

Duke Law Journal

VOLUME 58

MARCH 2009

NUMBER 6

LITIGATION DISCOVERY CANNOT BE OPTIMAL BUT COULD BE BETTER: THE ECONOMICS OF IMPROVING DISCOVERY TIMING IN A DIGITAL AGE

SCOTT A. MOSS†

ABSTRACT

Cases are won and lost in discovery, yet discovery draws little academic attention. Most scholarship focuses on how much discovery to allow, not on how courts decide discovery disputes—which, unlike trials, occur in most cases. The growth of computer data—e-mails, lingering deleted files, and so forth—increased discovery cost, but the new e-discovery rules just reiterate existing cost-benefit proportionality limits that draw broad consensus among litigation scholars and economists. But proportionality rules are impossible to apply effectively; they fail to curb discovery excess yet disallow discovery that meritorious cases need. This Article notes proportionality’s flaws but rejects the consensus blaming bad

Copyright © 2009 by Scott A. Moss.

† Associate Professor, University of Colorado Law School (scott.moss@colorado.edu). The author thanks the following: for reviewing and editing drafts of this Article, Professors Robert Bone, Roberto Corrada, Richard Marcus, Martha Minow, Alan Morrison, Paul Ohm, Nantiya Ruan, Michael Waggoner, and Phil Weiser; for offering feedback on presentations of this Article, the professors, attorneys, and judges who attended conferences and colloquia at the University of Colorado Law School, the Benjamin N. Cardozo School of Law, the Boston University Law School, the Chapman Law School, the Colorado Employment Law Faculty working group, the Workshop on Judges and Judging at American University Washington College of Law, the 2007 Colloquium on Labor and Employment Law, and the 2007 Midwest Law and Economics Association Meeting; and for excellent work as research assistants, University of Colorado Law School students Tiffany Howard and Lori Mayall.

rulemaking or judging. Rather, proportionality requires impossible comparisons between discovery value and cost before parties gather the evidence. Like other arguments that procedural rulings should depend on case merits, this Article notes how discovery has more probative value in close cases—yet a case's merits are unclear during discovery because the court cannot yet examine all the evidence.

In game theory terms, parties with discovery disputes cannot convey case merit credibly; courts have too little information, so low-merit parties can claim high merit, and courts are compelled to act as if all cases of a similar type warrant similar discovery. In this “pooling equilibrium,” ruling the same on all cases in the “pool,” regardless of merit, is courts' best strategy but a suboptimal one, yielding too much discovery in low-merit cases, too little in higher-merit ones. Thus, the quest for better discovery has disappointed not because of bad rules or decisions, but because courts and parties are stuck in a pooling equilibrium with information-timing circularity: optimal evidence gathering requires merits analysis, which in turn requires evidence gathering.

As a solution, courts could defer close decisions on possibly useful but costly evidence until meritorious cases separate from the pool, turning pooling into separating equilibria. Summary judgment can be this separation: cases going to trial after summary judgment not only have higher average merit than the pool of all filed cases, but are disproportionately likely to be the sort of close calls in which juries struggle to reach verdicts. No one yet has proposed post-summary judgment discovery to redress the costly discovery dilemma because summary judgment typically occurs only after all discovery, but high-cost evidence can be an exception to that usual sequence: cases surviving summary judgment are close calls warranting more fact gathering, so some costly discovery regularly denied should be allowed after summary judgment. Thus, the existing debate is too focused on limiting the amount of discovery; it should instead focus more on timing costly discovery optimally, to try to limit discovery to cases in which it is truly needed. Existing rules give courts discretion to use this proposal, but a new rule could minimize the risk of misusing the proposal to deny more discovery. This Article concludes by briefly noting how economic analyses must consider the details and information timing of the litigation process.

TABLE OF CONTENTS

Introduction	892
I. The Rise of Proportionality Limits on Litigation Discovery— and Their Disappointing Results	899
A. What Is Old Is New Again: Proportionality Requirements as a Solution to Tech-Driven Discovery Excess	899
B. The E-Discovery Rules: Just More Disappointingly Conventional and Ineffective Proportionality Limits	905
II. An Economic Analysis of Why Proportionality Limits, Though Popular, Cannot Be Optimal	909
A. The Consensus: Limit Discovery Based on Cost-Benefit Proportionality Principles	909
B. An Economic View of Discovery: Helping Factfinders Assess Case Value and Merit (L and p)—Which Makes Discovery Decisionmaking Circular	910
1. Size of Case (L): Difficult to Determine in Many Cases ..	912
2. Probative Value of the Evidence Sought (Δp): Difficult to Assess before Fully Analyzing That and Other Evidence	914
3. Probability the Claim Has Merit (p): Difficult for Court to Assess before Seeing All the Evidence and the Parties' Arguments about That Evidence. ...	915
III. Solving the Pooling that Prevents Better Discovery Decisions: In Close Calls about Costly Discovery, Preserve the Evidence But Delay the Discovery until after Summary Judgment	926
A. Replacing the Pooling Equilibrium with a Separating Equilibrium as a Case Progresses	926
B. Summary Judgment as the Key “Separating” Process that Allows Courts to Distinguish Cases by Level of Merit (p)	929
1. Stage I—Start of the Case, before Discovery, Motions, or During-Litigation Settlements	930
2. Stage II—After Dismissal Motions and Early Settlements (Including Settlements during Discovery)	932
3. Stage III—After Summary Judgment Motions	934

C. The Prescription: In Close Calls, Preserving the Evidence but Delaying the Discovery until after Summary Judgment	934
D. The Devil is in the Details: Making Workable the Proposal for Post-Summary Judgment Revisiting of Discovery the Court Had Denied Earlier	939
1. Concern Number One: Summary Judgment May Not Clarify Case Merits.....	939
2. Concern Number Two: Courts Should Use Alternatives Such as Sampling and Cost Shifting.....	941
3. Concern Number Three: Judges Might Excessively Deny Discovery and Grant Summary Judgment.....	944
4. Concern Number Four: Judicial Reluctance to Delay Trial to Reargue a Discovery Dispute	950
5. Concern Number Five: Loss of Evidence while Discovery is Delayed.....	951
Conclusion.....	953

INTRODUCTION

Most litigators agree that “discovery . . . is the battleground where civil suits are won and lost.”¹ “It is in discovery that the facts are developed,” sometimes leading to a dismissal of the case on summary judgment² but other times to “vital evidence” that yields a trial win or a “favorable settlement[.]”³ Discovery’s cost attests to its importance: in federal cases, discovery comprises half of all litigation costs; in the most expensive 5 percent of cases, discovery amounts to 90 percent of litigation cost and totals 32 percent of the amount in controversy.⁴ Yet on no other topic is there more disconnect between

1. Joseph D. Steinfield & Robert A. Bertsche, *Recent Developments in the Law of Access – 1998*, 540 P.L.I./PAT. 53, 107 (1998); see also *Hollander v. Am. Cyanamid Co.*, 895 F.2d 80, 85 (2d Cir. 1990) (“Because employers rarely leave a paper trail—or ‘smoking gun’—[of] discriminatory intent, . . . plaintiffs often must build their cases from pieces of circumstantial evidence which cumulatively undercut the [defendant’s] credibility . . . [R]efusal to compel [discovery] . . . deprived Hollander of evidence potentially helpful to . . . assembl[ing] such a quantum of circumstantial evidence . . .” (citations omitted)); Jon W. Green & Kyle M. Francis, *Age Discrimination in Employment: A Plaintiff’s Perspective*, 522 P.L.I./LIT. 227, 238 (1995) (“Discovery is the most crucial phase of age discrimination actions.”).

2. Steinfield & Bertsche, *supra* note 1, at 107.

3. Green & Francis, *supra* note 1, at 238.

4. Thomas W. Willging et al., *An Empirical Study of Discovery and Disclosure Practice Under the 1993 Federal Rule Amendments*, 39 B.C. L. REV. 525, 544–46, 548–49 (1998) (noting

the academy and bar; discovery controversies often are “not something that law professors pay a lot of attention to, but lawyers do.”⁵ Although discovery is the topic of numerous bar publications and conferences,⁶ it draws far fewer academic articles than hotter civil rights issues,⁷ even though discovery is crucial to civil rights litigation.⁸ Most of the limited academic writing focuses on *how much discovery* is too much for parties to request, at least without paying more of the production cost.⁹ Academics rarely focus on *how courts decide* discovery disputes (which, unlike trials, occur in most lawsuits), frustrating judges and parties alike. More academic focus is critical; if cases are won and lost in discovery, then they are really won and lost in discovery decisions.

The difficulty of discovery decisions took on a new salience once computerization brought litigation into a “brave new world”¹⁰ of costly “e-discovery” in which much evidence (perhaps 90 percent of corporate data) is digital, not paper,¹¹ and in which the *best* evidence can be e-mail sent in unguarded moments¹² or still-lingering “deleted”

also that only 15 percent of cases had *no* discovery, and 46 percent of those still had informal evidence exchange).

5. Richard Marcus, Assoc. Reporter, Fed. Judicial Conference Advisory Comm. on Civil Rules, Comment at the American Association of Law Schools 2008 Meeting, Section on Civil Procedure (Jan. 4, 2008).

6. For example, the Practising Law Institute’s annual conference, *Electronic Discovery and Retention Guidance for Corporate Counsel*, yields a lengthy symposium issue. See Symposium, *Electronic Discovery and Retention Guidance for Corporate Counsel 2007*, 766 PLI/LIT. 13 (2007) (17 articles); Symposium, *Electronic Discovery and Retention Guidance for Corporate Counsel 2006*, 747 PLI/LIT. 9 (2006) (13 articles); Symposium, *Electronic Discovery and Retention Guidance for Corporate Counsel 2005*, 733 PLI/LIT. 9 (2005) (14 articles).

7. An *ExpressO* search of top fifty law school specialty reviews found thirteen on women’s or gender issues, seven on constitutional or civil rights, and only one on litigation.

8. For a discussion on the need for discovery to prove hidden discriminatory intent circumstantially, see *supra* note 1 and accompanying text.

9. For a discussion of the proposals to limit discovery and cost-shifting proposals, see *infra* Parts I.B, II.A.

10. Patricia A. Bronte, *Managing Electronic Discovery Successfully in Insurance Coverage Litigation*, 758 PLI/LIT. 55, 75 (2007) (arguing for attorneys to “master the brave new world of electronic discovery”); Richard L. Marcus, *E-Discovery & Beyond: Toward Brave New World or 1984?*, 25 REV. LITIG. 633, 634 (2006).

11. Harvey L. Kaplan, *Electronic Discovery in the 21st Century: Is Help on the Way?*, 733 PLI/LIT. 65, 67 (2005).

12. See, e.g., *Siemens Solar Indus. v. Atl. Richfield Co.*, No. 93 Civ. 1126 (LAP), 1994 WL 86368, at *2 (S.D.N.Y. Mar. 16, 1994) (recounting the plaintiff’s discovery of e-mails “reveal[ing] beyond peradventure” that the defendant praised its new product yet knew it “was not commercially viable”).

incriminating files.¹³ In 2006, the Federal Rules of Civil Procedure clarified that “electronically stored information” (“ESI”) is discoverable unless it is “not reasonably accessible because of undue burden or cost.”¹⁴ The “reasonably accessible” limit reflects worry about e-discovery’s “enormous costs . . . becoming the single most expensive facet of litigation.”¹⁵ The cost of e-discovery has two key components: (1) quantity—with businesses exchanging 2.5 trillion e-mails annually, 2 million at a typical company,¹⁶ and with computer files often remaining recoverable after deletion,¹⁷ the amount of attorney time needed to review discovery, and the potential for discovery disputes, has increased; and (2) inaccessibility—digital data “can be expensive or virtually impossible to recover” due to “outmoded storage media and software, and dispersion of information.”¹⁸ Because e-discovery can cost tens or hundreds of thousands of dollars in even fairly typical cases,¹⁹ “[i]t is hard to overstate the importance and the degree of anxiety generated by electronic discovery It is not just in the world of big business; it is in the world of organizations generally, large data producers.”²⁰

Yet this brave new world may not be so new. Applying old litigation rules “to new technology presents additional challenges,”²¹ but “this is not the first time someone has argued that the discovery rules are no longer suitable for . . . contemporary discovery.”²² After all, when the federal rules were enacted, “[t]he photocopy machine

13. See Marnie H. Pulver, Note, *Electronic Media Discovery: The Economic Benefit of Pay-Per-View*, 21 CARDOZO L. REV. 1379, 1409–12 (2000) (recounting examples of these sorts of digital evidence).

14. FED. R. CIV. P. 26(b)(2)(B). For the details of these amendments, see *infra* note 74.

15. Bronte, *supra* note 10, at 59; see also Christopher S. Rugaber, *E-Documents Subject to Stricter Storage*, SUN-SENTINEL (Fort Lauderdale, Fla.), Dec. 2, 2006, at 3D (noting that companies pay e-discovery consultants over \$1 billion a year and must have lawyers review e-mail and “things more difficult to track, like digital photos . . . [on] cell phones and information on removable memory cards”).

16. Kaplan, *supra* note 11, at 67.

17. Pulver, *supra* note 13, at 1409–12.

18. William R. Maguire, *Current Developments in Federal Civil Discovery Practice: Setting Reasonable Limits in the Digital Era*, 754 PLI/LIT. 169, 175 (2007).

19. See *infra* notes 87–88.

20. Panel Discussion, *Managing Electronic Discovery: Views from the Judges*, 76 FORDHAM L. REV. 1, 4 (2007) (comments of Lee H. Rosenthal, J., United States District Court for the Southern District of Texas).

21. Richard L. Marcus, *Confronting the Future: Coping with Discovery of Electronic Material*, 64 LAW & CONTEMP. PROBS. 253, 258 (Spring/Summer 2001).

22. *Id.* at 254.

did not exist,”²³ and like all new technologies, it solved some problems but created new ones. As with the discovery-multiplying effect of computers, “the volume of paper created grew enormously when photocopies became ubiquitous.”²⁴ At the same time, as one court marveled in now-quaint terms, “the modern convenience of photocopying lessens . . . burdensome transportation of . . . documents,”²⁵ just as computers can make document exchange as easy as e-mailing attachments.²⁶

Because e-discovery is just another instance of an old problem—technology increases costs by permitting more discovery—just as photocopiers led to cost-benefit proportionality limits on discovery,²⁷ computers led to similar limits. The resulting e-discovery rules paralleled existing proportionality rules, asking courts to balance whether good cause justifies e-discovery that imposes a high “burden or cost.”²⁸ Such proportionality limits long have been prescribed by law and economics scholars—not just the economists most skeptical of litigation cost, runaway tort liability, and extortionate settlements, but also others proposing cost-benefit requirements that force parties to pay more of the cost of seeking heavy discovery.²⁹ With contrary views relegated to the periphery of the economic debate, discovery limits are a topic of unusual consensus among civil procedure scholars, economists, and the judicial and other governmental bodies that repeatedly undertake rule changes and other efforts to rein in discovery.³⁰

Dissenting from the consensus, this Article contends that proportionality limits cannot be implemented effectively. Sometimes

23. *Id.* at 266.

24. *Id.*

25. *Lemberger v. Westinghouse Elec. Corp.*, 76-C-552(JW), 1976 WL 834, at *5 (E.D.N.Y. Nov. 1976).

