of the prescriptive approach is to limit defendant's incentives to settle frivolous claims. Somewhat less obviously, the consequentialist analysis also indirectly implies superior participation on the part of litigants, once we accept the notion that "meaningful participation" works both ways, and that defendants are also entitled to something more than a nominal right to appear in court.

B. Existing Pleading Literature

Before tackling the pleading problem prescriptively, a brief review of the existing doctrinal and economic literature is also in order.

Until recently, the scholarly literature on pleading standards was remarkably thin, with only a few significant pieces written from the 1930s through the early 2000s. Widespread scholarly interest in pleading is a remarkably recent phenomenon, tracing its birth to the Supreme Court's

[150] *See infra* notes 176–82 and accompanying text.

[151] *See* Solum, *supra* note 139, at 25–60.

2007 opinion in *Bell Atlantic Corp. v. Twombly*.[152] The source of renewed academic interest in pleading is obvious: In *Twombly*, the Supreme Court apparently puts a sacred cow out to pasture, describing key language from the Court's seminal pleading opinion as having "earned its retirement."[153] But, before *Twombly*, the literature was thin indeed.

In the 1920s and 1930s, Yale Professor and Dean (and later Judge) Charles Clark wrote a series of influential articles that effectively laid the groundwork for the national transition from "code" or "fact" pleading[154] to notice pleading.[155] Clark is correctly hailed as the chief architect and proponent of the original Federal Rules of Civil Procedure, and his then-radical approach to pleading was intended to be the cornerstone of the Federal Rules' approach to civil litigation.[156]

Taken together, Clark's voluminous pre-Rules writings on pleading standards are a scathing indictment of the then-current state of pleading law. Clark largely adopted the conventional wisdom that by the time of the Field Code's adoption in 1848, traditional common law pleading had become "an abstruse and involved science, based upon such technicalities that the movement for the so-called reformed or code pleading necessarily followed."[157] But in Clark's view, the "fact pleading" solution proffered by the Field Code and its progeny ultimately proved only a marginal improvement, evolving over time to require increasingly detailed filings, and resulting in the same risk attendant with common law pleading: that justice might be denied by complex, confusing, overly technical and sometimes unpredictable pleading requirements.[158]

Clark's criticism of the status quo eventually evolved into a broad endorsement of notice pleading, a concept codified in the "short and plain

---

[152] 127 S. Ct. 1955 (2007).

[153] *Id.* at 1669.

[154] Several states had retained common law pleading as well.

[155] CHARLES E. CLARK, CASES ON MODERN PLEADING (West Pub. Co., rev. ed.) (1952).

[156] *See, e.g.*, Richard L. Marcus, *The Puzzling Persistence of Pleading Practice*, 76 TEX. L. REV. 1749 (1998) [hereinafter Marcus, *Pleading Practice*].

[157] Charles E. Clark, *Simplified Pleading*, 27 IOWA L. REV. 272, 275 (1941). Clark did note, however, that the complexity and technicality of common law pleading was only part of the story. In Clark's retelling, the common law pleading system remained simple and direct for many categories of claims, but was brought into disrepute by lawyers' pleading practices in other case types. *See id.*

[158] *See id.* at 277 (decrying the risk of dismissal because of a "lawyer's mistake, induced perhaps by technical ignorance or even by lack of clarity of the decisions").

statement" phrasing of Rule 8(a)(2).[159] Ultimately, Clark's vision of pleading achieved near-universal acceptance among both courts and commentators. With very few exceptions, pre-*Twombly* scholarly follow-ups to Clark's work were broadly sympathetic to his view of pleading standards; the handful of significant post-Clark articles mostly criticized perceived statutory or common law shifts away from notice pleading after the adoption of the Federal Rules.[160]

Twombly has rekindled scholarly interest in pleading standards.[161] Though the occasional commentator supports all or part of *Twombly*'s apparent retreat from the broadest possible conception of notice pleading,[162] the thrust of most post-*Twombly* articles has largely been the same: notice pleading is a fundamental characteristic of U.S. litigation, and the Court's retreat in *Twombly* is indefensible.[163]

The existing doctrinal literature on pleading standards has roots in all three of the broad "procedural justice" categories discussed above. By far the most common normative justification proffered in support of a proposed pleading standard is that it will yield more accurate results than the alternative. Given the plaintiff-centric focus of most pleading scholarship, this accuracy preference is typically expressed in terms of the risks

[159] FED. R. CIV. P. 8(a)(2). The first significant scholarly endorsement of notice pleading appears to have been published in 1918, shortly before Clark began his serious writing on the subject. *See generally* Clarke B. Whittier, *Notice Pleading*, 31 HARV. L. REV. 501 (1918).

[160] *See generally, e.g.*, Fairman, *supra* note 1 (criticizing judicial and congressional imposition of heightened pleading standards on both "original intent" and "participation" grounds); Fairman, *Notice Pleading*, *supra* note 7 (suggesting that transsubstantive notice pleading is a "myth" and implicitly criticizing departures from notice pleading as historically and normatively flawed); Richard L. Marcus, *The Revival of Fact Pleading Under the Federal Rules of Civil Procedure*, 86 COLUM. L. REV. 433 (1986) [hereinafter Marcus, *Fact Pleading*] (criticizing courts' revival of fact pleading standards and suggesting increased role for summary judgment rather than stricter pleading standards); Marcus, *Pleading Practice*, *supra* note 156 (concluding that then-recent statutory and common-law developments did not justify departure from notice pleading approach).

[161] *See generally* sources cited *supra* note 97.

[162] *See generally, e.g.*, Max Huffman, *The Necessity of Pleading Elements in Private Antitrust Conspiracy Claims*, 10 U. PA. J. BUS. & EMP. L. 627 (2008); EPSTEIN, *supra* note 3.

[163] *See, e.g.*, Spencer, *supra* note 97, at 433 (describing *Twombly* as "an unwarranted interpretation of Rule 8 that will frustrate the efforts of plaintiffs with valid claims to get into court"). *See generally, e.g.*, Thomas, *supra* note 97 (concluding that *Twombly* is unconstitutional under the Seventh Amendment, and suggesting that *Twombly*'s result is undesirable on normative grounds as well).

associated with rejecting plaintiffs' potentially meritorious claims before they have an opportunity to develop their cases.

Other scholars appear to adopt a "balancing" approach to the question. For these scholars, pleading standards are part of a bigger cost-minimization or value-maximization story. To date, most scholarship adopting a balancing approach has concluded that any problems inherent in a liberal pleading standard are best resolved at other points in the pretrial process.[164] As we have seen, these conclusions are suspect at best.

Finally, some scholars seem at least implicitly to endorse a "participation" theory of procedural justice, in which access to the courts (again, primarily for plaintiffs) is seen as having value independent of its effect on outcomes.[165] None of these perspectives is inherently incorrect. But because the existing literature tends to focus on only the plaintiff's side of the equation, these perspectives are incomplete.

There are essentially two problems with the existing doctrinal pleading literature: First, it largely fails to consider the bilateral economic incentives facing litigating parties. Second, to the limited extent the existing doctrinal articles do attempt to engage with the "frivolous suit" problem, or with a cost-balancing approach more generally, their defense of universally liberal pleading standards typically relies upon solutions that are inherently unworkable, sometimes for multiple reasons.[166]

The single most important consideration in crafting a pleading standard is the extent to which expected pretrial cost disparity influences parties' behavioral incentives. And because the problems associated with significant cost disparities cannot readily be solved through improved case management alone,[167] summary judgment proceedings,[168] or post-resolution sanctions regimes,[169] the pleading standard is effectively the only moving part left to adjust.

---

[164] *See generally, e.g.*, Marcus, *Fact Pleading*, *supra* note 160.

[165] *See generally, e.g.*, Fairman, *Notice Pleading*, *supra* note 7.

[166] *See supra* notes 14–68 and accompanying text.

[167] *See supra* notes 1468 and accompany text. As discussed below, more effective case management may be possible with a stricter pleading standard and the salience occasioned by the early pleading standard hearing this Article proposes.

[168] Summary judgment comes too late, and without compensation for pretrial expenditures. *See supra* notes 66–68 and accompanying text.

[169] *See supra* notes 46–61 and accompanying text.

C. Economic Literature

A similarly thin literature assesses the economic impact of pleading rules. Though scholars have developed a rich and complex economic perspective on the issue of frivolous litigation generally,[170] they have devoted remarkably little time to analyzing the critical role pleading standards play in filtering claims. Indeed, there appear to be only two significant preexisting economic analyses of pleading standards.

Professor Robert Bone devotes a full chapter to pleading standards in his book analyzing the economics of civil procedure.[171] But Bone's text is designed as a general introduction to the economic analysis of procedural problems.[172] Thus, his pleading standards analysis is primarily a stylized pedagogical example demonstrating the economic tradeoffs inherent in various procedural standards. Nonetheless, Bone's approach to the pleading problem is helpful and instructive, if ultimately incomplete.