26. See Richard Allan Horning, *Electronically Stored Evidence: Answers to Some Recurring Questions Concerning Pretrial Discovery and Trial Usage*, 41 WASH. & LEE L. REV. 1335, 1344 (1984) (“[D]uplicat[ing] electronic data is relatively inexpensive . . . compared with the cost of re-inputting . . .”).

27. See FED. R. CIV. P. 26(b)(2)(C)(iii). For a discussion of this proportionality rule, see *infra* Part I.A.

28. FED. R. CIV. P. 26(b)(2)(B). For a discussion of this e-discovery rule, see *infra* Part I.A.

29. See *infra* Part II.A.

30. For a discussion of the proportionality amendments and reform efforts, including efforts by U.S. attorneys general, that preceded them, see *infra* notes 42–43. For a discussion of the Manual for Complex Litigation proposals published by Federal Judicial Center, see *infra* notes 49–51 and accompanying text. For a discussion of the e-discovery amendments, see *infra* notes 74–77 and accompanying text.

they do too little, failing to curb discovery excess or allowing costly discovery on meritless claims; other times they do too much, disallowing discovery that meritorious cases need. These opposing errors do not average out any more than it is on average comfortable to have one foot frozen and one foot on fire. Additionally, when product quality is hard to determine, bad products drive good ones out of the market;³¹ analogously, if bad cases get too much discovery and good cases too little, court dockets will have more bad cases and fewer good cases. Consequently, suboptimal discovery skews the mix of cases.

After noting how proportionality rules have disappointed at regulating discovery and how e-discovery rules seem headed for a similar fate, this Article parts company with those blaming bad rulemaking or judges too timid to curb discovery.³² Rather, the problem is that proportionality rules ask the impossible: judges must decide when discovery cost is proportional to some measure of “value” that includes both *evidence value* to jury deliberation and *case value* to the parties and society. This yields a fundamental information-timing problem: discovery disputes occur before parties marshal all the evidence, so how can courts measure the value of particular evidence, much less case merits? Like other arguments that litigation procedure rulings cannot truly be independent of case merits,³³ this Article notes how discovery has more probative value in close-call cases than in the strongest and weakest cases (in which more evidence is less likely to affect case outcome). Case merits, though critical to discovery decisions, typically remain hidden in a cloud of uncertainty during discovery because the court is not yet able to sift fully through the evidence and arguments.

This Article applies economics and game theory to analyze courts’ decisions on litigation discovery disputes. Due to the information costs (including time) of assessing case merit during discovery, courts often cannot tell which plaintiffs’ braggadocio is cheap talk and which reflects real case merit. As a result, courts must ignore parties’ merits arguments and adjudicate discovery disputes as if all cases of a similar type in the pool (that is, cases arising under the same statute that are neither facially frivolous nor obvious winners) warrant similar discovery. In game theory terms, courts’ discovery

31. See *infra* Part II.B.3.c.

32. See *infra* Part II.B.3.c.

33. See *infra* note 130 and accompanying text.

rulings should be based on claims' merit, but merit cannot be communicated effectively; as a result, those rulings must be based on the average value of all cases in the pool. In this pooling equilibrium, the best available strategy for courts is to rule the same on all cases in a pool regardless of case merit—even though these rulings are suboptimal in the sense of yielding too much discovery in low-merit cases and too little discovery in high-merit cases.

Under this analysis, the quest for better discovery limits has disappointed not because of bad decisionmaking or bad rulemaking, but because courts and parties are stuck in a pooling equilibrium. This is a fundamental information-timing problem inherent in the discovery stage of litigation: optimal evidence-gathering decisions require more merits analysis, but merits analyses require more evidence gathering. As in the folk song, “There’s a Hole in the Bucket”—in which the hole is fixable only with a machine requiring water poured from the broken bucket³⁴—the problem is a classic circularity; the problem prevents the solution.

Deferring close decisions on potentially useful but costly evidence until case merit is clearer—until meritorious cases distinguish themselves, turning a pooling equilibrium into a separating equilibrium—is one possible solution to the pooling equilibrium. Fortunately, litigation has such a point: after summary judgment. A case reaching trial, having survived summary judgment, has a reasonable probability of merit: even without adopting the old theory that tried cases have fifty-fifty odds, cases reaching trial are more likely than others to be close calls, and they certainly have higher average merit than the pool of all filed complaints. More evidence, like costly electronic data, has more value to the jury in close-call cases than in very weak or strong cases.

Accordingly, much of the scholarly debate on discovery misses the mark by focusing on *how much to limit* costly discovery, such as with proportionality rules and numerical caps.³⁵ This Article suggests focusing on *when* in litigation to allow costly discovery. Specifically,

34. Harry Belafonte and Odetta are probably the most famous pair to have sung this duet. HARRY BELAFONTE, *A MAN AND HIS MUSIC* (RCA Records 1990) (track 5). The author first heard it sung by muppets on *The Muppet Show* in the 1970s but, sadly, cannot find a video clip or proper citation.

35. See Marcus, *supra* note 21, at 256 (“The retrenchment effort . . . [has] two themes . . . the principle of proportionality and quantitative limits . . .”). For a discussion of rules applying proportionality and numerical limits, see *infra* notes 43, 122 and accompanying text.

decisions regarding costly discovery should be postponed until after summary judgment, to ensure that costly discovery is imposed only in cases with a greater probability of merit. Although this proposal might enable judges to deny or postpone more discovery, any discovery denied or postponed under this proposal is probably already denied based on judges' proportionality discretion. Thus, the main utility of this proposal would be to explain how courts could *allow more* discovery, only after summary judgment, of helpful but costly evidence *that courts often disallow* and declare nondiscoverable.

Nobody has suggested solving the dilemma of costly discovery with post-summary judgment discovery, which might seem to be a counterintuitive idea;³⁶ summary judgment typically comes only after all discovery is completed. Unusually costly evidence should be an exception: surviving summary judgment means a case likely is the sort of close call warranting more fact gathering, so courts should allow truly costly discovery, like the heavy e-discovery that they commonly disallow, only once a case survives summary judgment.³⁷

Interestingly, no rule change is required to implement this Article's proposal that courts revisit denials of burdensome discovery if a case survives summary judgment. Existing rules give courts broad case management authority, including power over the extent and sequence of discovery and summary judgment. Thus, this proposal could not only improve litigation discovery, but also provide a welcome answer to courts' riddle of how to rule on proportionality without circular, premature case-merit evaluations. A new rule would be advisable, though, to minimize the risk of courts misusing the proposal to deny discovery excessively.

This Article's conclusion then offers a brief broader point about economic analysis. Fitting into a line of scholarship analyzing litigation as a series of points in time when information emerges, this Article suggests that for economic analysis of litigation to provide accurate diagnoses and useful recommendations, it must do more

36. The exception is that courts allow prediscovery limited-scope summary judgment motions in certain cases, such as those limited to governmental immunity defenses. *See infra* note 199 and accompanying text.

37. This Article does not discuss *all* discovery reforms, just one problem that cannot be fully fixed (the impossibility of optimal discovery decisions) and a partial fix well targeted to that information-timing problem (postponing some discovery until summary judgment). For a discussion of other proposals, such as discovery sampling or cost shifting, see *infra* notes 218–24 and accompanying text.

than just prescribe cost-benefit comparisons; it must consider the timing-and-stages nature of litigation, such as by delving into the details of discovery, prelitigation settlement, and other events short of trials and dispositive motions.

I. THE RISE OF PROPORTIONALITY LIMITS ON LITIGATION DISCOVERY—AND THEIR DISAPPOINTING RESULTS

A. *What Is Old Is New Again: Proportionality Requirements as a Solution to Tech-Driven Discovery Excess*

The most surprising aspect of the e-discovery rules is how *unnew* those rules are. This is not the first time new rules have targeted discovery excesses that new technology facilitated. The federal rules long have prescribed “liberal discovery”³⁸ that the producing party must pay for itself³⁹ and a broad relevance standard—that discoverable evidence need not be admissible, only “reasonably calculated to lead to the discovery of admissible evidence.”⁴⁰ These principles arose in an older, prephotocopier era, when typical discovery was just on-premises review of original evidence.⁴¹

Like the computer revolution, the older photocopier revolution facilitated the higher levels of discovery that still exist—massive document demands and “paper dump” responses—prompting a powerful countermovement at the highest levels of the legal establishment, including the U.S. attorney general and an American

38. See, e.g., *Swierkiewicz v. Sorema N.A.*, 534 U.S. 506, 512 (2002) (disallowing the prediscovery dismissal of a discrimination claim because the federal rules’ “simplified notice pleading standard relies on liberal discovery rules and summary judgment motions . . . to dispose of unmeritorious claims”).

39. See *Oppenheimer Fund, Inc. v. Sanders*, 437 U.S. 340, 358 (1978) (“[T]he presumption is that the responding party must bear the expense of complying with discovery requests . . .”).

40. FED. R. CIV. P. 26(b)(1); see also *Oppenheimer Fund*, 437 U.S. at 351 (noting that discovery relevance “has been construed broadly to encompass any matter that bears on, or that reasonably could lead to other matter that could bear on, any issue that is or may be in the case”).

41. The early discovery decisions following the 1938 adoption of the Federal Rules of Civil Procedure illustrate this limited nature of early-to-mid-twentieth century discovery. See, e.g., *Harris v. Sunset Oil Co.*, 2 F.R.D. 93, 93 (W.D. Wash. 1941) (ordering documents produced “at [parties’] respective places of business during reasonable office hours”); *Compagnie Continentale D’Importation v. Pac. Argentine Braz. Line, Inc.*, 1 F.R.D. 388, 389 (S.D.N.Y. 1940) (“[Document] inspection should be held at defendant’s convenience . . . at its place of business.”); *Gielow v. Warner Bros. Pictures, Inc.*, 26 F. Supp. 425, 426 (S.D.N.Y. 1938) (ordering party “to exhibit the documents . . . but at the office of the complainant; not to be removed therefrom”).

Bar Association “Special Committee on Abuse of Discovery.”⁴² The federal judges on the Judicial Conference ultimately enacted the proportionality rule requiring that discovery

shall be limited by the court . . . [if] the burden or expense of the proposed discovery outweighs its likely benefit, considering the needs of the case, the amount in controversy, the parties’ resources, the importance of the issues at stake in the litigation, and the importance of the proposed discovery in resolving the issues.⁴³

The sense of urgency behind proportionality limits has not dissipated; the Supreme Court in 1998 “signalled the importance of the proportionality concept in some cases by quoting portions of Rule 26(b)(2) and observing that ‘[it] vests the trial judge with broad discretion to tailor discovery narrowly.’”⁴⁴

The timing of the earlier proportionality movement is striking: coming in the 1980s and 1990s, it shortly preceded the widespread adoption in the 1990s and 2000s of e-mail, networked computers, and the Internet. To be sure, although computers were less ubiquitous before the 1990s, major companies⁴⁵ and government entities⁴⁶ long have computerized their data. But courts often ignored cost-based objections to computer discovery, blaming computer-using parties for “a system of record-keeping which conceals rather than discloses relevant records, or makes it unduly difficult to identify or locate

42. See, e.g., Wayne D. Brazil, *The Adversary Character of Civil Discovery: A Critique and Proposals for Change*, 31 VAND. L. REV. 1295, 1332 (1978) (noting that the Judicial Conference and American Bar Association Committee had the “primary concern . . . [of] curbing the cost of discovery and reducing the ways discovery can be abused”); Richard L. Marcus, *Discovery Containment Redux*, 39 B.C. L. REV. 747, 753–68 (1998) (recounting drives to limit discovery); Stephen N. Subrin, *Fishing Expeditions Allowed: The Historical Background of the 1938 Federal Discovery Rules*, 39 B.C. L. REV. 691, 744–45 (1998) (discussing the proportionality amendments that later became part of Rule 26(b)(2)(C)(iii)).

43. FED. R. CIV. P. 26(b)(2)(C)(iii).

44. 8 CHARLES ALAN WRIGHT, ARTHUR R. MILLER, & RICHARD L. MARCUS, *FEDERAL PRACTICE & PROCEDURE: CIVIL* § 2008.1, at 54 (2d ed. Supp. 2008) (quoting *Crawford-El v. Britton*, 523 U.S. 574, 598 (1998)).

45. See, e.g., *Dunn v. Midwestern Indemn.*, 88 F.R.D. 191, 193, 196 (S.D. Ohio 1980) (granting the plaintiffs, who had claimed a widespread “pattern or practice” of discrimination, a search that could take thousands of hours of the defendant’s “computer systems, including access to and information about . . . equipment, raw data, programs, data management systems, and the by-products of their analyses”).

46. See, e.g., *United States v. Greenlee*, 380 F. Supp. 652, 658 (E.D. Pa. 1974) (denying a criminal defendant a requested weeks-long search of Internal Revenue Service computers that would have created risks of security breaches, privacy violations, and “serious interruption of the operations of the IRS”).

them” and thereby rejecting parties’ “impossibility contentions insofar as they are grounded in the peculiar manner in which [they] maintain their computer systems.”⁴⁷

When computer discovery was a rarity, courts could ignore its cost. Once it was common, courts faced the reality of tactical demands for costly searches. As in the 1970s, technology raised discovery cost and volume,⁴⁸ inspiring efforts to limit discovery. The Manual for Complex Litigation in 1995 recommended “a cost-benefit approach” to stop “fishing expedition[s],”⁴⁹ proposing “conditioning particular discovery on payment of its costs by the party seeking it”;⁵⁰ the Manual did not detail as broad a range of relevant factors as did later e-discovery proposals, but it did stress that with discovery costs increasing, courts should give renewed consideration to cost-shifting options that had not previously been in wide use.⁵¹

Yet not all courts have followed the Manual for Complex Litigation,⁵² and one later case quickly became the leading word on e-discovery. In 2003, *Zubulake v. UBS Warburg LLC (Zubulake I)*⁵³ set out a multifactor test to determine whether discovery should be allowed or denied, partially allowed, or allowed conditioned upon cost shifting.⁵⁴ The *Zubulake I* test essentially was a more detailed cost-benefit analysis comparing the costs and benefits of discovering the disputed evidence,⁵⁵ with cost evaluated in light of the parties’ resources, the amount in controversy, and the “relative ability of each

47. *Dunn*, 88 F.R.D. at 198.

48. *See supra* note 15 and accompanying text; *infra* notes 87–88.

49. Pulver, *supra* note 13, at 1386.

50. MANUAL FOR COMPLEX LITIGATION (THIRD) § 21.433 (1995).

51. *Id.*

52. *See, e.g., In re Brand Name Prescription Drugs Antitrust Litig.*, 94 C 897, 1995 U.S. Dist. LEXIS 8281, at *2–3 (N.D. Ill. June 13, 1995) (mem.) (granting class action plaintiffs’ motion to compel the defendant to produce computer-stored e-mail at the defendant’s own expense, estimated at \$50,000 to \$70,000, and expressly rejecting an alternative suggested by the Manual of Complex Litigation).