Bone builds his analysis atop the classic Rosenberg/Shavell ("R-S") nuisance suit model.[173] Thus, though pretrial cost asymmetry is explicitly considered,[174] Bone's conclusions are limited somewhat by the assumptions in the R-S model. Bone's analysis focuses on the effects of two different but largely undefined pleading standards—"notice pleading" and "strict pleading"—upon the parties' incentives in frivolous lawsuits.[175] Bone explores these incentives in connection with three initial states of information: (1) plaintiffs and defendants both know ex ante that the suit is

[170] *See supra* note 104 and sources cited therein.

[171] *See* BONE, *supra* note 9, at 125–57.

[172] *See, e.g.*, Thomson West Foundation Press Product Lines, http://www.westacademic.com/Professors/FoundationPress/ProductLines.aspx?tab=2 (last visited Dec. 24, 2008) (describing the "Turning Point" series of explanatory texts).

[173] *See* BONE, *supra* note 9, at 45–50, 150–55 (following D. Rosenberg & S. Shavell, *A Model in Which Suits Are Brought for Their Nuisance Value*, 5 INT'L REV. L. & ECON. 3 (1985)). This Article's analysis is structurally similar to the Rosenberg/Shavell approach, save that it (1) models external costs independently, and (2) makes different assumptions about payoffs at different decision nodes, driven by the realities of U.S. pretrial practice.

[174] *See id.* at 45.

[175] The term "frivolous" is notoriously difficult to define, as Bone and others acknowledge. *See, e.g.*, *id.* at 41–43 (acknowledging definition problem, and defining a frivolous suit as one that either (1) is known by the plaintiff to be without substantive merit before filing, or (2) is filed by plaintiff without her conducting reasonable pre-filing investigation that would have revealed information sufficient to put the suit into Category 1).

frivolous, (2) plaintiff knows the suit is frivolous but defendant does not, and (3) defendant knows the suit is frivolous but the plaintiff does not.[176]

With respect to Category 1, relying upon R-S, Bone ultimately concludes that because "answering is seldom more costly than filing, the model predicts that few frivolous plaintiffs will find it worthwhile to sue."[177] As discussed below, though Bone's conclusion logically follows from the assumptions in the R-S model, it can in a few categories of cases be quite incorrect. The process of answering even a known frivolous lawsuit can be staggering in the real world of U.S. civil litigation, because the U.S. litigation process does not provide defendants with exit points in between pleading and post-discovery summary judgment. These costs are not only significant in absolute terms, but can also be substantial relative to the plaintiff's filing costs, especially for case types in which plaintiff's expected pretrial expenditures are comparatively low. Thus, Bone's conclusion that "few frivolous plaintiffs will find it worthwhile to sue"[178] may be true across the great run of civil litigation, but for certain cases involving significant pretrial cost disparity, even mutual foreknowledge of frivolousness will be insufficient to deter filing. Moreover, when the defendant's absolute costs of answering (better described as defendant's pretrial costs-to-disposition) are large, inefficient settlements can be correspondingly large as well.[179]

In Bone's Category 2, the plaintiff knows ex ante that her claim is frivolous, but the defendant does not. Bone acknowledges that this category produces significant risk of frivolous filings, but rejects the conclusion that pleading standards would affect the plaintiff's incentives.[180] According to Bone, the plaintiff "will simply fabricate the necessary allegations," even under a strict pleading regime.[181] But Bone may have ignored the potentially significant value of forcing plaintiffs to fabricate in circumstances where cost disparity creates perverse incentives. In

[176] *See id.* at 150–55. Bone explicitly rejects consideration of the fourth possible state of information, neither plaintiff nor defendant knows the claim is frivolous, but acknowledges that "mutual ignorance also invites frivolous filings that can produce high costs in equilibrium." *See id.* at 150 n.34.

[177] *Id.* at 150.

[178] *Id.*

[179] The R-S model also fails to consider the impact of external costs of settlement upon litigants; these costs have interesting implications for litigation decisions as well.

[180] *Id.* at 151.

[181] *Id.*

particular, a stricter pleading standard will in turn decrease the defendant's anticipated pretrial costs, because the discovery necessary to disprove more concrete allegations will be less than that required to defeat notice-pled allegations at summary judgment.

Finally, Bone considers situations in which the defendant knows that the claim is frivolous, but the plaintiff does not. In these situations, Bone concludes that a heightened pleading standard may be appropriate in such cases when investigation costs are moderate.[182] Bone is correct to point out the possible costs of a heightened standard as to these cases—some number of ultimately meritorious suits will never be filed—but he does not consider the hybrid pleading approach offered by this Article.

In sum, though Bone's basic approach is sound, the model upon which he bases his conclusions departs from economic reality in ways that significantly affect the parties' incentives. Moreover, Bone's simplified approach fails to consider whether a change in pleading standard may in fact affect the parties' expected costs of litigating to disposition, and thus their game theoretical incentives. Also, Bone effectively assumes transsubstantivity, basing his conclusion on the intuition that the small category of cases for which he believes strict pleading would be superior do not counterbalance the large category of cases for which notice pleading is the better option.

Keith Hylton has recently added to the literature on pleading economics.[183] But Hylton takes a significantly different approach to the problem. Building to some extent upon the two-stage "frivolous suit" model of Lucian Bebchuk,[184] Hylton ultimately suggests what amounts to a sliding scale pleading standard that should "vary with the evidentiary demands of the associated legal standards and the social costs of litigation."[185]

Hylton's approach is effectively unilateral; he suggests that "the threshold merit level [required to survive a motion to dismiss] should increase as litigation becomes less productive as a deterrent and more costly

---

[182] *Id.* at 155.

[183] *See generally* Keith N. Hylton, *When Should a Case Be Dismissed? The Economics of Pleading and Summary Judgment Standards*, 16 SUP. CT. ECON. REV. 39 (2008).

[184] *See id.* at 48–50; *see also* Lucian Arye Bebchuk, *A New Theory Concerning the Credibility and Success of Threats to Sue*, 25 J. LEGAL STUD. 1 (1996).

[185] Hylton, *supra* note 183, at 62.

to society."[186] That is, Hylton's approach largely ignores the possibility that civil litigation might effectively self-regulate as plaintiffs and defendants look to their own expected litigation costs in addition to their opponents. Instead, Hylton's prescription is based largely upon absolutes. He recommends dismissal at the pleading stage for claims where the "claims and evidence asserted at the pleading stage are insufficient to meet the merit requirement at the summary judgment stage" and the "discovery-enhanced" merit level (i.e., the merit of the claims after discovery) is not "unambiguously greater than the summary judgment merit requirement."[187]

In Hylton's model, the "summary judgment merit requirement" increases as the social cost of litigation increases.[188] Hylton suggests different pleading standards depending on the ratio of false convictions to false acquittals and the total nominal costs of litigation to the parties and society.[189] So Hylton envisions a stricter pleading standard when "the plaintiff's claims impose relatively high costs on the defendant."[190]

Hylton offers several concrete examples in defense of his proposed standard. He first considers fraud claims as an example of claims for which the underlying legal standard "imposes a heavy burden on plaintiffs."[191] He then justifies a stricter pleading standard for such claims on the grounds that it will "ensure that a claim admitted into court was likely to meet the evidentiary requirements of the legal standard" and "reduce the frequency of socially wasteful litigation."[192] Similarly, Hylton suggests that fraud claims carry high social costs, particularly in terms of damage to the defendant's reputation and the disruption in commerce that would result if contracts were subject to frivolous claims of fraud, and that these higher social costs also justify stricter pleading standards.[193] Hylton offers a similar analysis as to antitrust claims.[194]

Hylton's two-stage, sliding scale approach is certainly worthy of further study and consideration. But as a threshold matter, Hylton's decision not to

[186] *Id.* at 50.
[187] *Id.* at 51–52.
[188] *Id.* at 52.
[189] *Id.*
[190] *Id.*
[191] *Id.* at 58.
[192] *Id.*
[193] *Id.*
[194] *See id.* at 59–63.

incorporate a game theoretical analysis of the litigation process may ultimately result in overtreatment of the disease. Specifically, for claim types characterized by high net social costs but relative pretrial cost parity between plaintiffs and defendants, a stricter pleading standard may be unnecessary; the plaintiff in such cases has substantial incentive to file only claims she reasonably believes have substantive merit. Addressing Hylton's fraud example directly, if the plaintiff's expected pretrial costs under a notice pleading regime are equal to or greater than defendant's,[195] there is little risk that she will file a fraud claim without reasonable expectation of success on the merits. We do not need a high pleading burden to keep out frivolous or abusive claims in that situation.

Applying a game theory filter to the social cost problem more generally yields a similar result. Though it is certainly true that "an allegation of fraud could severely damage a business,"[196] a game theoretical approach suggests that defendants would be unwilling to incur these costs as to frivolous claims. When there is relative cost parity among plaintiff and defendant, the defendant has a strong incentive to pursue frivolous claims to disposition in order to protect that reputation. Because the plaintiff would be aware that the defendant's threat to litigate is credible, she would be deterred from filing such claims in the first place. The same is true to some extent with respect to "disruption in commerce" costs as well.[197] Finally, a sliding scale standard of the sort Hylton proposes may be difficult to implement in practice, as courts struggle to generate consistency and predictability in a world of nearly infinite variation.

In sum, though the existing economic literature on pleading standards advances the ball, it fails to fully consider the impact of bilateral economic incentives on the parties' litigation decisions.

---

[195] *See supra* Part IV.C.2.

[196] Hylton, *supra* note 183, at 58.

[197] *Id.* Defendants may not always have an incentive to internalize fully the "disruption in commerce" costs Hylton identifies. That category of costs may be subject to greater or lesser collective action problems, depending upon the impact a single suit would have upon the defendant's perception of its ability to contract in the future.

D. Prescribing Balance

1. Reimagining Transsubstantivity

If the model accurately captures the essence of pretrial economics, its implications are clear: a unitary transsubstantive pleading standard may do more harm than good. To the extent there exist some cases or claim types for which pretrial costs are typically balanced or favor the defendant, the model demonstrates that the pleading standard should err on the side of correcting informational asymmetries. For those claim types, the risk of strike suit is minimal, and thus the pleading standard can and should legitimately focus upon correcting plaintiff's presumed knowledge deficit. Notice pleading often works.

The calculus is vastly different when there is substantial pretrial cost disparity favoring the plaintiff. For those claim types, the risk of cost arbitrage is real, thus suggesting that a stricter pleading standard may be appropriate. But those claims are also likely to carry a concomitantly higher risk of Type II error; claims involving higher discovery costs for defendants are generally claims in which defendants control more of the relevant information, so the risk of erroneous dismissal is higher for these claims as well.

Given the challenges associated with designing a pleading standard for high-risk claims, and given the relatively straightforward analysis of low-risk claims, it may be preferable to use cost disparity as a dividing line, and to apply different default pleading standards on each side of the divide. The economics of pleading suggest that most garden-variety civil claims should be subject to traditional notice pleading, but that high-risk claims are better addressed through a stricter default pleading standard.

Given the fractured reality of our modern transsubstantive approach, a bifurcated standard based upon cost-disparity criteria may actually satisfy the goals of transsubstantivity more effectively than existing pleading regimes. The analysis of this Article ultimately implies replacing claim specific pleading standard enhancement (e.g., securities fraud claims) with a truly universal standard that diverges only on the basis of cost and not claim type.

2. Notice Pleading for Low-Risk Claims

Traditional notice pleading is more than adequate for the majority of existing claim types. Any category of claims that is not likely to engender

substantial cost disparities favoring plaintiffs is unlikely to invite inefficient claims and settlements. With concerns about cost asymmetries irrelevant, the only remaining concern as to these claims is the possible informational asymmetry favoring defendants in most cases.

Notice pleading is well-suited to address this asymmetry. Traditional notice pleading requires only that the plaintiff provide the defendant notice of the transactions, occurrences, or events that constitute the wrong, and that the plaintiff state what she wants from the lawsuit. This liberal standard, while potentially problematic for claims subject to a high risk of frivolous suit, allows plaintiffs to proceed in low risk suits even without knowing all of the details of their claims. Thus, for the great majority of existing claims, the economic analysis of pleading standards suggests that no change is necessary.

3. Strict Pleading for High-Risk Claims

By contrast, a notice pleading standard may not be preferable for high-risk claims. As the name implies, notice pleading requires only that the plaintiff give the defendant sufficient notice of her claim to allow the defendant to prepare a defense. Under a notice pleading regime, pretrial cost disparity will likely be at its highest, because notice pleading does not require the plaintiff to allege details around which the court can sequence or limit discovery.

To mitigate Type I risk, it would be economically preferable if the default rule required the plaintiff to allege facts tending to support each element of her claim. This stricter pleading standard need not and should not be overly formal; *Twombly*'s plausibility requirement may suffice.[198] But formal or informal, strict pleading is superior to notice pleading in high-risk situations along a number of dimensions.

[198] In context, *Twombly* effectively revives fact pleading, at least as to antitrust conspiracy claims. The factual practices alleged by the plaintiffs would have been illegal if agreed to among the defendants, thus if the plaintiffs had alleged "plausible" facts as to the existence of an agreement, they would have alleged facts in support of every element of a claim. Fact pleading is an all-or-nothing proposition; fact pleading is effectively meaningless without requiring facts supporting every element. *See infra* notes 231–45 and accompanying text for further discussion of the possible contours of a stricter standard.

a. Strict Pleading Lowers Three Forms of Social Cost

i. Inefficient Filing and Settlement: The Type I Problem

Strict pleading has obvious economic advantages over notice pleading. At its most basic level, by requiring the plaintiff to provide more detail, strict pleading can shrink defendant's expected internal costs of litigation substantially.[199] If a claim otherwise susceptible to cost arbitrage risk satisfies a stricter pleading standard, the additional details in the pleading will become the focal point of factual inquiry in the case. Compared to notice pleading, the specificity of stricter pleading will help courts overcome their own rational ignorance.[200] By focusing on the facts that make plaintiff's allegation plausible, the court can sequence much of the expected cost to defendants such that truly frivolous claims will be subject to summary judgment at substantially lower cost. [201]

All of this in turn reduces the problematic disparity between the plaintiff's expected litigation costs and the defendant's expected litigation costs. Because the risk of Type I error is directly related to the difference between plaintiff's expected costs and defendant's, decreasing defendant's expected costs of litigation in turn reduces the risk of pretrial cost arbitrage. The lower defendant's expected pretrial litigation costs, the less potential plaintiffs will be interested in filing strike suits.[202]

ii. Administrative and Other Social Costs Also Decrease

Strict pleading also likely decreases other social costs traditionally associated with litigation. Under strict pleading, some lower number of cases will be brought, and fewer cases will survive the pleading stage. Thus, one would expect the deadweight administrative costs of civil litigation—costs associated with running the courts—to be lower under

[199] In cases with substantial Type I risk, defendants are likely to be quite motivated to seek judicial intervention. When the relevant facts supporting the claim are clearly alleged, several of the traditional precipitants of judicial disengagement are less likely to factor in.

[200] *Cf.* Easterbrook, *supra* note 20, at 638, 644–45 (noting that with a return to fact pleading, judicial management of pretrial costs becomes more plausible).

[201] *See* Hylton, *supra* note 183, at 51–52 (arguing for symmetry between dismissal and summary judgment standards).

[202] A stricter, fact-driven pleading standard for such claims will provide the court with the ability to conduct meaningful case management; the bifurcated regime envisioned by this Article, complete with a cost/bond hearing if necessary, will provide opportunity and incentive.

strict pleading than under notice pleading. As important, the cases that survive a relatively more rigorous pleading standard are likely to be less expensive to develop and try in absolute terms, especially for defendants. Strict pleading limits the overall systemic cost of litigation.

4. Wrongful Dismissals: The Type II Problem

Compared to notice pleading, strict pleading should be superior along at least three cost dimensions for high-risk claims: Type I error costs, administrative costs of the litigation system, and direct litigation costs for cases passing the pleading bar all should be lower than if notice pleading were the standard. The critical question, then, is whether the benefits of strict pleading are associated with an increase in Type II error costs.

Any tightening of the pleading standard carries with it the risk of additional Type II error. If any chance exists that an apparently frivolous pleading—that is, a claim with no obvious objective factual support apparent at the pleading stage—may nonetheless turn out to be valid, then a move from notice pleading to strict pleading carries with it some additional risk of Type II error.

It is extremely difficult to measure relative error rates; in fact, the model predicts settlement of even wholly frivolous cases when cost conditions are right, so the fact of settlement in earlier cases is of no value in assessing the viability of underlying claims.[203] And in the absence of a natural experiment, change in relative error rates occasioned by a shift in pleading standards is particularly difficult to measure or predict.[204] Because measurement of error rates is so difficult, this Article suggests a solution designed to limit Type II error risk when good-faith plaintiffs otherwise would be unable to meet heightened strict pleading requirements in high cost disparity (and thus high-risk) contexts.

---

[203] If settlement amounts were made public, there might be some potential for quantitative analysis, but strike suits do not necessarily encourage settlements in which settlement amounts are disclosed publicly. *See generally* Scott A. Moss, *Illuminating Secrecy: A New Economic Analysis of Confidential Settlements* 105 MICH. L. REV. 867 (2007).