53. *Zubulake v. UBS Warburg LLC (Zubulake I)*, 217 F.R.D. 309 (S.D.N.Y. 2003). Judge Scheindlin, who authored *Zubulake I*, issued several later relevant opinions: *Zubulake v. UBS Warburg LLC (Zubulake II)*, 230 F.R.D. 290, 290 (S.D.N.Y. 2003); *Zubulake v. UBS Warburg LLC (Zubulake III)*, 216 F.R.D. 280, 281 (S.D.N.Y. 2003); *Zubulake v. UBS Warburg LLC (Zubulake IV)*, 220 F.R.D. 212, 214 (S.D.N.Y. 2003); and *Zubulake v. UBS Warburg LLC (Zubulake V)*, 229 F.R.D. 422, 424 (S.D.N.Y. 2004).

54. *Zubulake I*, 217 F.R.D. at 322.

55. *Id.* Factors focusing on the benefit of discovering the benefit include the following: “The extent to which the request is specifically tailored to discover relevant information;” “The availability of such information from other sources;” “The importance of the issues at stake in the litigation;” and “The relative benefits to the parties of obtaining the information.” *Id.*

party to control costs.”⁵⁶ *Zubulake I* never fully harmonized e-discovery law and practice, however. Not qualifying as “final” orders, discovery rulings ordinarily are nonappealable,⁵⁷ so “few trial court decisions regarding the scope and logistics of discovery . . . [reach] the appellate level.”⁵⁸ Because *Zubulake I* and virtually all e-discovery opinions are nonbinding district court precedent, *Zubulake I*, though “widely regarded as the leading case authority” on e-discovery,⁵⁹ has not drawn universal adherence.⁶⁰

Also helpful are the e-discovery reform proposals of the nongovernmental Sedona Conference Working Group on Best Practices for Electronic Document Production.⁶¹ Sedona paralleled the Manual for Complex Litigation and *Zubulake I* on some points, such as the permissibility of cost shifting,⁶² but it added other

56. *Id.*

57. 28 U.S.C. § 1291 (2006) (authorizing appeal only of “final decisions” in cases); *see also*, e.g., *Church of Scientology of Cal. v. United States*, 506 U.S. 9, 18 n.11 (1992) (“As a general rule, a district court’s order enforcing a discovery request is not a ‘final order’ subject to appellate review. A party that seeks to present an objection to a discovery order immediately to a court of appeals must refuse compliance, be held in contempt, and then appeal the contempt order.”).

58. Shira A. Scheindlin & Jeffrey Rabkin, *Electronic Discovery in Federal Civil Litigation: Is Rule 34 up to the Task?*, 41 B.C. L. REV. 327, 378 (2000).

59. Benjamin D. Silbert, Note, *The 2006 Amendments to the Rules of Civil Procedure: Accessible and Inaccessible Electronic Information Storage Devices, Why Parties Should Store Electronic Information in Accessible Formats*, 13 RICH. J.L. & TECH. 3, ¶ 18 (2007), <http://law.richmond.edu/jolt/v13i3/article14.pdf>.

60. *See, e.g.*, *Multitechnology Servs., L.P. v. Verizon Sw., No. Civ.A. 4:02-CV-702-Y*, 2004 WL 1553480, at *1 (N.D. Tex. July 12, 2004) (ordering cost shifting to requesting party even though *Zubulake I* “weighs against shifting any expense” because “*Zubulake* is a district court opinion without binding authority”); Panel Discussion, *supra* note 20, at 24 (comments of James C. Francis IV, J., United States District Court for the Southern District of New York) (discussing use of multifactor tests in e-discovery disputes and arguing that “it depends on whether you adopt Judge Scheindlin’s view [in *Zubulake*] of a hierarchy or whether you think . . . [the] factors will probably play out differently in different cases. I am resistant to the hierarchy approach because my fear is that the factor at the top of the hierarchy will almost always wash out the other[s]”).

61. WORKING GROUP ON ELEC. DOCUMENT RETENTION & PROD., SEDONA CONFERENCE, THE SEDONA PRINCIPLES: BEST PRACTICES RECOMMENDATIONS & PRINCIPLES FOR ADDRESSING ELECTRONIC DOCUMENT PRODUCTION (2d ed. 2007) [hereinafter SEDONA], available at http://www.thesedonaconference.org/content/miscFiles/TSC_PRINCP_2nd_ed_607.pdf. Sedona produced a draft in 2003, revised it after public comment in 2004, updated it minimally in 2005, and produced a 2007 Second Edition incorporating the 2006 e-discovery amendments to the Federal Rules. *See* THOMAS Y. ALLMAN, THE SEDONA PRINCIPLES AFTER THE FEDERAL AMENDMENTS: THE SECOND EDITION (2007), at 1 n.3, 2–3 (2007), <http://www.thesedonaconference.org/content/miscFiles/2007SummaryofSedonaPrinciples2ndEditionAug17assentforWG1.pdf>.

62. SEDONA, *supra* note 61, at 67.

proposals focused on the technical details of electronic data. For example, Sedona distinguished certain data types as presumptively discoverable or not, like “active data” (ordinary files) and backup or deleted “legacy data” accessible only at high cost.⁶³ It also proposed certain best practices for e-discovery, most notably that “[d]iscovery requests for electronically stored information . . . be as clear as possible.”⁶⁴ *Zubulake I* cited Sedona’s early work,⁶⁵ and a subsequent report authored by a Sedona editor notes that courts have cited its earlier work, “helping provide *de facto* ‘national standards.’”⁶⁶ Yet courts have not adopted Sedona wholesale; one topic Sedona addressed heavily, preservation of data, was “relegated . . . to evolving case law” because the e-discovery rules authors would not enact preservation standards.⁶⁷

Moreover, *Zubulake I* and Sedona, although thoughtful and useful guides to e-discovery, promise limited impact. They do not aim for a paradigm shift, instead relying on status-quo methods to limit discovery. Like *Zubulake I*, Sedona accepted as its touchstone the old “‘proportionality’ standard” of assessing costs “in light of the nature of the litigation and the amount in controversy.”⁶⁸ Sedona rejected other ideas, like Texas’s mandatory cost shifting (making requesting parties pay for e-discovery not “reasonably available” in the “ordinary course” of business⁶⁹) in favor of a merely permissive suggestion that cost “‘may’ (instead of ‘should’)” be shifted.⁷⁰ *Zubulake I* likewise said courts should order cost shifting only for “relatively inaccessible [data], such as in backup tapes” that are costly to recover,⁷¹ and “close calls should be resolved in favor of the

63. “Active data is typically stored on local hard drives, networked servers, and distributed devices or offline archival sources from which information can be accessed without a special restoration effort.” ALLMAN, *supra* note 61, at 6–7. “Information stored solely for disaster-recovery purposes, ‘legacy’ data retained in obsolete systems, and deleted or fragmentary information that can be restored only through extraordinary efforts” ordinarily are “unduly burdensome and costly to access.” *Id.* at 5.

64. *Id.* at 25.

65. *E.g.*, *Zubulake v. UBS Warburg LLC (Zubulake I)*, 217 F.R.D. 309, 320 n.61, 321 n.67 (S.D.N.Y. 2003).

66. ALLMAN, *supra* note 61, at 2.

67. *Id.*

68. *Id.* at 5.

69. TEX. R. CIV. P. 196.4.

70. ALLMAN, *supra* note 61, at 9.

71. *Zubulake v. UBS Warburg LLC (Zubulake I)*, 217 F.R.D. 309, 324 (S.D.N.Y. 2003).

presumption” against cost shifting,⁷² with cost shifting denied if evidence is sufficiently useful.⁷³

The 2006 Federal Rules amendments⁷⁴ followed at least the spirit of *Zubulake I* and *Sedona* in stressing cost-benefit proportionality limits⁷⁵: when digital or electronically stored information is “not reasonably accessible because of undue burden or cost . . . the court may nonetheless order discovery . . . if the requesting party shows good cause, considering the limitations of Rule 26(b)(2)(C),”⁷⁶ the proportionality rule. The Advisory Committee’s notes to these new rules built on *Zubulake I* and other case law by prescribing an essentially similar cost-benefit analysis instructing courts to look to various factors relevant to the likely benefit and cost of a disputed discovery request.⁷⁷

72. *Id.* at 320.

73. *See id.* at 322 (noting factors such as the “extent to which the request is specifically tailored to discover relevant information . . . [and] [t]he relative benefits to the parties of obtaining the information”); *see also* AAB Joint Venture v. United States, 75 Fed. Cl. 432, 444 (2007) (“Defendant shall bear the costs of restoration of the initial sample of back-up tapes and screening the sample to identify responsive documents. The parties will then have an opportunity to argue . . . [whether] additional restoration of back-up tapes is likely to lead to production of relevant evidence and consequently who should bear the cost . . .”).

74. Three rules primarily control e-discovery. FED. R. CIV. P. 26(a)(1)(A)(ii) (requiring initial disclosures to give locations of documents and “electronically stored information” (ESI)); *id.* 34(a)(1)(A) (deeming ESI part of “document” demands and allowing ESI testing or sampling); *id.* 34(b) (allowing parties to “specify the form . . . in which electronically stored information is to be produced,” with default rule that ESI be produced as “ordinarily maintained” or in form “reasonably usable”); *id.* 45 (allowing ESI discovery from nonparties).

75. *See* Panel Discussion, *supra* note 20, at 9–10 (comments of Lee H. Rosenthal, J., United States District Court for the Southern District of Texas) (explaining that it is “clear that the key is proportionality” in the new e-discovery rule, because its “good cause determination must be based on the proportionality limits that [already] have been in the rules”).

76. FED. R. CIV. P. 26(b)(2)(B).

77. *Id.* 26 advisory committee’s note. According to the Advisory Committee,

Appropriate considerations may include: (1) the specificity of the discovery request; (2) the quantity of information available from other and more easily accessed sources; (3) the failure to produce relevant information that seems likely to have existed but is no longer available on more easily accessed sources; (4) the likelihood of finding relevant, responsive information that cannot be obtained from other, more easily accessed sources; (5) predictions as to the importance and usefulness of the further information; (6) the importance of the issues at stake in the litigation; and (7) the parties’ resources.

Id.

B. The E-Discovery Rules: Just More Disappointingly Conventional and Ineffective Proportionality Limits

With the 2006 e-discovery rules following earlier proposals by the Manual for Complex Litigation, *Zubulake I* and other district courts, the Sedona Group, and academic writings,⁷⁸ the rules evolved in a highly decentralized fashion. Decentralization, though chaotic, often has the virtue of yielding a flourish of creative, varied independent efforts. Oddly, the major proposals to limit technology-inspired discovery proliferation since the 1970s have been similar, even derivative, cost-benefit proportionality rules; at best, these proposals contain a mild, permissive suggestion of cost shifting in the rarest, most costly discovery situations.⁷⁹

But the main problem with proportionality limits on discovery is not that they are old news. Rather, the problem is that such limits never have worked terribly well and appear unlikely to work well for e-discovery. Although the idea of proportionality has gained momentum for decades, led by powerful forces in the judiciary and the bar, proportionality is widely criticized as having been ineffective at convincing judges to rein in discovery excess. “Whatever the theoretical possibilities,” the proportionality rule “created only a ripple in the caselaw,” a leading civil procedure treatise notes; “no radical shift has occurred.”⁸⁰

Even if the proportionality rule did not yield a major shift, courts do deny discovery for cost reasons. To be sure, some plaintiffs get very broad discovery; *Kozlowski v. Sears, Roebuck & Co.*,⁸¹ for example, allowed discovery of all prior defective-product complaints over pleas that the discovery would require a costly search of voluminous records.⁸² The court explained that “most courts have

78. See, e.g., Horning, *supra* note 26, at 1344 (proposing that courts more often require production of digital rather than paper data, as well as that courts more often allow requesting parties to require that producing parties put that data into specific forms and facilitate interpretation of complex data); Martin H. Redish, *Electronic Discovery and the Litigation Matrix*, 51 DUKE L.J. 561, 615–18 (2001) (proposing more cost shifting to parties requesting costly e-discovery); Pulver, *supra* note 13, at 1386 (same).

79. See *supra* notes 42–44 and accompanying text.

80. WRIGHT ET AL., *supra* note 44, § 2008.1, at 121; see also Ronald J. Hedges, *A View from the Bench and the Trenches: A Critical Appraisal of Some Proposed Amendments to the Federal Rules of Civil Procedure*, 227 F.R.D. 123, 127 (2005) (“[T]he proportionality principle of Rule 26(b)(2) . . . is not being utilized by judges . . .”).

81. *Kozlowski v. Sears, Roebuck & Co.*, 73 F.R.D. 73 (D. Mass. 1976).

82. *Id.* at 76–77.

held that the existence and nature of other complaints in product liability cases is a proper subject for pre-trial discovery,” and it would be unfair “[t]o allow a defendant whose business generates massive records to frustrate discovery by creating an inadequate filing system, and then claiming undue burden.”⁸³

But however persuasive *Kozlowski*'s logic, its permissiveness with costly discovery is the exception, not the rule. Courts *often* disallow discovery on relevant matters for reasons of cost, burden, and inconvenience, such as in decisions denying plaintiffs' requests for discovery on similar instances of misconduct in claims of discrimination⁸⁴ or other wrongdoing;⁸⁵ courts similarly have rejected defendants' efforts to discover information a government agency used to enact a disputed regulation.⁸⁶ As to e-discovery in particular, the

83. *Id.* at 75–76.

84. *See, e.g.,* Sallis v. Univ. of Minn., 408 F.3d 470, 477–78 (8th Cir. 2005) (upholding, in claims of racially discriminatory failure to promote and hostile work environment, a ruling that “discovery must be limited, in both its temporal and geographical reach,” to just complaints in plaintiff's department for one year, and that plaintiff could not obtain discovery of complaints at all departments, despite the plaintiff's argument that complaints “were contained in an easily accessible, central database, and he experienced discrimination at the hands of other [university] departments,” because the request was “unduly burdensome”); Balderston v. Fairbanks Morse Engine Div. of Coltec Indus., 328 F.3d 309, 317, 319, 320 (7th Cir. 2003) (upholding, in claim of “systematic elimination of older employees”, denial of discovery of statistical data, including records of all employees defendant terminated over nine years, because of courts' “substantial discretion to curtail the expense and intrusiveness of discovery” in limiting . . . broad discovery of personnel files” (quoting Gehring v. Case Corp., 43 F.3d 340, 342 (7th Cir. 1994))); EEOC v. D.C. Pub. Sch., 217 F.R.D. 12, 13, 15 (D.D.C. 2003) (denying, in a claim of age discrimination in the termination of a teacher during a reduction in force, a request for “teaching disciplines of each teacher . . . [in the] academic year” when the plaintiff was terminated, when data was “perhaps retrievable [only] from a search of every personnel file,” which “would be oppressive”); Lee v. Executive Airlines, Inc., 31 F. Supp. 2d 1355, 1356 (S.D. Fla. 1998) (denying a race discrimination plaintiff's requests as to all employees “disciplined but not terminated for . . . time card infractions” (plaintiff's alleged offense) over five years, when discovery “would require extensive searches of files outside of the locations” plaintiff worked); Aramburu v. Boeing Co., 885 F. Supp. 1434, 1442–44 (D. Kan. 1995) (denying a race and disability discrimination plaintiff discovery that would take 240 hours to procure 1,500 personnel files, because the “plaintiff's need for the information” was “disproportionate” to the burden).

85. *See, e.g.,* Kowalski v. Stewart, 220 F.R.D. 599, 602 (D. Ariz. 2004) (denying a motion to compel discovery in a prisoner's claim of denial of court-ordered medical care because of the burden of “photocopying, organizing, and taking adequate measures to ensure prisoner confidentiality for the previous thirteen years of prisoner complaints”); Green Constr. Co. v. Kan. Power & Light Co., 732 F. Supp. 1550, 1554 (D. Kan. 1990) (denying a discovery request that would have required examining “nearly 62,400 bond claims” because the discovery's relevance was outweighed by the burden of examining the bond claims, which lacked any “index[ing] or filing code system”).

86. *See, e.g.,* United States v. Duke Energy Corp., 214 F.R.D. 392, 393 (M.D.N.C. 2003) (denying the defendant, an energy company challenging an Environmental Protection Agency

limited case law has been mixed, but given the level of angst about costly e-discovery underlying these rules, they may be disappointing to their advocates. Whatever the mix of decisions allowing⁸⁷ and disallowing⁸⁸ costly e-discovery, however, a more fundamental problem remains with both the original proportionality rule and the e-discovery rules.

Although denying relevant discovery due to cost may be defensible pragmatically, it is an unsatisfying concession that litigation accuracy inevitably is limited due to the cost of finding and analyzing evidence needed for accurate verdicts or settlements. Less accuracy is troubling not only morally but economically. Failing to impose liability on the guilty because evidence of guilt is too costly insufficiently deters misconduct and insufficiently assures that parties

(EPA) air regulation, discovery from another federal agency whose “personnel may have been present when some decisions were made by the EPA,” because although “statements and positions taken by any EPA employee are relevant,” the “burden to the [plaintiff] far outweighs the relevance”); Wyoming v. U.S. Dep’t of Agric., 208 F.R.D. 449, 454 (D.D.C. 2002) (denying the plaintiff, in a suit against a federal agency for violating rules on issuing regulations, an order for nonparty witnesses to produce documents regarding those regulations, when the court saw the request as expensive and unduly burdensome).

87. Some cases allow costly e-discovery when justified by high case stakes. *See, e.g.*, PSEG Power NY, Inc. v. Alberici Constructors, Inc., No. 1:05-CV-657, 2007 WL 2687670, at *1, *9–10 (N.D.N.Y. Sept. 7, 2007) (mem.) (ordering the plaintiff, in a \$4.4 million construction contract claim, at cost the of \$40,000 to \$200,000, “to produce all electronically stored emails, numbering approximately 3000, conjunctively with their corresponding attachments as ‘married’ documents”).

Other cases allow costly e-discovery despite modest case stakes when the information appears valuable. For example, in *W.E. Aubuchon Co. v. BeneFirst, LLC*, 245 F.R.D. 38 (D. Mass. 2007), regarding a claim that an employee benefit administrator breached its fiduciary duty, the court found that the data sought—thousands of employee claims stored electronically as unindexed images—were “not reasonably accessible,” *id.* at 43, when it could cost \$80,000 and 4,000 hours to recover 34,000 requested claim forms and medical bills, though plaintiff then narrowed its request to 3,000 claims, *id.* at 41, 44. Yet even though the discovery was burdensome and the “importance of the issues at stake” was low, the court allowed the discovery because the information was “clearly an integral part of the litigation . . . not only to BeneFirst’s culpability, but also to the amount of damages.” *Id.* at 44.

88. Courts also have denied, or denied unless plaintiffs paid the bulk of the cost, seemingly high-relevance e-discovery, *see, e.g.*, *Wiginton v. CB Richard Ellis, Inc.*, 229 F.R.D. 568, 577 (N.D. Ill. 2004) (requiring class action harassment plaintiffs to pay 75 percent of a \$249,000 e-mail search for known pornographic and other harassing e-mails), and have denied costly discovery even in high-stakes litigation in which the request seemed insufficiently essential, *see, e.g.*, *Best Buy Stores, L.P. v. Developers Diversified Realty Corp.*, 247 F.R.D. 567, 569–72 (D. Minn. 2007) (denying, in a claim that landlords caused actual damages of \$800,000, “enhanced damages” for fraud, and “long-term economic impact” on parties’ relationship, defendants’ request for plaintiff’s database on other landlords’ lease charges, because data was not in searchable format and required restoration, costing \$124,000 plus \$27,823 per month, and defendant could compile the data from paper discovery).

internalize costs (such as the costs of pollution) they impose on others.⁸⁹ Conversely, imposing liability on the innocent because exculpatory evidence was too costly yields ill-targeted deterrence of innocent activity; imposing pollution liability on a nonpolluting business just disincentivizes that socially useful commerce.⁹⁰

Thus, proportionality rules can be criticized equally for allowing entirely opposite errors, both false negatives (failing to detect and halt discovery abuse) and false positives (finding disproportionate some costly discovery that actually is justified by high evidentiary value and case merit). Erroneous pro-plaintiff rulings unjustifiably increase litigation costs and pressure defendants to settle unmeritorious cases; conversely, erroneous pro-defendant rulings deny plaintiffs the ability to press meritorious claims successfully.⁹¹

If the e-discovery rules are likely to yield the sort of uninspiring results seen after the original proportionality rule that so much of the legal establishment demanded, the question becomes how powerful forces attempting to respond to a hugely costly phenomenon have proven so impotent for so long?

89. As to the economic value of litigation accuracy, see generally Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307 (1994); Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399 (1973).

90. Posner, *supra* note 89, at 402–06, 410–15. Although discovery is most commonly analyzed as a way plaintiffs get information from defendants, the opposite can be true as well. Defendants can destroy plaintiffs' claims by pressing them for the details of, and facts supporting, their allegations; defendants also often press defenses that are based on the plaintiff's conduct (and thus that require the defendant to seek discovery from the plaintiff), such as contributory negligence in tort cases, mitigation of damages in contract cases, and evidence of other misconduct that would have justified a challenged firing in employment discrimination cases.

91. Cf. Robert G. Bone & David S. Evans, *Class Certification and the Substantive Merits*, 51 DUKE L.J. 1251, 1287 (2002). As Professor Bone and Dr. Evans noted as to rulings on class action certification,

Judges make mistakes. They grant certification when it should be denied, and they deny certification when it should be granted. . . . An erroneous grant creates unnecessary administrative and litigation costs and . . . unjustified settlements. An erroneous denial adds to plaintiffs' litigation costs and can make it harder for plaintiffs to recover.

Id.

II. AN ECONOMIC ANALYSIS OF WHY PROPORTIONALITY LIMITS, THOUGH POPULAR, CANNOT BE OPTIMAL

A. *The Consensus: Limit Discovery Based on Cost-Benefit Proportionality Principles*

For a field featuring so much controversy, discovery has featured an odd degree of consensus among analysts in disparate fields. Chicago School economists like Professor Richard Epstein, skeptical of whether much litigation is worth the potential for high cost and abuse, blame malleable balancing tests generally, including “underregulated” discovery, for the “inexorable expansion of [tort] liability.”⁹² Judge Frank Easterbrook, fretting about “impositional (excessive, abusive)” discovery that induces settlement by imposing high costs on defendants,⁹³ advocates in part “limit[ing] discovery to matters admissible at trial”⁹⁴—a drastic change from allowing any discovery “relevant to any party’s claim or defense,” including material that is *inadmissible* but “reasonably calculated to lead to . . . admissible evidence.”⁹⁵ Surprisingly, many non-Chicago School economists have similar qualms about discovery. Professors Robert Cooter and Daniel Rubinfeld propose that after a “reasonable” amount of discovery in a case, the cost of responding to discovery requests should shift to the requesting party.⁹⁶ Though disagreeing with that proposal, Professor Bruce Hay notes how heavy discovery, by scaring defendants into settling early, can counterproductively lead to *less*, not more, disclosure of illegality.⁹⁷

There are dissenting voices criticizing the drive to narrow discovery, often with arguments that limiting discovery favors defendants over plaintiffs or that drives to limit discovery are

92. Richard A. Epstein, *The Risks of Risk/Utility*, 48 OHIO ST. L.J. 469, 469, 476 (1987).

93. Frank H. Easterbrook, Comment, *Discovery as Abuse*, 69 B.U. L. REV. 635, 637, 644 n.26 (1989); see also Richard A. Epstein, *Was New York Times v. Sullivan Wrong?*, 53 U. CHI. L. REV. 782, 809 (1986) (criticizing the “ability of a well-heeled or determined plaintiff to hound a defendant in discovery”).

94. Easterbrook, *supra* note 93, at 644 n.26.

95. FED. R. CIV. P. 26(b)(1).

96. Robert D. Cooter & Daniel L. Rubinfeld, *An Economic Model of Legal Discovery*, 23 J. LEGAL STUD. 435, 459 (1994); Robert D. Cooter & Daniel L. Rubinfeld, *Reforming the New Discovery Rules*, 84 GEO. L.J. 61, 69–76 (1995). For other cost-shifting proposals, see *supra* note 78.

97. Bruce L. Hay, *Civil Discovery: Its Effects and Optimal Scope*, 23 J. LEGAL STUD. 481, 510–14 (1994).

premised on exaggerated fears of discovery abuse by plaintiffs.⁹⁸ Relatedly, there have been empirical findings that discovery excess may be confined to exceptional cases.⁹⁹ But those voices have been on the periphery of the consensus in favor of more proportionality-based limits on discovery—a consensus spanning (as this Section discusses) civil proceduralists, economists, and the judges who have enacted rule changes imposing new discovery limits since the 1970s.

Proportionality limits can be optimal, though, only if courts can perform the needed economic cost-benefit analyses passably well. What this Article seeks to add to the scholarship on the proportionality rule is that courts cannot undertake the needed analyses well—which means that discovery limits are doomed to be suboptimal.

B. An Economic View of Discovery: Helping Factfinders Assess Case Value and Merit (L and p)—Which Makes Discovery Decisionmaking Circular

The purpose of broad discovery, in economic terms, is well established: “[a] full exchange of the information . . . enabl[es] each party to form a more accurate, and generally therefore a more convergent, estimate of the likely [case] outcome.”¹⁰⁰ The case law uses similar logic to justify broad discovery.¹⁰¹ This Article focuses not only on how discovery helps parties assess cases, but also on how *judges* decide discovery disputes—a matter more rarely analyzed in

98. See, e.g., Linda S. Mullenix, *Discovery in Disarray: The Pervasive Myth of Pervasive Discovery Abuse and the Consequences for Unfounded Rulemaking*, 46 STAN. L. REV. 1393, 1396 (1994) [hereinafter Mullenix, *Discovery in Disarray*]; Linda S. Mullenix, *The Pervasive Myth of Pervasive Discovery Abuse: The Sequel*, 39 B.C. L. REV. 683, 685 (1998); Jeffrey W. Stempel, *Ulysses Tied to the Generic Whipping Post: The Continuing Odyssey of Discovery “Reform,”* 64 LAW & CONTEMP. PROBS. 197, 252 (Spring/Summer 2001) (arguing against 1990s-era discovery rules amendments “constricting the availability of information and continuing the late twentieth-century evolution favoring defendants over plaintiffs”).

99. See Willging et al., *supra* note 4, at 547 n.34 (noting that discovery cost is unusually high in the costliest 5 percent of cases); see also Mullenix, *Discovery in Disarray*, *supra* note 98, at 1397, 1432–42 (reviewing a number of empirical studies and finding “a surprisingly low incidence of discovery in federal civil litigation”).

100. RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 571 (6th ed. 2003); see also ROBERT G. BONE, *CIVIL PROCEDURE: THE ECONOMICS OF CIVIL PROCEDURE* 203 (2003) (characterizing discovery similarly).

101. See, e.g., *Hickman v. Taylor*, 329 U.S. 495, 507 (1947) (“Mutual knowledge of all the relevant facts gathered by both parties is essential to proper litigation.”).

economic terms.¹⁰² Such judicial decisionmaking draws little academic attention, probably because almost all discovery decisions are unappealable district court decisions¹⁰³ and because only experienced litigators recognize that if cases are won and lost in discovery, they really are won and lost *in litigating discovery motions*, dueling motions to compel discovery¹⁰⁴ and for protective orders.¹⁰⁵

A court must undertake a cost-benefit analysis to decide, as the rules require, whether the value of particular discovery is proportional to its cost (both dollar cost and nonpecuniary burdens). It must assess the cost of finding and producing the evidence, and it must compare that cost to the benefit of having that evidence. Assessing cost often is feasible; parties litigating discovery regularly detail the time and dollar costs of producing disputed evidence.¹⁰⁶

Assessing the *benefit* of particular discovery is the tricky part. As this Section discusses, for truly accurate judicial decisionmaking, a court must consider not only the probative value of the particular evidence and the size of the case (as the rules command), but also—contrary to the conventional wisdom on discovery decisionmaking—the likelihood that the case is meritorious, that the plaintiff will prevail at trial. In economic terms, the court’s proportionality determination bases on the following three variables:

- L , the size of the case, typically the amount in controversy but also possibly the value of nonmonetary relief;
- p , the probability that the plaintiff will win if the case goes to trial; and
- Δp , the probative value of the evidence (the difference the disputed evidence makes to p).

Yet, as this Section discusses, each of these three variables can be difficult or impossible for courts to assess during discovery. This difficulty is why this Article offers a diagnosis of pessimism about

102. For one article that does undertake just such an analysis of a range of procedural matters, including but not limited to discovery decisions, see Robert G. Bone, *Who Decides?: A Critical Look at Procedural Discretion*, 28 CARDOZO L. REV. 1961, 1996–2000 (2007). Bone analyzes procedural decisions based on a view of the judge as not an “umpire” but “a player in the game” whose decisions interact with those of the parties’ to affect the course of litigation. *Id.*

103. See *supra* note 57 and accompanying text.

104. See FED. R. CIV. P. 37(a).

105. See *id.* 26(c).

106. For a collection of cases in which parties detailed and argued before the court the costs and burdens of responding to discovery requests, see *supra* notes 84–88.

courts' ability to make accurate proportionality decisions on discovery disputes.

1. *Size of Case (L): Difficult to Determine in Many Cases.* Under the proportionality rule, in assessing the likely benefit of discovery, courts should take into account "the amount in controversy . . . [and] importance of the issues at stake."¹⁰⁷ This makes economic sense; discovery offers less benefit in low-value cases. More evidence-gathering expense is justified in a case that might result in millions of dollars changing hands; it is harder to justify similar evidence-gathering expense in a small-claims dispute.¹⁰⁸

The "amount in controversy" (dollar value) and "importance of the issues" (nonmonetary value), however, can be uncertain until trial; plaintiffs often press claims that present a strong argument for some amount of damages (such as recovery of out-of-pocket losses) with a weaker claim for additional damages (such as punitive damages).¹⁰⁹ Worse, case value can be subjective;¹¹⁰ what is the value of an injunction stopping seal clubbing, sexual harassment, or other illegality? Courts do assign damages awards for complex, nonpecuniary harms like torture.¹¹¹ Still, it remains wholly subjective whether evidence that might help win an injunction stopping seal clubbing is proportional to a month-long, million-dollar data search.

With case value often subjective, one problem with proportionality is "finding principled criteria for differentiating

107. FED. R. CIV. P. 26(b)(2)(C)(iii).

108. One caveat: If there are *many similar* low-value claims, then discovery in any one such case might be quite valuable, even if the claims are not aggregated into a class action, so long as the information disclosure in one case yields benefits for the others similarly situated, either by reducing other litigants' discovery costs or by disclosing illegality before it occurs (and thereby saving not only litigation cost, but also the cost of the illegality).