[204] One long-term advantage of the bond requirement proposed below may be the accumulation of data regarding Type II error rates in high-risk cases. Since the defendant's incentive to settle on cost-of-defense grounds alone is eliminated, summary judgment rates, settlement rates, and verdicts would provide meaningful data as to the actual incidence of Type II risk.

a. Balancing the Equation: A Bond Requirement

There will be some ultimately meritorious high-risk claims for which plaintiffs cannot initially satisfy a heightened fact pleading standard. Viewed ex ante from the plaintiff's perspective, such claims are essentially a matter of luck, subjective belief, or probability. If the plaintiff cannot plead plausible details, the fact that the claim ultimately has value may be wholly fortuitous. Alternatively, the plaintiff may genuinely believe her claim has a positive expected value, but that belief may be based upon subjective criteria that cannot be articulated in terms of the plausibility facts supporting the claim. Finally, despite the absence of plausible factual knowledge, plaintiff may nonetheless rationally believe her claim has value on the basis of population compliance data.[205]

Consider the plaintiff's thought process in the latter context: "In my experience or according to reliable sources, some percentage *X* of all similarly situated defendants are guilty, causing damages of $*Y*. Though I have no objective or subjective reason to believe this particular defendant is guilty, it is nonetheless rational to file suit because my expected value of filing suit is *X* times $*Y*, which exceeds my expected costs of litigation $*Z*."

To the extent we are concerned that any dismissal or failure to file will be unacceptably common with the adoption of strict pleading in high-risk contexts, those risks can be ameliorated by allowing plaintiffs to opt out of the strict pleading requirement by posting a bond to be forfeited to the defendant in the event the claim was without merit. The following sections detail the theoretical underpinnings of a bond requirement and offer thoughts on real-world implementation of the scheme.

i. Economics of the Bond Requirement

Acceptance of notice pleading to mitigate Type II error risk carries with it a substantially increased risk of Type I error in the form of cost arbitrage suits. The model presented in Part IV predicts that the risk of inefficient settlements increases as the disparity of defendant's pretrial costs (offset by settlement externalities) to plaintiff's pretrial costs increases.[206] If the equation can be balanced in favor of the defendant, the risk of cost arbitrage decreases.

---

[205] Here, "population compliance" refers to aggregate rates of compliance and noncompliance with legal duties among similarly situated parties.

[206] *See supra* discussion at Part IV.

A risk-neutral plaintiff with a genuine but wholly subjective or probabilistic belief that her claim has a positive expected value should be willing to opt out of strict pleading standards (which she cannot, by definition, satisfy) by posting a bond in order to proceed through discovery on a notice pleading standard.[207] In order to accomplish its goals, the bond would first need to be large enough to overcome any disparity between defendant's expected pretrial disposition costs for a frivolous suit and plaintiff's expected pretrial disposition costs for the same claim. In addition, upon dismissal or rejection of the plaintiff's claim—preferably at summary judgment or before—the bond should be surrendered to the defendant rather than to the court. If the defendant knows ex ante that he will receive the bond amount at disposition of a frivolous claim, he would no longer have an incentive to settle instead of litigating to disposition. Thus, the theoretically optimum bond would equal the difference between the plaintiff's expected costs of litigation on one hand and the defendant's expected pretrial costs less his external costs of settlement on the other. Put another way, the bond should be equal to the expected strike suit settlement value of a frivolous claim.

In our hypothetical case between risk-neutral Peggy and risk-neutral Donald, Peggy's pretrial costs are $5, Donald's pretrial costs are $12, and Donald's external costs of settlement are $3. Without the bond, Donald's settlement indifference value is $9 and the case ultimately settles for $4. If Peggy were required to post a bond in the amount of $4 to proceed under a notice standard, pleading, she would have no incentive to bring the claim unless her expected value of claim was probabilistically positive. For a wholly frivolous claim with a $4 bond posted, Donald would be indifferent between proceeding to disposition, where his net costs would be $8 ($12 in costs, partially offset by a $4 bond payment) and settlement of $5 (where his total costs would also be $8: $5 in settlement plus $3 in externalities). Because Donald could then counterbalance the full amount of his indifference value by imposing $5 in pretrial costs upon Peggy, Peggy would have no ex ante incentive to bring the claim absent good-faith belief in its substantive merit.

[207] Risk-neutral in this context means that the plaintiff is indifferent between any particular actual payout and a probabilistic payout of the same expected value. Thus, a risk-neutral lottery participant would be indifferent between receiving a certain $1 cash payout and a chance to win $1,000,000 with 0.000001 a probability. Risk-seeking plaintiffs (those willing to trade a certain payout for a *lower* probabilistic payout) may be willing to post bonds even when they have no subjective or objective reason to believe their claim has merit.

ii. Practical Challenges and Solutions

Within the model, the bond requirement solves the Type I problem by completely eliminating a plaintiff's incentives to arbitrage the difference between its own expected litigation costs and the defendant's. But the model and the real world diverge in several potentially important ways; real-world implementation of a bond requirement must address those differences and the challenges they create.

The challenges associated with implementation of a bond solution fall into three categories: expectation, calculation, and location. In the real-world, the parties may have different beliefs as to their own expected costs and the expected costs of their adversaries. Even if this challenge can be overcome, it may be quite difficult to calculate expected costs ex ante such that the bond amount is set at an appropriate level. And finally, significant practical challenges are associated with the location of the bond forfeiture itself: at what point and under what conditions should the defendant be entitled to recover?

iii. Differing Expectations

In the real world, litigating parties do not often share identical ex ante perceptions as to their own expected costs or their adversaries' expected costs. In our scaled-down hypothetical world, for example, a defendant may expect to spend $10 defending a claim through summary judgment, while the plaintiff believes that the defendant will spend only $7. Differing real-world expectations can play havoc with the efficient resolution of claims. If the model itself were adapted to allow differing expectations without simultaneously permitting iterative settlement negotiations, those differing expectations regarding the parties' pretrial litigation costs could ultimately derail settlement entirely.[208]

One real-world mechanism for addressing this concern would instead address the difference in expectations. There is, presumably, a "right answer" to the pretrial costs question; at the end of discovery, both plaintiff and defendant will have expended certain sums on the litigation. Thus, it might be possible to mitigate the effects of expectation differences by bringing the parties' expectations closer together. Perhaps a judicial

[208] For example, consider the situation in which the plaintiff's estimate of defendant's pretrial costs is higher than defendant's own estimate. If the model allows only for a single settlement offer from plaintiff, the offer will be too high for defendant to accept.

determination of expected costs after hearing would be sufficient; even if it fails to align the parties' expectations perfectly in real life, it is likely to bring them closer together, and may in any event offer a useful fiction for the purposes of establishing a bond amount.[209]

iv. Difficult Calculations

Assuming the parties' differing expectations ultimately can be harmonized, it still may be difficult to calculate the appropriate bond amount. A court tasked with determining the relevant cost disparity is quite likely to face claims of high cost from both plaintiffs and defendants. The plaintiff has an incentive to overstate her expected costs because this will decrease the ex ante disparity upon which the bond is based; the defendant faces incentives to overstate costs because it increases that disparity and thus the bond. In addition, pretrial costs are often difficult to estimate ex ante even when the parties are acting in complete good faith.

The most attractive solution to the calculation problem may be to put the adversary system to work. By requiring the parties to litigate the issue of pretrial costs before a judge or magistrate, evidence relevant to expected costs can be brought to light and tested. The evidence most relevant to this determination would be cost data for similar previous cases. Though there is certainly reason to believe that the litigants would reach for the outer limits in arguing that their current case will be costly to litigate, data on past cases may be surprisingly reliable. In any litigated case, the defense attorneys' incentives to maximize fees are counterbalanced by their long-term desire to obtain and retain clients; excessive defense costs over time tend to lose clients and decrease firm profits. Plaintiff's attorneys billing by the hour are subject to similar pressures. In addition, plaintiff's attorneys working on contingency fee face a slightly different set of competing incentives. Though they may be interested in generating hours to improve their attorneys' fees in "lodestar" cases, that incentive is counterbalanced by the fact that they maximize their recovery and minimize their opportunity costs by doing as little work as possible for any given level of return. Thus, though the parties may stretch and push with respect to why their current

[209] It is worth noting that the single-iteration binary settlement negotiation of the model does not reflect real-world practice; to the extent the parties continue to differ as to cost expectations, they may continue to negotiate and persuade. These ongoing negotiations and the concomitant shifts in expectations can produce settlements when a single-iteration binary model suggests none is possible; a judicial fact-finding exercise may prove an adequate substitute.

case is different (behavior judges are fully competent to evaluate on the merits in any event), data on earlier similar cases may be relatively reliable.[210]

Calculation of cost disparities will be difficult and, to some degree, inherently speculative. But an adversarial process backed by data from previous cases should help the court arrive at a reasonable estimate, in addition to helping the parties harmonize their own expectations regarding the likely costs of litigation.

v. Case-by-Case or Claim Type?

Should the "fact pleading with an opt-out bond regime" be applied on a case-by-case basis or by claim type? In a world without transaction costs, case-by-case is best. Not all antitrust claims will present Type I risk, and a claim type approach will necessarily involve some errors along the margins. But transaction costs are a real and persistent feature of litigation, and it will be far from costless to determine whether any specific individual claim in fact raises cost arbitrage concerns.[211] Thus, it may be preferable to develop a claim-type approach over time, under which the default presumption is that notice pleading will apply, with specific high-risk claim types subject to the heightened fact pleading standard.[212] That said, as an initial matter at least, it may be necessary to develop a body of judicial knowledge through case-by-case analysis. And transaction costs can be limited to some degree by providing defendants with economic disincentive to invoke the higher pleading standard absent substantial cost disparity.

Finally, assuming it is possible to identify claims subject to fact pleading and opt-out bond posting, when should the plaintiff forfeit her bond? Any line we might draw will be potentially arbitrary on the margins,

---

[210] Taken to its logical extreme, it is unlikely that attorneys will attempt to increase their costs systematically simply to be in a better position when a pleading bond issue arises.

[211] *See generally* Isaac Ehrlich & Richard A. Posner, *An Economic Analysis of Legal Rulemaking*, 3 J. LEGAL. STUD. 257 (1974); *see also* Easterbrook, *supra* note 20, at 640–41.

[212] One alternative to rigid categorization would be analogous to antitrust law's per se/Rule of Reason continuum. Using that framework, certain claim types would be presumptively subject to fact pleading, others presumptively would not, and intermediate cases could get a "quick look" or truncated pleading standard analysis. Intuitively, even this regime is likely to be expensive; further research might explore whether the additional accuracy benefits would be worth the transaction costs. Another option might be to create a rebuttable presumption of fact pleading, affording the plaintiff an opportunity to explain why the pretrial cost disparity for a generally high-risk claim would be insufficient to trigger frivolous pleading risks in her specific case.

but as an initial matter, it may be preferable to designate summary judgment as the forfeiture line. Thus, a plaintiff who has opted out of fact pleading would surrender her bond to the defendant in the event that she dismissed her claim before summary judgment, or lost her claim at summary judgment. Viewed in the context of a complaint that is allowed to proceed despite the absence of plausible factual allegations, summary judgment, granted only if there is "no genuine issue as to any material fact,"[213] comes closest to capturing what should be the triggering conduct: filing of a frivolous claim.[214]

E. The Bond Hearing

1. Practically

Moving from theory to practice is never easy. Before attempting to flesh out either the contours of the bond hearing or the economic incentives it would engender, it is worth addressing one significant question up front: is a bond hearing of this sort practical? What indication do we have that the process has real-world potential?

This question can be further subdivided into two additional questions. First, is the solution proffered by this Article politically feasible? And second, assuming it is adopted, can courts implement it effectively and efficiently?

a. Political Feasibility

Though no significant change to well-established (and largely transsubstantive) procedural rules will come without tension and debate, there is reason to believe that the dual-standard approach this Article proposes is politically feasible. First and foremost, the time is ripe for such a change. The upheaval caused by *Twombly* and the credible arguments on either side of its purported plausibility requirement bespeak an environment in which a compromise might be possible.[215] As discussed above, the bond hearing opt-out offers the best of both worlds, especially when compared

[213] *See* FED. R. CIV. P. 56(c).

[214] This is not to suggest that any similar consequence is appropriate when a properly pleaded claim is nonetheless dismissed at the summary judgment stage.

[215] *See generally* Bell Atl. Corp. v. Twombly, 127 S. Ct. 1955 (2007).

against a "universal *Twombly*" alternative under which pleading standards for certain case types are tightened without an opt-out opportunity.

Second, there exists a well-traveled path for reformers to follow in the amendment of the Federal Rules of Civil Procedure. The Rules are relatively easily amended, and the notoriously ambiguous pleading standard in Rule 8(a)(2) has already long existed side-by-side with the heightened pleading standards set forth in Rule 9.[216] Rules recognition of a cost disparity-driven standard split is not an unthinkable departure. Moreover, it is possible that the amended standard could in fact replace the oft-criticized heightened pleading standard in Rule 9; an effective cost disparity-driven rule would largely supplant the need for claim-specific pleading standards.[217]

Third, though it may be beyond the power of the Supreme Court to mandate a bond hearing without formal rulemaking, the quasi-common law status of the pleading standard is virtually unique among the Federal Rules, which strive for clarity over flexibility as to most other rules. Thus, it seems that the Court could at the very least mandate a heightened pleading standard for high cost disparity cases following in the general vein of cases like *Twombly*. To the extent the Court is called upon to do so in the near future, any resultant standard split is likely to encourage a response from Congress or the Rules Committee.

b. Implementation

Though there are several potential challenges inherent in any attempt to implement a bifurcated standard with a bond hearing proposal, courts have sufficient resources and institutional experience to perform the task. Courts perform similar tasks in connection with a variety of matters. For example, Rule 65 allows courts to issue preliminary injunctions or temporary restraining orders only if the moving party provides "security in an amount that the court considers proper to pay the costs and damages sustained by any party found to have been wrongfully enjoined or restrained."[218] Thus, every preliminary injunction or temporary restraining order hearing necessarily should involve a structurally similar argument regarding the

[216] *See* FED. R. CIV. P. 8(a)(2); *id.* 9.

[217] This goes for the statutory heightened pleading requirements in the PSLRA, certain civil rights cases, etc. as well.

[218] FED. R. CIV. P. 65.

value of the provisional relief sought.[219] Civil courts also require bonds in other contexts, including appeals. In addition, although the reality of Rule 11 practice suggests that Rule 11 is a bad candidate to solve the pleading problem,[220] Rule 11 hearings have a similar character; the judge may impose sanctions, but they are "limited to what suffices to deter repetition of the conduct or comparable conduct by others similarly situated."[221] Thus, though courts have not often applied Rule 11 to deter misconduct systemically by balancing economic incentives, they have already been given the explicit authority to do so; the bond hearing would simply move up the determination to a point in the litigation where it is likely to have actual effect upon the parties' ex ante cost expectations.

Moreover, the specific types of information the courts must consider in a bond hearing are already well within their institutional competence to assess. For the last dozen years or more, courts have examined and passed judgment on similar information in the context of discovery dispute hearings, especially those concerning the collection, review, and production of electronically stored information.[222] For reasons discussed at length above, mere case management without a shift in pleading standards is unlikely to curb abuses in high cost disparity situations.[223] But courts are increasingly tasked with examining and evaluating the parties' expected expense claims in civil litigation.

Finally, the proposed structure of the bond hearing and the case-specific and longitudinal incentives of the parties will likely substantially mitigate the most serious challenges associated with the bond hearing process. The proposal does not solve every potential problem (nor does it wholly eliminate the risk of cost disparity-driven opportunistic pleading), but it

[219] Courts admittedly have not been overly consistent in their application of Rule 65(c). *See generally, e.g.*, Note, *Recovery for Wrongful Interlocutory Injunctions Under Rule 65(c)*, 99 HARV. L. REV. 828 (1986).

[220] *See supra* notes 46–61 and accompanying text.

[221] *See* FED. R. CIV. P. 11(c)(4).

[222] *See generally* Zubulake v. UBS Warburg LLC, 231 F.R.D. 159 (S.D.N.Y. 2005); Zubulake v. UBS Warburg LLC, 382 F. Supp. 2d 536 (S.D.N.Y. 2005); Zubulake v. UBS Warburg LLC, 229 F.R.D. 422 (S.D.N.Y. 2004); Zubulake v. UBS Warburg LLC, 230 F.R.D. 290 (S.D.N.Y. 2003); Zubulake v. UBS Warburg LLC, 220 F.R.D. 212 (S.D.N.Y. 2003); Zubulake v. UBS Warburg LLC, 217 F.R.D. 309 (S.D.N.Y. 2003); Zubulake v. UBS Warburg LLC, 216 F.R.D. 280 (S.D.N.Y. 2003).

[223] *See supra* notes 14–68 and accompanying text.

does offer a useful starting point for crafting a workable solution to the pleading problem.