109. See *Davis v. Ross*, 107 F.R.D. 326, 327, 330 (S.D.N.Y. 1985) ("[P]lausible claims for punitive damages can easily be made in many actions . . . [T]he amount of damages will always be in issue; plaintiff seeks one million dollars in compensatory damages, and evidence must be introduced to demonstrate that the award should be more than nominal." (emphasis omitted)).

110. See, e.g., *Lyons v. Mobil Oil Corp.*, 554 F. Supp. 199, 201 (D. Conn. 1982) (holding that the prevailing party wins attorneys' fees unless it obtains only nominal damages because "[i]njunctive relief is an important part of the [statutory] scheme . . . regardless of whether . . . damages are awarded"); Robert G. Bone, *The Process of Making Process: Court Rulemaking, Democratic Legitimacy, and Procedural Efficacy*, 87 GEO. L.J. 887, 912 (1999) ("Assigning values to substantive interests is both difficult and controversial. Reasonable people disagree, for example, about the relative importance of the different interests protected by the Constitution.").

111. See, e.g., *Hilao v. Marcos*, 103 F.3d 767, 787 (9th Cir. 1996) (awarding damages in a human rights class action).

between various types of cases”; “[w]here . . . are judges expected to find the criteria and analytical structure for making such judgments” as whether more discovery is warranted on a high-dollar contract claim than on a low-dollar discrimination claim for mainly injunctive relief?¹¹² These problems can be mitigated with rules of thumb as to “the amount of discovery normally permissible in certain types of cases”¹¹³:

[A] search for discriminatory intent in a civil rights case may be seen as involving constitutional values . . . [and] broader discovery . . . than in a personal injury or commercial case. On the other hand, judicial experience indicating that in certain civil rights cases . . . further discovery is unlikely to shed additional light . . . might lead a judge to place limits . . .¹¹⁴

Although “patterns of appropriate discovery . . . may emerge which can normally be followed unless the particular facts warrant otherwise,”¹¹⁵ “[o]bviously these judgments will not be easy.”¹¹⁶ Nor will they be optimal; even the best rule of thumb treats cases similarly despite relevant differences between them.¹¹⁷

This difficulty estimating L is not the main topic of this Article’s analysis, but the partial solution, assuming the same average value for all cases in the same pool, returns in this Section as a similar imperfect solution to the problem of estimating p , the probability that the plaintiff would win at trial.

112. Edward F. Sherman & Stephen O. Kinnard, *Federal Court Discovery in the 80’s—Making the Rules Work*, 95 F.R.D. 245, 276 (1982).

113. *Id.* at 279.

114. *Id.*

115. *Id.*

116. *Id.*

117. I should not overstate the point; the relevant rules of thumb could be made as accurate as possible by refining them for various subsets of cases. *Cf.* Bone, *supra* note 102, at 1996 (noting, as to various procedural matters for which judges might lack the information necessary for accurate rulings, that “it is possible to find criteria, such as type of claim, amount in controversy, number of parties, and so on, to sort different case types with reasonable clarity and efficiency”). For example, rather than just say that employment discrimination cases are fact intensive and thus warrant more discovery, the relevant rule of thumb could allow more discovery in *certain kinds* of discrimination cases, such as incumbent employees’ termination and promotion claims (which typically depend on detailed evaluations of years of employee performance), but not rejected applicants’ claims (in which all the employer knew was the applicant’s interview and paper application, not years of performance). Still, even the best rule of thumb is just the best probabilistic generalization, one that yields suboptimal results in nonconforming cases.

2. *Probative Value of the Evidence Sought (Δp): Difficult to Assess before Fully Analyzing That and Other Evidence.* The most important consideration in a discovery dispute is the probative value of the evidence— Δp , the difference (Δ) the evidence makes to the probability (p) the plaintiff will win at trial. The proportionality rule asks courts to assess the likely benefit of discovery, taking into account “the needs of the case...and the importance of the discovery in resolving the issues,” which are aspects of probative value.¹¹⁸ Complexity of the issues,¹¹⁹ another proportionality factor, also goes to probative value: in a simple case (that is, “did the defendant sign the contract?”), more evidence has little probative value. Rather, more evidence is most useful in cases about technical matters,¹²⁰ hidden intent,¹²¹ and so forth.

Yet courts may have difficulty discerning the probative value of evidence before discovery of that evidence. Nobody knows in advance what a witness will say in a deposition, making it difficult to assess the probative value of going beyond the ten-deposition limit.¹²² The same holds for searching computers or voluminous paper files: the party opposing discovery will have to make its proportionality cost-benefit argument *before* the search, so the court will not see the fruits of the discovery before having to rule on discovery’s likely benefit.¹²³

“In the absence of any information about [the] evidence,” Judge Richard Posner noted in discussing how parties anticipate opposing evidence, the only option is to “assume that such evidence . . . is of average helpfulness.”¹²⁴ For example, when a court must decide whether a data search for similar stock trades is worth the cost, all it

118. FED. R. CIV. P. 26(b)(2)(C)(iii).

119. See *infra* note 134 and accompanying text.

120. The rules and cases on expert witness admissibility expressly rely on this logic. See, e.g., *Andrews v. Metro N. Commuter R.R. Co.*, 882 F.2d 705, 708 (2d Cir. 1989) (“For an expert’s testimony to be admissible . . . , it must be directed to . . . scientific, technical, or specialized knowledge and not to lay matters which a jury is capable of understanding and deciding without the expert’s help.”).

121. Employment discrimination cases are the paradigmatic example. See sources cited *supra* note 1.

122. FED. R. CIV. P. 30(a)(2)(A) (requiring “leave of court” for more than ten depositions); *id.* 30(d)(1) (providing that “a deposition is limited to 1 day of 7 hours” absent leave of court).

123. See BONE, *supra* note 100, at 229 n.36 (noting that to expand discovery past presumptive limits, courts must assess the value of greater discovery, which “is bound to be difficult in the absence of precise knowledge of what the discovery will reveal”).

124. POSNER, *supra* note 100, at 571.

knows is whether *these kinds* of searches, in *these kinds* of cases, are *usually* fruitful for plaintiffs (or defendants).¹²⁵ As when assessing the size (*L*) of cases with subjective value, courts assessing probative value have little other than Professors Sherman and Kinnard's idea that "patterns of appropriate discovery in certain cases may emerge" based on "rules-of-thumb for determining the amount of discovery normally permissible in certain [case] types."¹²⁶ These rules about particular case types, however, can be fairly indeterminate. In some "civil rights cases," for example, "further discovery is unlikely to shed additional light on the issues"¹²⁷; on the other hand, perhaps such cases' complex intent questions require more extensive evidence gathering.¹²⁸ The tension between these competing views of discovery in civil rights cases shows how imperfect such rules of thumb can be, even if they are the best among the imperfect alternatives available to judges.

Thus, even the most relevant proportionality factor—probative value—can be difficult for courts to analyze, especially if they cannot see the evidence before ruling (such as in determining whether upcoming deposition testimony will include enough useful content). As this Section moves on to discuss, it is just as hard for courts to analyze *p*, the variable capturing the merit of the case (the probability that the plaintiff would win at trial), and, disturbingly, most courts do not even see merit as relevant to discovery decisions.

3. *Probability the Claim Has Merit (p): Difficult for Court to Assess before Seeing All the Evidence and the Parties' Arguments about That Evidence.*

a. Why Courts Do Not Consider Case Merit in Making Discovery Decisions: The Conventional Wisdom that Discovery is Unrelated to Case Merit. The Federal Rules of Civil Procedure were created, and initially interpreted, on a consensus view, in the words of a famous

125. The plaintiff, the party with the burden of proving its opponent's misconduct, usually is the one seeking more discovery. Defendants may seek extensive discovery to prove misdeeds by plaintiffs, *see* sources cited *supra* note 90; such a defendant is in a position akin to that of a plaintiff, seeking evidence to prove its opponent's misdeeds, which is why this Article takes as its paradigmatic example plaintiffs seeking evidence to prove allegations of misconduct that defendants deny.

126. Sherman & Kinnard, *supra* note 112, at 279.

127. *Id.*

128. One example of a case that presents a complex intent question is *Hollander v. American Cyanamid Co.*, 895 F.2d 80, 85 (2d Cir. 1990). *See supra* note 1.

article by Professor Robert Cover, that “the procedure available in our courts of general jurisdiction,” including the federal courts, “is assumed to be largely invariant with substance.”¹²⁹ That view of litigation procedure has drawn increasing criticism,¹³⁰ beginning with Professor Cover, and some of the Federal Rules do mandate inquiries into the merits, though primarily rules governing substantive (rather than purely procedural) pretrial rulings on the merits, such as motions for dismissal,¹³¹ for summary judgment,¹³² or for preliminary injunctions.¹³³ But for discovery in particular, it remains received wisdom that proportionality rulings do not depend on the case merits (p , the odds the plaintiff would win at trial); rather, the factors entering into proportionality rulings include case size (L , amount in controversy and importance of issues) and issue complexity (Δp , probative value).¹³⁴ The proportionality rule and its Advisory Committee’s note detail various factors for assessing the likely benefit of requested discovery, none of which relates to case merits; all the factors relate to case size, probative value, the parties’ resources, and whether the evidence is available elsewhere.¹³⁵

Courts rarely say *anything* about case merits in deciding discovery disputes. When courts do discuss case merits in adjudicating

129. Robert M. Cover, *For James Wm. Moore: Some Reflections on a Reading of the Rules*, 84 YALE L.J. 718, 732 (1975) (recounting in depth, and criticizing, that prevailing view); *see also* Bone, *supra* note 110, at 894–95 (recounting the prevailing view underlying enactment of the Federal Rules that “procedure was normatively distinct from and subordinate to substantive law . . . [so that] the design of a procedural system was mainly a technical exercise in perfecting administrative machinery . . . enforc[ing] the substantive law (whatever that law might be) . . . [because] the values relevant to procedural rulemaking were not substantive in nature”).

130. *See, e.g.*, Bone & Evans, *supra* note 91, at 1282–83 (“Insofar as the argument assumes that it is possible to mark a sharp divide between procedure and substance, it ignores decades of judicial frustration grappling with the procedure/substance dichotomy.”); Cover, *supra* note 129, at 732–33 (“It is by no means intuitively apparent that the procedural needs of a complex antitrust action . . . and an environmental class action to restrain the building of a pipeline are sufficiently identical to be usefully encompassed in a single set of rules which makes virtually no distinctions among such cases in terms of available process.”).

131. FED. R. CIV. P. 12 (providing for review of merits at the pleading stage).

132. *Id.* 56 (providing for review of merits at the pretrial stage).

133. *Id.* 65 (providing that “likelihood of success” on the merits is a factor in judges’ decisions whether to grant motions for preliminary injunctions).

134. As to L , *see supra* Part II.B.1. As to p , *see supra* Part II.B.2.

135. FED. R. CIV. P. 26(b)(2)(C)(iii) (assessing “likely benefit” by “the needs of the case, the amount in controversy, the parties’ resources, the importance of the issues at stake in the action, and the importance of the proposed discovery in resolving the issues”); *id.* advisory committee’s note (listing similar factors and noting that “cases in public policy spheres, such as employment practices, free speech, and other matters, may have importance far beyond the monetary amount”).

discovery disputes, it almost always is to *disclaim* any consideration of the merits, as in this classic passage from a case decided soon after the 1938 enactment of the Federal Rules of Civil Procedure:

[D]efendant is really arguing . . . that the issue raised by the plaintiff is irrelevant, not that the interrogatories are irrelevant to the issue. . . . Whether or not the plaintiff is right is immaterial at this stage. . . . [T]o ask the Court to decide the whole case on answers to interrogatories involves a misconception of the office of discovery procedure.¹³⁶

There are rare exceptions to the rule that courts do not consider case merit in discovery decisions, but they typically occur in two circumstances. First, Congress may mandate a sequencing of discovery and a merits inquiry, most notably as it did in the Private

136. *Love v. Metro. Life Ins. Co.*, 8 F.R.D. 583, 584 (E.D. Pa. 1948); *see also Carrizosa v. Stassinis*, No. C 05-2280 RMW (RS), 2006 WL 1581953, at *1-2 (N.D. Cal. June 6, 2006) (“[P]ropriety of discovery does not turn on one party’s belief that the claims are without merit. . . . This [discovery] motion . . . does not turn on the merits . . . but on the relevance of the materials requested to such claims.”); *Maher v. Monahan*, No. 98 Civ. 2319 (JGK)(MHD), 2000 WL 777877, at *4 (S.D.N.Y. June 15, 2000) (“We need not . . . address the ultimate merits of plaintiff’s claim in order to assess the immediate discovery dispute.”); *Natural Res. Def. Council v. Curtis*, 189 F.R.D. 4, 8 (D.D.C. 1999) (“[P]ermitt[ing] discovery and leaving [aside] the question of the sufficiency of plaintiffs’ case as a matter of law . . . is the way . . . courts handle such matters. . . . [Plaintiffs] are not required to establish a legally sufficient case . . . as a condition of securing discovery”); *United States v. Clean Harbors*, No. C-89-109-L, 1995 WL 155007, at *3 (D.N.H. Feb. 21, 1995) (“[T]he motion . . . deals solely with parameters of discovery and does not touch or address the merits of the case.”); *In re Gupta Sec. Litig.*, No. 94-1517 FMS (FSL), 1994 WL 675209, at *2 (N.D. Cal. Sept. 24, 1994) (“In considering whether to stay discovery pending . . . a motion to dismiss, a court should not weigh the relative merits.”); *In re First Constitution S’holder Litig.*, 145 F.R.D. 291, 294 (D. Conn. 1991) (deeming case merits irrelevant to a decision to stay discovery pending dismissal motion and stating that “[t]his judicial officer has a great uneasiness in reviewing . . . [the] complaint *and* the pending motion to dismiss and in second-guessing which one is likely to be the more meritorious”); *Chubb Integrated Sys. v. Nat’l Bank of Wash.*, 103 F.R.D. 52, 59 (D.D.C. 1984) (“[O]n questions of discovery, typically, courts do not determine the legal sufficiency of claims.”); *Paramount Film Distrib. Corp. v. Ram*, 15 F.R.D. 404, 405 (E.D.S.C. 1954) (“Plaintiffs’ objections to the remaining interrogatories are based upon their contention that the allegations by the defendants . . . do not constitute a valid defense but so far as [the court is] aware no motion has been made by the plaintiffs to strike this defense of the defendants. [The court] know[s] of no authority that [it] ha[s] to strike such a defense [on its] own motion. The defense, therefore, until stricken is valid. [The court] cannot say that the interrogatories are not relevant”); *V.D. Anderson Co. v. Helena Cotton Oil Co.*, 117 F. Supp. 932, 945 n.9 (E.D. Ark. 1953) (“[I]t is no objection to an interrogatory that it relates to a defense or claim which is insufficient in law. It is not ordinarily the function of the court in passing upon objections to interrogatories to decide ultimate questions.”); *Laird v. United Shipyards, Inc.*, 1 F.R.D. 772, 773 (S.D.N.Y. 1941) (“The validity of defenses need not be determined upon a motion to limit an examination before trial, where the matter sought to be inquired into is relevant”).