2. Procedure

What, then, would the bond hearing look like? This inquiry has two distinct components. First, we must consider the general contours of the process. What are the primary waypoints along the way to case disposition, and how should the flow of the case be structured? Second, we must consider the practical specifics of the process. How will the court take and consider evidence? How much cost disparity is enough? How do we define "strict pleading" for those rare cases in which it becomes the default assumption? I take each component in turn.

a. General Contours of The Process

To get a handle on the process, we must go back to the very beginning of a hypothetical litigated case. The flow chart below offers a possible implementation of the cost disparity paradigm:

[INTENTIONALLY LEFT BLANK FOR CHART]

Several key features of this proposed process bear further discussion. First, the process assumes that the plaintiff's initial pleading satisfies the minimal notice pleading requirements under, for example, a traditional interpretation of *Conley*.[224] Second, in order to mitigate the risk that the defendant will play his own cost imposition games, the process envisions a sort of nondiscretionary unilateral English Rule for hearing costs associated with determining the appropriate pleading standard. For most commonly litigated case types, traditional pleading practice satisfies not only the minimal requirements of notice pleading, but also whatever heightened requirement we might choose to impose for high cost disparity claims. If the defendant challenges a pleading that satisfies the heightened requirement, he should be required to pay the plaintiff's costs of defending that pleading.[225]

Similarly, if after hearing the court determines that the defendant has failed to meet the requisite cost disparity threshold, then the defendant would again be required to pay the plaintiff's (substantially higher) costs associated with the evidentiary hearing. This is in some ways a draconian rule, and one that may in fact deter defendants from challenging the sufficiency of pleadings in all but the most significant cases of disparity. But given the Type II risk associated with heightened pleading standards, a default rule of this sort seems an appropriate starting point for system design purposes.

b. Specific Considerations

i. How much Disparity, and Measuring External Costs

No matter how elegant the economic model, real-world implementation is necessarily going to be imprecise. This is particularly true where, as here, the court must estimate the parties' future costs in order to make its decision. [226] Moreover, the bond hearing suggested by this Article's model

[224] *See generally* Conley v. Gibson, 355 U.S. 41 (1957).

[225] One possible exception to this requirement might allow defendants to avoid payment of hearing costs if the court chooses to manage discovery actively to limit cost disparity that would otherwise exist. Though it may be difficult for a court to manage discovery effectively in the absence of actual detail in the pleading, it is possible that for some claim types, the court would be able to act, once the defendant has brought the cost disparity to the court's attention.

[226] Nonetheless, we routinely ask courts and juries to predict the future in connection with damages awards and electronic discovery disputes, to name two examples.

raises two additional complicating questions: (1) how much cost disparity should trigger a bond requirement, and (2) can we or should we attempt to estimate external costs?

At the level of pure economic incentives, the model suggests that any net difference in expected pretrial costs favoring plaintiff in turn creates an incentive for opportunistic pleading.[227] Thus, one possible approach would require the court to impose a nondiscretionary bond in the amount of any demonstrated cost disparity under notice pleading. But there are several potential practical problems with such an approach.

First and foremost, given our longstanding commitment to liberal court access and the admittedly imprecise nature of the cost disparity analysis, a strict "any difference" policy could chill potentially useful litigation activity. In addition, at the very least, it would be inefficient to encourage parties to litigate the pleading standard when that determination would itself cost more than the bond arising out of it.

These factors suggest that courts should not impose a bond requirement unless the cost disparity analysis suggests a substantial incentive for plaintiffs to behave opportunistically. In other words, the bond should only be triggered if the disparity is greater than some substantial threshold amount. Note that this is not the same as asking judges to pass on whether the particular plaintiff in question seems to be behaving opportunistically. Rather, the analysis is economic only, and looks to the incentives engendered by the parties' cost expectations alone.

At this early juncture, it is difficult if not impossible to identify any relative or absolute cost disparity threshold that should trigger a bond requirement. But a few guidelines may be helpful in framing future debate on this issue.

First, the disparity should be great enough that any concern regarding the binary determination (whether cost disparity exists in the problematic direction) is minimized. Though the court will not be certain that its calculations are perfectly accurate, the disparity should be big enough that the court is convinced that some incentive-affecting disparity exists.

Second, the disparity needs to be large enough to offset any concerns regarding the impact that the parties' external costs might have on the calculation. Unfortunately, real-world judges likely cannot rely upon the parties to provide accurate, good-faith ex ante estimates of the external costs associated with filing or settlement of a claim. Nor can the court

[227] *See supra* discussion at Part IV.

assess these costs accurately and precisely on its own. First, though external filing and settlement costs are real, and though parties and their lawyers spend significant time thinking about them as they plan litigation strategy, they are extremely difficult to quantify. More important, the parties do not face any obvious incentives to tell the truth to the court when presenting their evidence of external costs, and the traditional adversarial crucible is unlikely to yield a significantly refined product. As discussed above, litigating parties do face incentives that encourage truth-telling with respect to their expected internal pretrial costs.[228] A variety of factors combine to suggest that parties will not routinely overstate their expected attorneys' fees or other "hard" litigation expenditures to a significant degree.

The same cannot be said for the parties' external costs. External costs are inherently indeterminate; the fact that they are not readily susceptible to either ex ante or ex post public calculation provides the parties with an opportunity to misstate their expected external costs in court proceedings. Defendants, for example, face a substantial incentive to understate their expected external costs. To the extent they can persuade a court that early settlement of the litigation in question is not likely to have follow-on consequences (in the form of copycat litigation, reputational harm, etc.), this increases the cost disparity modeled above, and thus increases the court's perceived strike settlement value.

The plaintiff's incentives point in the opposite direction, and with an added twist. A plaintiff seeking to minimize or eliminate a bond requirement would like to persuade the court that she will suffer substantial external consequences if she files a frivolous suit.[229] If the plaintiff could extract, for example, a $1 million settlement from a defendant but would suffer $2 million in external costs as a result, then filing a frivolous claim would not be worth it. Thus, the plaintiff's incentive is to overstate her external consequences of filing.

Moreover, the plaintiff's "external costs" story is at once both simpler and more complicated than that. In the typical case, a plaintiff will not

[228] *See supra* note 210 and accompanying text.

[229] Note that the external consequences to the plaintiff from filing a frivolous suit are likely to be zero or close to zero. If she successfully extracts a settlement from the defendant, regardless of the underlying substantive merit of the claim, judges will remain rationally ignorant of the reasons for the settlement, and there will be little or no information available from which others might assess the merits of the claim.

suffer any significant external consequences if she successfully extracts a settlement from a defendant. Even the most conscientious judge will remain rationally ignorant of the reasons for a settlement; it is not unlikely that a judge who initially saw a complaint as borderline or likely frivolous would change her mind when the parties settle. At the very least, settlement would tend to encourage judicial (and broader public) agnosticism as to the claim's underlying merits. Thus, though the model appropriately incorporates the plaintiff's external costs, those costs are likely to be quite low if all of the other economic incentives hold, and the defendant chooses to settle rather than litigate to disposition.[230]

For this reason, any early attempt to implement a cost disparity-driven pleading regime should probably not attempt to quantify the parties' external costs. It is safer by far to insist instead that the internal cost disparity be sufficiently large to offset the likely external consequences of filing and settlement (most of which will be borne by the defendant). Though precise quantification of external costs is neither advisable nor even possible, a competent judge should be able to estimate likely external costs sufficient to determine whether the internal cost disparities overwhelm external cost effects.

Finally, the disparity should be large enough to make the bond hearing process worthwhile. The bond hearing process will not be costless. Both parties will expend resources preparing for and participating in the hearing, and the court will invest its own resources as well. Thus, from a social cost standpoint, it would make very little sense to allow a bond hearing if net social welfare would actually be greater without one. At the same time, the economic model suggests that defendants will have no incentive to seek a bond hearing if it will cost them more than would a settlement resulting from cost disparity. If a defendant would be willing to settle plaintiff's claim for $50,000, then it would make no economic sense for him to pursue a bond hearing that would cost him more than that.

But what about a scenario in which defendant's bond hearing costs would be less than his expected settlement costs or the bond amount, but aggregate hearing costs (all economic deadweight loss) would exceed the bond amount?

[230] A future extension of the model might incorporate the plaintiff's probability-adjusted expected external costs, accounting for differing expectations between the parties as to expected pretrial costs.

ii. Setting the Bond

No magic formula exists for setting the bond amount in the real world. As discussed above, the calculation process is inherently imprecise, even as to the parties' internal litigation cost estimates; moreover, it may not be possible for a court to incorporate external cost terms into its analysis. Given that, how can the bond hearing be of any use at all?

Actually, the cost estimation process embodied in the bond hearing is enormously valuable, because it provides an anchoring point around which the court can craft an appropriate bond. By exploring cost disparity in a deliberate, systematic way, the bond hearing will provide a salient range within which the judge can exercise her discretion without being overly influenced by sympathy, emotion, or other factors irrelevant to the bond determination.

As discussed above, Rule 11 sanctions hearings provide a useful and humbling object lesson in the value of a bond hearing.[231] Despite Rule 11's explicit authorization of sanctions in an amount necessary to deter conduct among similarly situated parties in the future, courts rarely if ever impose sanctions for frivolous suits that in fact fully counteract the economic incentives giving rise to frivolous claims.[232] In fact, sanctions in even the hundreds of thousands of dollars are virtually unheard of. But the economics of pleading suggest that occasionally a bond in the millions of dollars may be necessary to ensure good-faith in a notice pleading environment.

Courts will not be able to estimate the bond amount with the economic precision implied by the model. But the bond hearing will help to ensure that the bond is in the right ballpark, and that it will have some actual effect upon the parties' litigation incentives and behavior.[233]

3. The Heightened Standard

Judge Clark certainly did not anticipate the cost disparity problem identified in this Article. That said, he was unassailably correct about one thing: heightened factual pleading standards present enormous practical and

---

[231] *See* FED. R. CIV. P. 11.

[232] *Id.*

[233] Given the substantial burden imposed by such a bond, and given the rarity with which the bond proceeding/heightened pleading standard is likely to be employed, it may be advisable to allow interlocutory appeal of the court's decision for both parties.

theoretical challenges.[234] Given that, what exactly should the default pleading standard be for claims presenting the requisite risk of opportunistic filing?

No magical phrasing or standard characterization is available to short-circuit the analysis. Fact pleading has long had different meanings in different times and in different places, and even if we were to adopt a specific historical incarnation of fact pleading, there would still be inherent ambiguity in the categories and requirements. For example, many fact pleading standards became hopelessly entangled in debates regarding the character of factual allegations, and as Judge Clark noted, distinctions between "law" and "fact" are rarely as clear-cut as one might hope.[235] Thus, we must generate a heightened pleading standard without specific reference to the fact pleading standards of the past.[236]

That said, the general idea behind heightened pleading in high-risk scenarios is to create factual focal points around which the parties and the court can build an efficient pretrial. The best way to accomplish this goal is to require the plaintiff to allege specific factual occurrences at a "5Ws"—who, what, when, where and why—level of specificity.[237] Applied to pleading, a 5Ws[238] approach would effectively require the plaintiff to do more than simply allege wrongdoing of a certain character. Instead, the plaintiff must include factual details that pin the wrongdoing to specific

---

[234] *See* Clark, *supra* note 157, at 277.

[235] *See generally id.*

[236] There is one potentially worthwhile distinction in the fact pleading regime: that between "evidentiary facts" and "ultimate facts." Though still susceptible to confusing interpretation and implementation, the distinction does capture the general gist of the problem. "Evidentiary facts" concerned the particular events of the case, e.g., "the defendant was driving west on Elm Street." By contrast, "ultimate facts" alleged the substance of a cause of action, e.g., "the defendant negligently struck plaintiff with his vehicle." Fact pleading required the pleading of ultimate facts rather than evidentiary. This gets things exactly backward. From an economic perspective, in cases of cost disparity plaintiffs should be required to plead *evidentiary* facts, not ultimate facts. Thus, one approach to the pleading problem would essentially turn fact pleading around; it would not be enough to allege that "defendants conspired in violation of the antitrust laws," but it might be enough to allege a series of evidentiary facts that if proven, tend to support that conclusion.

[237] *See, e.g.*, L.N. FLINT, NEWSPAPER WRITING IN HIGH SCHOOLS 47 (1917). From a pleading perspective, the less catchy "4Ws and an H" might be more appropriate; "why" is not typically relevant to the pleading process, but "how" often is.

[238] There are several additional formulations of this maxim, in addition to scholarly criticisms of its application to news-writing. For an example of the latter, see Philip F. Griffin, *The Correlation of English and Journalism*, 38 ENG. J. 189, 192–93 (1949).

dates, people and locations. Each of these factual allegations can then become a focal point for judicial management and limited pretrial inquiry; if that limited inquiry fails to produce summary judgment evidence, then the case can and should be dismissed at an early stage.

Consider a slightly streamlined version of the facts in *Twombly*. Under the most liberal view of notice pleading, an antitrust plaintiff might survive a motion to dismiss by alleging nothing more than, for example, a specific commercial practice of the defendant's was the result of a conspiracy among competitors. Pretrial inquiry regarding such an open-ended claim is likely to be similarly open-ended, with no realistic way for the judge to streamline discovery in a meaningful manner. By comparison, a 5Ws default standard would require the plaintiff to allege some details of the conspiracy: Where did it happen? Who was involved? When did the conspirators meet? What did they agree upon?

Despite their prolixity, the plaintiffs' pleadings in *Twombly* arguably did not satisfy a 5Ws standard.[239] Although the plaintiffs did offer substantial detail of the "what"—an alleged conspiracy among incumbent telecommunications providers not to compete with one another and to discriminate against non-incumbent competitors—they failed to allege any meaningful facts regarding the "who," the "where", or the "when" of the alleged conspiracy.[240]

What, then, would have satisfied a 5Ws requirement? The Supreme Court's announced standard in *Twombly* is too strict to be applied universally, and its apparent "quick look" to the merits is not an appropriate justification for the standard.[241] But the Court's choice of terminology in that case, plausibility, is helpful. A 5Ws pleading is one that contains enough factual detail to be plausible.

Thus, the addition of just a few additional factual allegations might have sustained the *Twombly* plaintiffs' claim. For example, if the complaint had alleged that the defendants initiated their conspiracy at a meeting at the Honolulu Hilton on June 23, 2003, that factual detail might have made the

[239] *See generally* Bell Atl. Corp. v. Twombly, 127 S. Ct. 1955 (2007).

[240] *Id.*

[241] *Id.* The point of a heightened pleading standard should be to reduce the *in terrorem* value of lawsuits, not by insisting that the pleadings demonstrate some objective indicia of merit, but by requiring pleadings that can be tested at relatively low cost when a notice pleading standard would generate disproportionately high costs for defendants.

claim plausible in a way that saved the pleading.[242] So too, if the complaint alleged that the defendants' commercial practices regularly changed in concert after, for example, regularly scheduled trade association meetings. Importantly, the point here is *not* to require the plaintiff to plead facts that suggest substantive merit, but rather to require facts around which a court can construct a more efficient, lower-cost pretrial.[243] If the *Twombly* plaintiffs had alleged a Honolulu meeting among various defendant executives, then the proof or disproof of that meeting would have been a logical focal point (and limiting factor) for discovery. If the plaintiffs had alleged a pattern of commercial practice changes after trade association meetings, the court could instead focus pretrial inquiry on those meetings.

The scientific literature offers a relatively close analogy: the heightened standard should require plaintiffs to plead easily "testable hypotheses" whose truth or falsity ultimately point to the existence or nonexistence of the underlying legal wrong.[244] In the social science literature, open-ended research questions are disfavored relative to testable hypotheses. Thus, "Are children more resilient than we think they are?" or even "Children are more resilient than we think they are" are too vague to be subject to rigorous social science research. By contrast, a genuinely testable hypothesis might look something like the following statement: "Children's scores on the Resilience Scale will negatively correlate with their parents' scores on the Perceptions of Child Resilience Scale."[245]

Though the analogy is imperfect, the distinction between notice pleading and strict pleading is similar. Under notice pleading, a claim that "defendant engaged in an antitrust conspiracy to fix prices" would be

[242] Perhaps "plausibility" is actually the wrong word; a pleading might still satisfy the 5Ws requirement (and thus be entitled to limited, focused discovery) if it alleged a conspiracy among telecom executives that took place on a tramp steamer bound from Havana to Miami with contraband Cuban cigars. The point is that the pleading provides a focal point for inquiry.

[243] There is some unavoidable overlap; however, courts will understandably be less concerned about letting cases move forward if the plaintiff has satisfied a heightened pleading standard. This is partially attributable to a likely decrease in pretrial costs, but it is also at least partially attributable to some increased likelihood of substantive merit. The point is that the Supreme Court's focus on "plausibility" is misplaced to the extent it is a proxy for an increased likelihood of defendant's ultimate liability.

[244] *See generally* Norman A. Desbiens, *The Presence of Hypotheses in the Scientific Literature*, 6 J. PHIL. SCI. & L. 50 (2006), *available at* http://www6.miami.edu/ethics/jpsl/archives/all/scientific_hypotheses.html.

[245] *See, e.g.*, Statistics Consulting Blog, http://statconsultant.blogspot.com/2008/01/stating-testable-hypotheses.html (Jan. 5, 2008, 11:55 EST).

sufficient to provide the requisite notice with few additional details. But to the extent discovery is analogous to scientific research, this is a remarkably vague and open question, and would not be an appropriate working hypothesis. By comparison, "defendants met to fix prices at the Honolulu Hilton on June 23, 2003" (1) can be proved or disproved with relative ease and certainty, and (2) points toward the existence or nonexistence of the underlying legal wrong. This is a relevant testable hypothesis. Again, the issue is not whether the pleading itself increases the likely merit of the plaintiff's claim, but rather whether it presents an easily (and cheaply) falsifiable hypothesis that impacts the substance of plaintiff's claim.