Securities Litigation Reform Act requirement of staying discovery¹³⁷ pending dismissal motions under a heightened pleading standard¹³⁸—but such a congressional mandate is rare. Second, courts sometimes analyze whether disputed discovery implicates such a weighty public interest that it must be disallowed absent a sufficient showing of merit¹³⁹—but this rare exception shows not that courts do consider case merits on discovery motions, but instead that a strong public interest can create a sort of privilege against disclosure that only sufficient case merit can overcome. Sporadic exceptions that prove the rule aside, the rule against considering case merit on discovery motions is quite well established and consistently followed.

b. Why Courts Should Consider Case Merit: Optimal Discovery Depends on Whether a Case is a Close Call. Whether or not the conventional wisdom is an accurate statement of how courts *actually* decide discovery disputes, it is dead wrong as to how courts *should* decide them. Accurate cost-benefit analysis of the value of evidence is impossible without considering case merits, because the benefit of evidence (helping a plaintiff¹⁴⁰ prove a case) is highest when the plaintiff's claim has enough merit that the factfinder is permitted, but not compelled, to rule for the plaintiff. In the *lowest*-odds cases, additional evidence has little value because it is unlikely to affect the outcome, which is why parties can move to dismiss before discovery—to avoid discovery when, given the lack of merit, “[n]o amount of discovery could change the legal reality [of] plaintiff's claim.”¹⁴¹

137. Private Securities Litigation Reform Act of 1995 (PSLRA), Pub. L. No. 104-67, § 101, 109 Stat. 737, 747 (codified as amended at 15 U.S.C. § 78u-4(b)(3)(B) (2006)) (“In any private action arising under this chapter, all discovery and other proceedings shall be stayed during the pendency of any motion to dismiss . . .”).

138. 15 U.S.C. § 78u-4b(1) (requiring that plaintiffs identify each false or misleading statement and specify why each was false or misleading).

139. Discovery decisions considering case merits tend to feature a high public interest that discovery would jeopardize, such as a public interest in newsgathering harmed by discovery from journalists, *see, e.g.*, *Apel v. Murphy*, 70 F.R.D. 651, 654 (D.R.I. 1976), or in avoiding publicity that would discourage reports of air accidents, *see, e.g.*, *United Air Lines, Inc. v. United States*, 26 F.R.D. 213, 219 (D. Del. 1960).

140. For a discussion of how the same analysis applies to defendants, *see supra* note 125.

141. *Federico v. United States*, 70 Fed. Cl. 378, 386 (2006) (dismissing the case before discovery for failure to state a claim because of the “legal reality. . . that federal employees who serve by appointment may not bring contract claims”); *see also, e.g.*, *Kloth v. Microsoft Corp.*, 444 F.3d 312, 324 (4th Cir. 2006) (holding the same, on a claim that the defendant “deprived consumers of competitive technology” when the claim of injury from deterred invention of new technology was “speculative and beyond the competence of a judicial proceeding,” so that “discovery would not change or inform the nature of the alleged injuries”).

Conversely, although it is the rare case that is so facially clear-cut that it requires no discovery at all (although there are such cases, such as a clear breach of a contract to pay a certain sum¹⁴²), the *highest*-odds cases more quickly reach the point in discovery after which additional evidence will be of little use.¹⁴³

Thus, the optimal amount of discovery depends on the odds that a claim will win: the closer the case is to having fifty-fifty odds, the more the jury needs additional evidence to help it decide the case, so more discovery should be allowed. The remainder of this Section offers some fairly brief economic modeling of this analysis; readers disinclined to mathematical or economic models can skip those portions of this Article without any problem, but the economic models are offered to show how traditional economic cost-benefit models of litigation could be improved by incorporating the points this Article adds.¹⁴⁴

Case merit (p) affects the optimal discovery amount because of the court's opposing goals: (1) limit discovery cost (C_D) and (2) limit the error cost (C_E) of incorrect verdicts. More discovery raises discovery cost (C_D) while lowering error cost (C_E). The latter has a diminishing marginal benefit; each additional piece of evidence likely is less useful (less helpful at preventing error) than the prior one (for example, the first deposition is the defendant company's key decisionmaker, the second is a key witness, the third is a peripheral witness).¹⁴⁵ The court must choose the discovery amount (Q , quantity)

142. See, e.g., *New Rochelle Dodge, Inc. v. Bank of N.Y.*, 511 N.Y.S.2d 663, 665 (App. Div. 1987) (granting the plaintiff prediscovery "summary judgment in lieu of complaint" when the defendant "acknowledged the debt" on retail installment contract); *Chase Manhattan Bank, N.A. v. Marcovitz*, 392 N.Y.S.2d 435, 436 (App. Div. 1977) (holding the same for debt to a law firm).

143. This point should not be overstated because a high-odds case quickly can become a low-odds case if denied sufficient discovery. The only (modest) point here about evidentiary value in close-call cases versus higher-odds cases is just that although both case types typically need discovery, the point at which additional evidence proves redundant is likely to come earlier in the high-odds case.

144. See BONE, *supra* note 100, at 89 n.63 (noting similarly that readers "whose algebra is a bit rusty can skip . . . [these] algebraic expressions without any problem").

145. See Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1482 (1999) ("[As] more evidence is obtained, the effect of additional evidence . . . will tend to decrease, especially if the search[] begins . . . with the most probative evidence."). A related reason evidence offers diminishing marginal benefits is that "[i]f the searcher cannot determine in advance which evidence is . . . fruitful, his search procedure will resemble random sampling, and as the size of a sample grows, the value of additional sampling . . . [is] at a falling rate." *Id.* at 1482–83.

that minimizes total cost, which is the sum of discovery and error cost ($C = C_D + C_E$). The court's decision is illustrated by the following relatively informal model, which is based on the classic incomplete information Cournot duopoly model.¹⁴⁶

Goal: $\min_Q C = C_D + C_E$. Explanation:

- Choose the amount of discovery (Q) that minimizes the sum of error costs and discovery costs.

Where: $C_D = aQ$. Explanation:

- Discovery's cost (C_D) is proportional to its amount (Q) times a constant (a) reflecting cost per unit discovery.

And where: $C_E = (b/Q)(p - p^2)$. Explanation:

- The term $p - p^2$ models error risk as highest for close calls ($p = 0.5$), lowest when merit is clear ($p=0$ or $p=1$).
- Error cost (C_E) drops at a declining rate as discovery increases (Q), as modeled by term b/Q (the constant b reflects evidentiary value per unit discovery).

Substituting the above expressions for C_D and C_E into $\min_Q C = C_D + C_E$ yields

$$\min_Q C = \min_Q aQ + (b/Q)(p - p^2).$$

Optimization conditions: choose Q^* , the optimal discovery amount, to minimize C (more discovery would increase cost more than it reduces errors, and vice versa):

$$(1) \frac{\partial C}{\partial Q} = 0 \text{ and } (2) \frac{\partial^2 C}{\partial Q^2} > 0$$

Calculating Q for condition (1), that is, the Q for which $\frac{\partial C}{\partial Q} = 0$:

$$\frac{\partial C}{\partial Q} = a - \frac{b}{Q}(p - p^2) = 0$$

$$Q^* = \sqrt{(p - p^2) \frac{b}{a}}. \text{ Explanation: Optimal discovery } (Q^*) \text{ depends on } p.$$

Calculating the term constituting condition (2):

146. In a Cournot model, two firms comprise a market; facing an inverse demand curve, each chooses production quantity based on the probability the other has high costs (low output) or low costs (high output). ROBERT GIBBONS, *GAME THEORY FOR APPLIED ECONOMISTS* 144 (1992). My model is analogous as to the probability a case has merit. With optimal discovery rising, and then falling as p increases, I model p as a continuous, not discrete, variable (and thus use derivatives for optimization calculations). See ROBERT COOTER & THOMAS ULEN, *LAW & ECONOMICS* 317 (4th ed. 2003) (noting preference for "develop[ing] theory using continuous variables").

$$\frac{\partial^2 C}{\partial Q^2} = 2b(p-p^2)/Q^3 > 0.$$

Explanation: The condition is satisfied because all components of the term are positive: b and Q are inherently positive, and $(p-p^2) > 0$ for all p from 0 to 1.

In sum, the optimal discovery amount (Q^*)¹⁴⁷ depends on the odds a claim will win (p). Optimal discovery is highest when the odds are close to fifty-fifty (the closer p is to 0.5) and lowest when merit level is more apparent (the closer p is to 0 or 1).

c. Why Courts Cannot Easily Consider Case Merit: In Discovery, Cases Are in a “Pooling Equilibrium” in which Parties Cannot Signal, and Courts Cannot Assess, Case Merit Effectively. For courts to make accurate decisions as to optimal discovery amounts, they must consider p , case merit—yet that may be the hardest task in a discovery dispute. This is a problem of decisionmaker difficulty interpreting information signals (parties’ claims as to case merit), so it is useful to model the situation with game theory, the branch of economics that is a “powerful tool for modeling information and studying its economic role.”¹⁴⁸ In game theory terms, a decision (here, court discovery rulings) must base on some measure of merit (often in game theory the value of a good for sale, but here the merits of parties’ claims) that parties try to communicate. But during discovery, it is hard for courts to tell which cases truly have merit because all the evidence has not yet been gathered. Even if all the evidence had been gathered, courts cannot review all of a case’s evidence (essentially holding a minitrial) just to resolve a discovery dispute. As a result, in discovery, those of low merit often can falsely signal high merit.¹⁴⁹

With limited potential for effectively communicating merit, the court’s decision must base on the average merit of all cases in the pool, such as the pool of all cases arising under the same statute that

147. To reiterate the preceding calculations: $Q^* = \sqrt{\frac{b}{a}(p - p^2)}$.

148. H. SCOTT BIERMAN & LUIS FERNANDEZ, *GAME THEORY WITH ECONOMIC APPLICATIONS* § 17.1, at 297 (1993); see also Ian Ayres, *Playing Games with the Law*, 42 *STAN. L. REV.* 1291, 1291 (1990) (“[T]he theory of games has increasingly dominated microeconomic theory.”).

149. See ERIC RASMUSEN, *GAMES AND INFORMATION: AN INTRODUCTION TO GAME THEORY* § 11.1, at 320 (4th ed. 2007) (noting that “signaling costs must differ” between those of high and low worth “for signaling to be useful”).

are neither facially frivolous nor obvious winners.¹⁵⁰ Premising discovery rulings on such broad, blunt proxies as case type is destined to yield inaccuracy in many individual cases, but this imperfect “pooling equilibrium,”¹⁵¹ or ruling the same regardless of case merit, remains courts’ best available strategy.¹⁵² Others have noted how parties discount each others’ bragging about the merits of their case or about a certain piece of evidence.¹⁵³ This Article adds that not only *parties*, but also *courts* deciding discovery disputes, face the same dilemma of receiving useless signals of merit.

Worse, judges might have an exaggerated (rather than accurate) perception of the extent to which, in discovery, they must assume all

150. Courts occasionally admit premising their rulings on broad hunches, of questionable accuracy in any individual case, about case merit based on proxies such as case type. *See, e.g., In re First Constitution S’holder Litig.*, 145 F.R.D. 291, 293 (D. Conn. 1991) (“Securities fraud actions are recognized as being particularly vulnerable to strike suits . . . [T]his action belongs to a class that is subject to strike suits . . .”). Most courts, however, deny considering case merit in discovery decisions. *See supra* note 136 and accompanying text. Courts also might see in the pleadings an apparent flaw that justifies limiting discovery until a dismissal motion. *See, e.g., Flaim v. Med. Coll. of Ohio*, 418 F.3d 629, 644 (6th Cir. 2005) (upholding limited discovery pending a dismissal motion, given that the complaint seemed highly questionable and a qualified immunity defense seemed promising).

151. BIERMAN & FERNANDEZ, *supra* note 148, § 18.2.6, at 337 (“[A] pooling equilibrium implies the informed player’s actions reveal nothing about what type of player he is.”). The situation actually is likely a *partially* pooling equilibrium because *some* parties can signal merit effectively, such as with a powerful piece of evidence unearthed early enough to submit to the court on a discovery motion. *See GIBBONS, supra* note 146, at 213–18 (discussing partially pooling equilibria). The pooling diagnosis remains because in many cases the evidence will be equivocal or disputed, and the court will have trouble sifting through both sides’ opposing arguments as to case merit, so many cases of varied merit levels will populate the same pool because they will feature signals (merits arguments) the court cannot distinguish without undertaking more effort than it typically can devote to a discovery dispute.

152. Alternatively, the situation could be viewed not as a pooling equilibrium in a signaling game (in which the party making the showing is informed but the party that must respond is not) but as a screening game (in which the party responding to the showing knows more than the party making that showing). *See, e.g., Andrew F. Daughety & Jennifer F. Reinganum, Found Money? Split-Award Statutes and Settlement of Punitive Damages Cases*, 5 AM. L. & ECON. REV. 134, 140–42 (2003). This Article’s focus on diagnosing and remedying the signaling problems that afflict many cases (that in many cases, parties know more than the judge which claims or defenses have merit) should not be seen as an argument that *all* cases have signaling problems. There presumably are cases of close to complete, or at least symmetrically incomplete, information. Further, not all asymmetric-information cases are best described as signaling rather than as screening games; presumably there are cases in which the judge must screen claims and defenses because the judge knows better than the litigants which claims or defenses have merit—such as cases turning on complex disputed legal interpretations in which the judge knows best what the judge’s views will be on certain arguments.

153. BONE, *supra* note 100, at 205 (noting that because “parties have incentives to misrepresent that they have favorable evidence when they do not . . . [and] verification is not always possible . . . [recipients] discount the truth of the information disclosed”).

cases are low-merit ones not warranting costly discovery. If the pool contains more low- than high-merit cases, it is rational for judges to presume, early in a case, that the case likely has low merit and so does not deserve costly discovery. Presumably judges change such opinions as more case information emerges. But as anyone who has argued politics knows, relevant information may not convince people to change their initial opinions due to common cognitive biases, such as the confirmation bias¹⁵⁴ (people's tendency to be "not equally open to all information, but more open to that which comfortably confirms their views, more inclined to spin disconfirming evidence to fit"¹⁵⁵) and the availability bias (the tendency to assume that easy-to-recall events are more likely than they really are¹⁵⁶). Thus, judges' early-stage inability to distinguish good and bad cases may persist even after they have enough information about a case to separate it from the pool: "a judge might more easily recall cases where discovery was abused, leading her to assign an excessively high probability of abuse in the case before her and therefore choose stricter discovery limits than the case warrants."¹⁵⁷ Even if judges, experts at evidence analysis and logical conclusions, are *less* prone to such biases, there is little reason to think them *immune* from these well-documented quirks in human cognition, especially in light of experimental evidence that judges make decisions with intuitive shortcuts prone to exactly these biases.¹⁵⁸

Although ruling the same on all cases in the pool is the best available judicial strategy, it is merely the best among imperfect

154. P.C. Wason, *On the Failure to Eliminate Hypotheses in a Conceptual Task*, 12 Q.J. EXPERIMENTAL PSYCHOL. 129, 138–39 (1960) (finding that after people make an initial, premature guess as to a numerical pattern, they skew their interpretation of later data to preserve that guess).