Thus, an assertion like, "On June 25, 2003, defendant ABC Corp. announced a price increase," would not be sufficient to qualify as a relevant testable hypothesis, because even if proved, it does little or nothing to advance the ball as to the underlying legal wrong: whether the defendants' pricing practices were the function of an intercompany agreement. By contrast, a pleading that details a pattern of trade association meetings followed by price increases would likely satisfy the standard.

Generic allegations of conspiracy or fraud, without more, are not readily susceptible to focused inquiry in the same way. In a world of liberal notice pleading, there is effectively no foundation around which a court can construct a more limited pretrial inquiry. Moreover, in the absence of the salience created by a bond hearing, judges will have little incentive to so limit discovery, even assuming it would be possible. By requiring that, in order to avoid a bond, the plaintiff must plead plausible "newspaper facts" or relevant testable hypotheses, the court will substantially reduce the costs associated with pretrial, thus reshaping the parties' ex ante cost expectations in a way that substantially mitigates the risk of opportunistic pleading.

It is important to note that the vast majority of litigated cases already satisfy the heightened standard. Breach of contract claims, personal injury tort claims, and most other forms of traditional civil litigation necessarily involve the pleading of a plausible factual nexus; they include details regarding the date of the accident, the identities of the contracting parties, the circumstances of the allegedly wrongful termination, etc. For the most litigated civil claims, the flow chart ends with the court's determination that the pleading satisfies even the heightened standard.

VI. CONCLUSION

The tension inherent in the selection of pleading standards makes it difficult to draw firm lines. Informational asymmetry and other advantages

favoring defendants certainly justify a liberal approach to case initiation. But there are countervailing concerns. In particular, when substantial pretrial cost disparities favor the plaintiff, she may file suit without regard for the suit's substantive merit, because defendants face an economic incentive to settle such suits rather than pursue them to disposition.

The economics of pleading thus suggest that the general preference for liberal pleading standards can sometimes go too far. To the extent notice pleading standards contribute to cost disparity by broadening the scope of pretrial inquiry and increasing defendant's pretrial costs, application of a stricter pleading standard may be appropriate when the cost disparity is sufficiently great. By requiring the plaintiff bringing a high-risk claim to satisfy a heightened pleading standard, the court can limit defendant's pretrial costs and thus reduce the cost disparities that can sometimes induce plaintiffs to file frivolous claims.

Any increase in Type II error occasioned by the adoption of bifurcated pleading standards can be addressed by inclusion of a bond requirement. Plaintiffs unable to satisfy the heightened pleading standard could nonetheless proceed through discovery on a notice pleading standard by posting a bond that would be forfeit to defendant if the plaintiff's claim is dismissed at summary judgment or before.

Though this solution does not wholly eliminate cost arbitrage risks, it represents a substantial improvement over both traditional notice pleading and the apparently universal heightened standard announced by the Supreme Court in *Twombly*. This Article's approach helps to break the Type I/Type II error deadlock and brings the law closer to the Supreme Court's enduring vision of pleading's important role in U.S. civil procedure: "to facilitate a proper decision on the merits."[246]

[246] *See* Conley v. Gibson, 355 U.S. 41, 48 (1957).

Appendix

This Appendix offers a slightly more formal version of the game theoretic model presented in prose form in Part IV of the text. As noted in Part IV, the model makes several standard assumptions (e.g., party rationality) and provides a single-period analysis. Because the model also assumes perfect and symmetrical information, the model is fully determined and initial actions can be predicted by backward induction.

THE BASIC GAME THEORETIC MODEL

The model focuses primarily upon the plaintiff's decision to file a lawsuit, reasoning backward from the defendant's likely actions in response to that filing. Let Π_p equal the plaintiff's expected internal pretrial costs, let Π_t equal plaintiff's expected trial/disposition costs, let γ equal the probability of a plaintiff's verdict at trial, and let x equal the amount of the verdict such that γx equals the consensus expected "merits" value of the claim. In the simplest case, the plaintiff will file suit if

$$\gamma x > \Pi_p + \Pi_t$$

That is, plaintiff will file suit if she rationally expects to recover more than her full litigation costs at trial.

But given the peculiarities of U.S. civil litigation, the defendant's expected payoffs at various junctures may be sufficient in themselves to encourage the plaintiff to file, regardless of the merits of plaintiff's claim. Assume, therefore, that γ equals 0, such that the expected value of plaintiff's claim γx is also zero.[247] Let S equal the proposed settlement amount, and let Δ_p equal the defendant's expected internal pretrial litigation costs. Further let Δ_t equal the defendant's expected internal trial or disposition costs, and let $e_{\Delta s}$ equal the defendant's expected external costs of settlement. If the plaintiff pursues her claim to disposition, the defendant's payoff will be

$$-\gamma x[0] - \Delta_p - \Delta_t$$

And if the plaintiff drops her claim after defendant refuses to settle, defendant's payoff will be

$$-\Delta_p$$

[247] Additional variations on this model might integrate positive expected value suits.

because the Rules allow for the full imposition of pretrial costs once the pleading stage is passed. By contrast, if the defendant settles the claim before incurring pretrial costs, its payoff will be

$$-S - e_{\Delta s}$$

Thus, viewed in a vacuum, the defendant has an incentive to settle any claim, regardless of the value of γx, for any amount S such that

$$S < \Delta_p - e_{\Delta s}$$

The plaintiff in turn has an incentive to file her lawsuit if her expected pretrial costs Π_p are less than defendant's settlement indifference value, or:

$$\Pi_p < \Delta_p - e_{\Delta s}$$

In that case, the defendant's threat to either pursue the litigation to disposition is not credible, because: (1) plaintiff's right to impose costs Δ_p upon defendant regardless of the merits of the claim ultimately makes disposition less attractive than settlement, once the external costs of settlement to the defendant are factored in; and (2) defendant cannot eliminate the full net value of a settlement by imposing reciprocal costs upon the plaintiff.

When these conditions hold, the model predicts a settlement equal to the difference between plaintiff's expected costs of suit and the defendant's indifference value:

$$S = (\Delta_p - e_{\Delta s}) - \Pi_p$$

To see why this is so, consider a hypothetical settlement negotiation between plaintiff and defendant in which plaintiff initially seeks to obtain the full value of the defendant's settlement/disposition surplus, making an offer S such that

$$S = \Delta_p - e_{\Delta s}$$

In that case, the defendant will respond by imposing pretrial costs upon the plaintiff up to Π_p. But the defendant's cost imposition strategy will not result in tit-for-tat response by the plaintiff. Every dollar in pretrial costs plaintiff actually imposes upon defendant reduces the defendant's willingness to settle by the same amount. If the defendant has spent $1 in pretrial costs, he need only spend $(\Delta_p - 1)$ to obtain dismissal (remember, the expected value of plaintiff's claim is assumed to be zero), and thus his new indifference value is $(\Delta_p - 1) - e_{\Delta s}$. Seeing this, the plaintiff will quickly recognize that although it remains economically preferable for the

defendant to settle rather than litigate, the defendant can nonetheless unilaterally impose pretrial costs on the plaintiff if he so desires (thus reducing the plaintiff's net settlement proceeds to the difference between plaintiff's expected internal pretrial costs and defendant's indifference value):

$$S = (\Delta_p - e_{\Delta s}) - \Pi_p$$

Thus, all else equal, a pleading standard that reduces or eliminates the difference between the plaintiff's internal trial costs and the defendant's indifference value concomitantly reduces the risk of opportunistic pleading of frivolous claims. As discussed in Part V, a fact pleading standard for high-risk claim types will often accomplish this goal.

FACT PLEADING BOND

In the event that a plaintiff bases her rational belief in the positive merits value of her claim upon only probabilistic evidence (e.g., general information about population compliance), it may be desirable to allow her claim to move forward upon the posting of a bond *B* that eliminates the risks of opportunistic pleading by balancing the parties' expected payoffs. This bond would be forfeited in its entirety to the defendant in the event of a summary disposition in defendant's favor, and thus the relevant payoffs would change. For the defendant facing a known frivolous claim, his payoff at disposition would be

$$-\Delta_p + B$$

Plaintiff's payoff at that juncture would be

$$-\Pi_p - B$$

To balance ex ante incentives, the bond would have to be set where the defendant becomes indifferent between pursuing the case to disposition and settlement at plaintiff's rational offer. In other words, the bond *B* equals plaintiff's equilibrium settlement offer *S*, which equals $(\Delta_p - e_{\Delta s}) - \Pi_p$. For any *B* greater than or equal to the difference between defendant's indifference settlement amount and plaintiff's pretrial costs, the plaintiff's expected net proceeds of a frivolous claim are zero or negative, because the bond combined with the defendant's ability to impose costs without retribution will make settlement, on net, a neutral or losing economic proposition. For simplicity's sake, I do not model bond-hearing costs

explicitly; they are likely to be extremely similar for both plaintiff and defendant in any given case, and would thus fall out of the equation.