155. Scott A. Moss & Daniel A. Malin, Note, *Public Funding for Disability Accommodations: A Rational Solution to Rational Discrimination and the Disabilities of the ADA*, 33 HARV. C.R.-C.L. L. REV. 197, 208 (1998) (noting the role of confirmation bias in perpetuating discriminatory stereotypes).

156. See, e.g., DAVID G. MYERS, SOCIAL PSYCHOLOGY 119–20 (3d ed. 1990); see also Moss & Malin, *supra* note 155, at 207 (noting the role of availability bias in perpetuating discriminatory stereotypes).

157. Bone, *supra* note 102, at 1988.

158. Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1, 16 (2007).

[In] our experimental research on judges[,] [w]e provide tests of judges' general reasoning skills as well as their decision-making skills in legal contexts. Our results demonstrate that judges, like others, commonly make judgments intuitively, rather than reflectively, both generally and in legal contexts.

Id. at 6.

strategies. In fact, it is an imperfect strategy that yields suboptimal results. In low-merit cases, ruling identically on all cases yields more discovery than justified by the need for more evidence to assess case merit. Ruling identically on all cases yields too little discovery in close-call cases; it disallows the extensive discovery that is justified when the case is a close call for the factfinder. In this scenario, the bad cases are treated too well and the good treated too badly, as in Professor George Akerlof's classic economic analysis of used car markets: due to "asymmetry in available information . . . good cars and bad cars must still sell at the same price—since it is impossible for a buyer to tell the difference"; as a result, "bad cars drive out the good because they sell at the same price as good cars."¹⁵⁹

As with product decisions, in litigation the bad may come to drive out the good. If courts allow most cases similar discovery—the amount appropriate for an average case—bad (weak) cases will take up too much time and money, whereas good cases lose for inability to gather enough evidence or may never be filed because they will be allowed only average-case discovery. The harder it is to dismiss bad cases quickly or reliably, (a) the more often those bad cases will settle for a nontrivial amount or (less often) yield a plaintiff's verdict, and (b) the more judicial attention those bad cases will take up, at the expense of the attention the good cases deserve. This is to say that bad cases will drive out good cases; court dockets and parties' litigation efforts may be filled with more bad cases and fewer good cases than they otherwise would have, absent this information problem in discovery. Consequently, the discovery problem this Article diagnoses—courts' inability to separate good and bad cases until after the discovery that accounts for so much litigation cost—may be a cause of the widely noted prevalence of frivolous litigation.¹⁶⁰

Could courts adjudicating discovery disputes undertake the necessary inquiries into the merits? Problematically, any merits analysis will be incomplete; it would lack at least some of the evidence because the analysis would be occurring during discovery and before the resolution of all discovery disputes. Moreover, even if all the evidence is in, the information costs of undertaking a merits analysis

159. George A. Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488, 489–90 (1970).

160. For views on litigation and discovery excess, see *supra* notes 81–91 and accompanying text.

to decide a discovery dispute is prohibitive; the court and the parties would have to spend a great deal of time, and the parties a great deal of money, holding a minitrial presenting and arguing about all the evidence and any allowable inferences. Thus, a merits analysis is necessary, but infeasible, for optimally accurate rulings on discovery disputes.

Under this analysis, the quest for better discovery limits has disappointed not due to bad decisionmaking or bad rulemaking, as many argue.¹⁶¹ Typifying arguments blaming rulemaking, Professor Thomas Rowe criticizes as too “vague” to “curb[] cost and excess”¹⁶² the narrowing of discovery, in the 2000 amendments to Rule 26, from material relevant to the “subject matter” to material relevant to “claims and defenses.”¹⁶³ One judge less charitably depicted “debating [that] difference . . . [as] the juridical equivalent to debating the number of angels that can dance on the head of a pin.”¹⁶⁴ Professor Henry Noyes likewise faults bad rulemaking and bad judging:

[The] e-discovery amendments are the fourth recent attempt to contain discovery. The three prior . . . relied on increased judicial discretion, mistakenly assuming that judges would act to limit discovery. . . . [H]owever, courts have continued to rely on the default policy of “liberal discovery.” . . . [T]he good cause standard is problematic both for the new e-discovery rules and for the existing discovery rules.¹⁶⁵

Professor Noyes concludes that “[t]he courts’ persistent reliance on the ‘liberal rules of discovery’ mantra will only be overcome with express instruction to limit discovery, which is absent from the e-discovery amendments.”¹⁶⁶

161. See, e.g., Henry S. Noyes, *Good Cause Is Bad Medicine for the New E-Discovery Rules*, 21 HARV. J.L. & TECH. 49, 71 (2007) (criticizing proportionality and e-discovery rules as too vague to rein in excess discovery that courts are too unwilling to limit); Redish, *supra* note 78, at 563–64 (noting that “the rules’ drafters and revisers over the years . . . have failed to fashion a discovery process that satisfies most people,” and specifically criticizing discovery rules for lacking more cost shifting or spoliation provisions); Thomas D. Rowe, Jr., *A Square Peg in a Round Hole? The 2000 Limitation on the Scope of Federal Civil Discovery*, 69 TENN. L. REV. 13, 14 (2001) (criticizing federal rules’ discovery limits as vague and therefore unable to change judicial decisionmaking).

162. Rowe, *supra* note 161, at 14.

163. *Id.*

164. *Thompson v. Dep’t of Hous. & Urban Dev.*, 199 F.R.D. 168, 172 (D. Md. 2001).

165. Noyes, *supra* note 161, at 51–52.

166. *Id.* at 52.

This Article, while agreeing that courts' discovery decisionmaking is suboptimal, and perhaps not disagreeing that different rules could help, disagrees as to whether better rules or better judicial decisionmaking truly could fix the problem. The relevance to optimal discovery (as this Section discusses) of case merit, amount in controversy, and evidentiary probative value means that some sort of "proportionality" inquiry is inevitable; one cannot evade the relevance of the proportionality considerations. Yet even with the best of all possible rules and judging, courts and parties would remain stuck in a pooling equilibrium; judges simply do not have the necessary information to make optimal decisions about exactly what discovery to allow. It is a fundamental information timing problem inherent in the discovery stage of litigation: optimal discovery depends on the merits, but the merits are knowable only after discovery. As in the folk song about the hole in the bucket fixable only with a machine requiring water poured from that bucket,¹⁶⁷ the problem is a classic circularity; the problem prevents the solution.

III. SOLVING THE POOLING THAT PREVENTS BETTER DISCOVERY DECISIONS: IN CLOSE CALLS ABOUT COSTLY DISCOVERY, PRESERVE THE EVIDENCE BUT DELAY THE DISCOVERY UNTIL AFTER SUMMARY JUDGMENT

A. *Replacing the Pooling Equilibrium with a Separating Equilibrium as a Case Progresses*

In a pooling equilibrium, decisions are suboptimal because it is hard to distinguish between the meritorious and the unmeritorious, as discussed above. More optimal decisions are possible in a separating equilibrium in which parties are forced to "reveal their types to the previously uninformed" decisionmakers.¹⁶⁸ Courts could make more accurate discovery decisions if they could better tell case merit, allowing more discovery in close-call cases that, being neither clear

167. See *supra* note 34 and accompanying text.

168. Ayres, *supra* note 148, at 1307. Separating equilibria actually may be suboptimal if the signals have no intrinsic value except as signals of merit (for example, obtaining a certain educational degree as a signal of work ethic or intellect). In such a separating equilibrium, the cost of signal acquisition (for example, time and tuition) could exceed the improved ability to separate those of high and low merit. See Michael Spence, *Job Market Signaling*, 87 Q.J. ECON. 355, 364-65 (1973). But this Article addresses forced disclosure of evidence a party *wishes to conceal*, so the problem of *wasteful acquisition* of signals is inapposite.

winner nor clear loser, warrant more extensive evidence gathering. But courts cannot separate the close calls from the broader case pool unless parties can credibly signal merit by citing and asserting the evidence supporting their positions.¹⁶⁹ During discovery, parties have not yet gathered and marshaled all their evidence, so low-, mid-, and high-merit cases are hard to distinguish. Due to parties' inability to signal merit level convincingly, courts are stuck with a pooling, rather than a separating, equilibrium.¹⁷⁰ The only way out is for courts to conduct minitrials in which parties argue case merits, detailing and offering interpretations of the evidence,¹⁷¹ but the information costs (in time and money) of that endeavor are prohibitive for resolving a discovery dispute.

A pooling equilibrium may become a separating equilibrium *over time* as more information emerges that illustrates distinctions among the pool—a point noted by game theory analyses of information problems outside the litigation context, such as analyses of information about product quality¹⁷² and corporate corruption.¹⁷³

169. See ROBERT H. FRANK, *PASSIONS WITHIN REASON: THE STRATEGIC ROLE OF THE EMOTIONS* 96–113 (1988) (discussing how signals can degenerate into cheap talk if listeners are uninformed and therefore unable to spot false signals); Michal Barzuza, *Lemon Signaling in Cross-Listing 27* (Oct. 1, 2007) (unpublished manuscript), available at <http://ssrn.com/abstract=1022282> (discussing investor efforts to distinguish “Type L” companies more susceptible to corruption and “Type H” ones less susceptible and arguing that “[t]here will be a separating equilibrium [if and only if] Type L firms choose not to mimic Type H firms”); Lucian A. Bebchuk, *Asymmetric Information and the Choice of Corporate Governance Arrangements 2* (John M. Olin Ctr. for Law, Econ. & Bus., Harvard Law Sch., Discussion Paper No. 398, 2002), available at <http://ssrn.com/abstract=327842> (noting how a pooling equilibrium can be destabilized if “better” actors can make tangibly different offers).

170. See RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 725 (7th ed. 2007) (noting how pooling equilibria occur when those with higher merit find it “difficult to separate themselves” from those with less).

171. See *supra* Part II.B.3.c.

172. The classic article is George A. Akerlof's *The Market for “Lemons”: Quality Uncertainty and the Market Mechanism*, which notes that in used car markets, “bad cars drive out the good because they sell at the same price as good cars,” but over time better information emerges:

After owning a specific car . . . the car owner can form a good idea of the quality . . . i.e., the owner assigns a new probability . . . that his car is a lemon. This estimate is more accurate than the original estimate. . . . But good cars and bad cars must still sell at the same price — since it is impossible for a buyer to tell the difference

Akerlof, *supra* note 159, at 489–90; see also Alan Mathios, *The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market*, 43 J.L. & ECON. 651, 651, 666–67 (2000) (noting how mandatory disclosure of food nutrition content shifted a market from partial pooling—nondisclosure by higher-fat dressings that varied greatly, in nutritional terms—to full separating).

Pretrial litigation is, at heart, a series of stages at which different information emerges. The paper pleadings stage, disclosing parties' allegations, is followed by pre-discovery dispositive motions (most commonly motions to dismiss for failure to state a claim¹⁷⁴ or for jurisdictional failings¹⁷⁵ or conceivably—but rarely—plaintiffs' motions for judgment on the pleadings¹⁷⁶) that disclose some of the parties' legal arguments and weed out the cases whose (lack of) merit¹⁷⁷ is clearest; those motions are followed by fact disclosures in discovery, which in turn are followed by summary judgment motions that further weed out weak claims, and finally followed by the trial that resolves remaining claims.¹⁷⁸

In sum, as a case progresses through the pretrial stages, it gets easier to distinguish it from the pool. This is why, even though most cases settle, some do not settle until some motion litigation or discovery; the outcomes of certain pretrial skirmishes, or disclosures in early-stage discovery (like the initial, key depositions), may allow parties to signal merit more meaningfully than they could earlier.¹⁷⁹ In this sense, moving from one litigation stage to the next—pleadings, dismissal motions, discovery, and so forth—is the bearing of the information costs necessary to separate by merit an initially hard-to-distinguish pool of cases.

173. See Barzuza, *supra* note 169, at 7–10 (discussing how a pooling equilibrium might become a separating equilibrium if law forces a decision on parties (that is, whether to list stock on an exchange imposing intrusive regulation) that high- and low-value companies decide differently, thereby credibly signaling their value).

174. See FED. R. CIV. P. 12(b)(6).

175. See *id.* 12(b)(1)–(2).

176. Motions for judgment on the pleadings are rare because “federal courts have followed a fairly restrictive standard in ruling on motions for judgment on the pleadings.” 5C CHARLES ALAN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 1368, at 222 (3d ed. 2004) (collecting cases); see also *id.* § 1367, at 207–08 (“[J]udgment on the pleadings only has utility when all material allegations of fact are admitted or not controverted in the pleadings and only questions of law remain . . .”).

177. To be clear, lack of merit could mean any number of ways that a case could lose, whether that the allegations were false, that the allegations were true but could not be supported sufficiently, that the allegations were true but some form of jurisdiction was lacking, or any other reason.

178. See *Swierkiewicz v. Sorema N.A.*, 534 U.S. 506, 510–14 (2002) (discussing this sequence of pretrial stages).

179. See BONE, *supra* note 100, at 90–91; Scott A. Moss, *Illuminating Secrecy: A New Economic Analysis of Confidential Settlements*, 105 MICH. L. REV. 867, 877 (2007).

B. Summary Judgment as the Key “Separating” Process that Allows Courts to Distinguish Cases by Level of Merit (p)

Summary judgment is the critical stage for redressing the case pooling equilibrium problem. Typically coming at the end of discovery, summary judgment is the next point, after most discovery disputes, when the court can meaningfully distinguish among cases. It is exactly the sort of minitrial—reviewing all the evidence to assess case merit—needed to decide discovery disputes accurately. In deciding summary judgment, courts allow to proceed to trial only claims a reasonable jury could decide either way,¹⁸⁰ weeding out both claims with the lowest probability of merit (summary judgment grants to defendants) and claims with the highest probability (grants to plaintiffs). After summary judgment, the only claims left are the close calls in which additional evidence is most useful; summary judgment separates those close-call cases from the pool.

Courts are stuck with a low-information pooling equilibrium until summary judgment, as illustrated by the following model. The estimated probability that a lawsuit is meritorious varies, as litigation progresses, based both on how many cases get weeded out of the pool at each litigation stage and on whether the reason cases are weeded out is that they lack merit. The following are the variables that influence estimates of the probability that a lawsuit is meritorious:

let:

d_1 = fraction of cases *dismissed before discovery* (on motions to dismiss)

d_2 = fraction *dismissed after discovery* (on summary judgment motions)

s_1 = fraction *settling before discovery disputes arise*

thus:

$d_1 + s_1$ = fraction not reaching the end of discovery or summary judgment

$1 - d_1 - s_1$ = fraction reaching the end of discovery (called “Stage II” in this Part)

$d_1 + s_1 + d_2$ = fraction not surviving past summary judgment motions

$1 - d_1 - d_2 - s_1$ = fraction *surviving summary judgment* and thus *going to trial* or *settling just before trial* (called “Stage III” in this Part).

180. See, e.g., *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 252 (1986) (allowing summary judgment only if no reasonable jury could find for the nonmovant); *Guilbert v. Gardner*, 480 F.3d 140, 145 (2d Cir. 2007) (noting that summary judgment is denied when the evidence, “in the light most favorable to the nonmoving party, is such that a reasonable jury could decide in that party’s favor”).

Some data and theory indicate a roughly fifty-fifty chance that plaintiffs will prevail in cases that reach trial after surviving dispositive motions (dismissal and summary judgment) and not settling.¹⁸¹ The fifty-fifty hypothesis draws legitimate criticism,¹⁸² but in weak form it remains useful: dispositive motions and settlements weed out many of the strongest and weakest claims, so the pool of cases reaching trial has a disproportionate share of the close-call claims. Whether cases reaching trial have 50 percent, 40 percent, or 30 percent odds is immaterial, because the key insight is that whatever their particular odds, those odds are higher than in the pool of all filed cases, which includes many cases of little or no merit. Further, at no stage before immediately pretrial (that is, after summary judgment) does the court have a meaningful sense of the merits.

Following is a discussion of what information the court has, or can infer, about case merit at three key stages of the path to trial: first, at the start of the case, before discovery, motions, or in-litigation settlement efforts (Stage I); next, after dismissal motions and early-litigation settlements, including during discovery (Stage II); and next, after summary judgment motions (Stage III).

1. *Stage I—Start of the Case, before Discovery, Motions, or During-Litigation Settlements.* At this early stage, all the court knows is that the parties' pleadings allege exactly opposite facts, and that there are various possible case outcomes: a pretrial finding that the case lacks sufficient merit, either on a motion to dismiss or on a

181. George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 18–19 (1984) (noting that as litigation features fewer trials, “the proportion of plaintiff victories will approach 50 percent” under certain assumptions, such as that the “plaintiff and defendant possess information that is on average of equal precision, and if the application of legal standards is, on the whole, coherent and predictable . . . [and] to the extent [there is a] cost advantage of settlement over litigation”). Professors Priest and Klein collect “substantial evidence” for their “selection hypothesis” that cases selected for trial will tend to be close calls. *Id.* at 31–53, 55 (recounting the evidence).

182. See, e.g., Steven Shavell, *Any Frequency of Plaintiff Victory at Trial is Possible*, 25 J. LEGAL STUD. 493, 494 (1996) (“[D]ata . . . does not support a tendency toward 50 percent plaintiff victories.”). Professor Shavell notes that the fifty-fifty hypothesis may fail under certain information problems or if most lawsuits are meritorious. *Id.* at 494, 499–500. These conditions seem likeliest in certain case types, such as those that are especially uncertain, and thus hard to settle (or dismiss when unmeritorious), because they arise under a new law. See, e.g., Ruth Colker, *The Americans with Disabilities Act: A Windfall for Defendants*, 34 HARV. C.R.-C.L. L. REV. 99, 100 (1999) (“[D]efendants prevail in more than ninety-three percent of reported ADA employment discrimination cases decided on the merits at the trial court level. Of those cases that are appealed, defendants prevail in eighty-four percent of reported cases. These results are worse than results found in comparable areas” (footnotes omitted)).

summary judgment motion; or that the case is the sort of close call that will survive dismissal motions; or that the parties will settle the case. This intuitive sense of the range of possibilities is easily formalized. The probability that a case is meritorious at Stage I (P_I) depends on the fraction of cases weeded out on dismissal and summary judgment (d_1 and d_2 , respectively) and weeded out via settlement (s_1), as well as the likelihood a settled case was meritorious (P_{st}):

P_I = sum of the fraction of cases with each possible outcome multiplied by the probability a case with that outcome is meritorious

$$P_I = (0)(d_1 + d_2) + P_{st}(s_1) + (.5)(1 - d_1 - d_2 - s_1) = .5 - s_1(.5 - P_{st}) - .5(d_1 + d_2)$$

Accordingly, the probability a case is meritorious at the start of litigation (P_I) is less than 0.5, except under two unlikely scenarios: (1) there would have to be few enough cases dismissed on motions (that is, low $d_1 + d_2$) that removing good cases from the pool by settlement dominates the opposite effect of removing weak cases by dismissal, contrary to (very rough) estimates that about one-third of federal cases are dismissed on motions;¹⁸³ and (2) settled cases would have to be on average highly meritorious (high P_{st}), contrary to the (limited) data indicating that many confidential settlements are for modest sums.¹⁸⁴

Thus, judges' likely intuition is that initially, the probability that a case is meritorious is low ($P_I < 0.5$), but that assessed probability increases during pretrial processes, eventually reaching 0.5, or at least some higher level than that of the average case filed, for cases surviving summary judgment (Stage III). The question is whether the court's estimate of case merit rises primarily from filing to the discovery stage (Stage I to Stage II) due to dismissal motions and

183. See Judith Resnik, *Failing Faith: Adjudicatory Procedure in Decline*, 53 U. CHI. L. REV. 494, 511–12 (1986) (“[S]ome 35 percent of all federal cases are disposed of by rulings on motions for dismissal or for summary judgment.” (citing information from the Administrative Office of the United States Courts)); Stephen C. Yeazell, *The Misunderstood Consequences of Modern Civil Process*, 1994 WIS. L. REV. 631, 636 (noting, based on a collection of decades of government data, that the proportion of cases resolved on “dispositive motions[, including] dismissals on the pleadings, summary judgments, and similar rulings that end a case . . . has remained quite constant over fifty years at about one-third of all federal civil cases”). These estimates are very rough, however, and efforts at more precise estimates, such as the percentage dismissed on 12(b)(6) motions, have yielded quite a varied range of figures, from 2 percent to 6 percent or higher. See Richard L. Marcus, *The Puzzling Persistence of Pleading Practice*, 76 TEX. L. REV. 1749, 1754 (1998).

184. See Minna J. Kotkin, *Outing Outcomes: An Empirical Study of Confidential Employment Discrimination Settlements*, 64 WASH. & LEE L. REV. 111, 111–12, 117 (2007).

early settlements, or primarily from discovery to trial (Stage II to Stage III) as a result of summary judgment motions; as discussed below, it is primarily the latter.

2. *Stage II—After Dismissal Motions and Early Settlements (Including Settlements during Discovery)*. At this stage, when most discovery disputes occur, the court has two pieces of information it lacked at the start of litigation: (1) the case survived pre-discovery dismissal motions; and (2) it did not settle early. But early dismissals and settlements, taken together, clarify little about case merit. Dismissal motions do not weed out all low-merit lawsuits, only the lawsuits in which the lack of merit is sufficiently clear on paper filings, given that the Supreme Court has cautioned against granting such motions too readily.¹⁸⁵ Settlements typically are confidential, preventing the court from knowing the terms of settlement or looking any further into the merits,¹⁸⁶ so the court knows nothing meaningful about the merits of settled cases.

The probability that a case is meritorious during discovery, Stage II (P_{II}) can be estimated by noting that the probability of merit of a just-filed (Stage I) case is the weighted average of the following possibilities: (1) that a case survives to reach discovery, Stage II (the fraction $1 - d_1 - s_1$ of all cases, with P_{II} probability of merit); (2) that a case loses on a dismissal motion (fraction d_1 , which by definition has zero probability of merit); and (3) that a case settles early (fraction s_1 , with P_{s1} probability of merit):

$$P_1 = (P_{II})(1 - d_1 - s_1) + (0)(d_1) + (P_{s1})(s_1)$$

$$P_{II} = (P_1 - P_{s1}s_1) / (1 - d_1 - s_1)$$

185. In *Swierkiewicz v. Sorema N.A.*, 534 U.S. 506 (2002), the Court held that motions to dismiss for failure to state a claim rarely should be granted in employment discrimination suits, *id.* at 514–15, one of the most common lawsuit types, *see infra* note 234 and accompanying text (discussing employment cases). The Court may have shown more willingness to allow such dismissals in *Bell Atlantic Corp. v. Twombly*, 127 S. Ct. 1955 (2007), which dismissed an antitrust complaint that insufficiently alleged conspiracy, *id.* at 1973–74. Yet *Twombly* denied abrogating *Swierkiewicz*, *id.*, and may be more of a heightened *antitrust* pleading standard than a major change to general standards for dismissal motions.

186. *See* BONE, *supra* note 100, at 19 (“Empirical research in this area is extremely difficult to conduct because most lawsuits settle and settlements mask evidence of frivolousness.”); Moss, *supra* note 179, at 867, 869 (noting the prevalence of confidentiality clauses in settlements). The one known study of confidential settlements found that in one federal district, the median confidential settlement size was \$30,000 in employment discrimination and \$181,500 in personal injury cases. Kotkin, *supra* note 184, at 144 & n.134. But most such settlements were late in litigation, after discovery or summary judgment, *id.* at 135, 145–49, so the study sheds only a little light on the merits of cases that settle early.

By discovery (Stage II), one knows somewhat, but not much, more about the probability that a case has merit. Knowing how many cases lose on motions to dismiss helps: the probability that a case reaching discovery by surviving pre-discovery motions has merit (P_{II}) is higher when more cases lose on dismissal motions (that is, high d_1), because weeding out unmeritorious cases leaves the remaining pool more meritorious. Although the effect of more dismissals (d_1) is knowable, the effect of higher settlement rates (s_1) is not, because we do not know the merits of settled cases:

- If early settling cases are *mostly unmeritorious* (for example, if defendants mostly pay small nuisance-value settlements of a few thousand dollars in weak cases¹⁸⁷), then early settlement weeds out weak cases, leaving the remaining pool (Stage II cases) of higher merit (that is, $P_{II} > P_I$).
- If early settling cases are *mostly meritorious* (that is, if defendants pay mostly to avoid liability and incriminating disclosures), then settlement decreases the average merit of cases in discovery (Stage II). If the merit of settled cases (P_{s1}) is high, then as the fraction of cases that settle (s_1) rises, the merit of cases reaching discovery (P_{II}) falls. With dismissals weeding out the unmeritorious while settlements weed out the meritorious, one cannot say which is higher, the average merit of the pool of filed cases (P_I) or the average merit of the pool of cases in discovery (P_{II}).¹⁸⁸

187. See, e.g., *Fletcher v. City of Fort Wayne*, 162 F.3d 975, 976, 978 (7th Cir. 1998) (finding that the plaintiffs were not “prevailing parties” due to the size of their \$2,500 to \$5,000 settlements and noting that settlement “for less than the costs of defense is a good working definition of a nuisance-value settlement, unless . . . the stakes of the case are themselves small”). See generally Moss, *supra* note 179, at 899–900 (noting that defendants in some cases “stick to a ‘nuisance-value’ offer (such as \$5000)” (citing *Fletcher*, 162 F.3d at 976)).

188. With settled case merit unknown, varied settlement frequency has indeterminate effects; one cannot tell whether increasing settlements leaves the remaining case pool higher or lower merit:

$$\begin{aligned} \frac{\partial P_{II}}{\partial s_1} &< 0? \\ \frac{\partial P_{II}}{\partial s_1} &= \frac{P_I - P_{s1} s_1}{1 - d_1 - s_1} \frac{\partial s_1}{\partial s_1} \\ &= (P_{s1}(d_1 - 1) + P_I) / (1 - d_1 - s_1)^2 \\ \frac{\partial P_{II}}{\partial s_1} &> 0? \\ &\text{if } P_{s1}(d_1 - 1) + P_I > 0 \\ P_{s1} &> P_I / (1 - d_1) \end{aligned}$$

In short, between filing and discovery (Stage I and Stage II), dismissal motions weed out weak cases while settlements weed out cases of unclear merit. It seems likely that the dismissal of weak cases dominates the theoretically possible effect of settling strong cases,¹⁸⁹ which would mean that Stage II cases have higher average merit ($P_{II} > P_I$). But with so little information about settlements,¹⁹⁰ one cannot make any truly confident statements. Accordingly, courts face much the same dearth of information about cases in discovery (Stage II) that they face as to just-filed cases (Stage I).

3. *Stage III—After Summary Judgment Motions.* At this stage, as discussed above, theory and data suggest that the remaining cases have a roughly fifty-fifty probability of merit (that is, $P_{III} = 0.5$), or at least have higher average merit than cases earlier in litigation, such as newly filed cases.¹⁹¹ With case merit largely unknowable at filing (Stage I) and during discovery (Stage II), Stage III—after summary judgment—is the first point in time at which courts meaningfully can assess case merit, and therefore the first point when cases exist largely in a separating rather than a pooling equilibrium. It is the stage when courts finally can know enough about case merit to decide discovery disputes accurately.

Yet delaying discovery decisions until summary judgment seems to conjure up the hole-in-the-bucket problem again: summary judgment should base on all the evidence, so how can evidence-gathering decisions wait until summary judgment? As discussed in Section C, there is room for a narrow but important practice of making some discovery decisions after summary judgment.

C. *The Prescription: In Close Calls, Preserving the Evidence but Delaying the Discovery until after Summary Judgment*

Because summary judgment motions ideally are evaluations of all the evidence, they typically come after all the evidence is gathered

• $\frac{\partial P_{II}}{\partial s_1} > 0$ only if P_{II} is higher than P_I discounted by the fraction of cases not dismissed.

189. For a discussion of the evidence of high rates of dismissals of weak cases and of the limited value of settled cases, see *supra* notes 184–85 and accompanying text.

190. See *supra* note 186 and accompanying text.

191. For a discussion of the Priest-Klein hypothesis and its critics, see *supra* notes 181–82 and accompanying text.

in discovery.¹⁹² Summary judgment before discovery closes is, and should be, exceptional because “discovery should precede consideration of dispositive motions when the facts sought to be discovered are relevant to consideration of the particular motion.”¹⁹³ For this reason, the summary judgment rule provides that if additional discovery is reasonably available, courts should not grant summary judgment without that discovery, but instead should “deny the motion [or] order a continuance to enable . . . other discovery to be undertaken.”¹⁹⁴ Courts granting summary judgment before completion of discovery risk reversal, as in *Gary Plastic Packaging Corp. v. Merrill Lynch, Pierce, Fenner & Smith, Inc.*,¹⁹⁵ which explained why complete discovery should precede a grant of summary judgment:

While summary judgment is a valuable procedural device . . . it is also a drastic remedy that cuts off the right to have one’s day in court. The harshness of the remedy is exacerbated when the trial court refuses to allow plaintiff to conduct discovery. Discovery serves important purposes, such as . . . fully disclosing the nature and scope of the controversy . . . framing the issues involved, and enabling parties to obtain the factual information needed to prepare for trial. . . . [S]ummary judgment should be sparingly granted . . . when discovery is incomplete and . . . defendants have

192. See *Natural Res. Def. Council v. Curtis*, 189 F.R.D. 4, 8 (D.D.C. 1999) (“[L]eaving the question of the sufficiency of plaintiffs’ case as a matter of law to a point *after discovery closes* is the way in which the federal courts handle such matters. Therefore, plaintiffs are correct . . . that they are not required to establish a legally sufficient case . . . of the applicability of [the Federal Advisory Committee Act] . . . as a condition of securing discovery and that resolution of the legal issues concerning that applicability is *premature until discovery ends*.” (emphases added)); *Bone & Evans*, *supra* note 91, at 1284 (“[T]he procedural system seems to favor postponing a serious evidentiary review until after substantial discovery has been completed. Summary judgment, for example, usually takes place only after the parties have had ample opportunity to uncover information and evidence.”).

193. *Coastal States Gas Corp. v. Dep’t of Energy*, 84 F.R.D. 278, 282 (D. Del. 1979) (citing *Canavan v. Beneficial Fin. Corp.*, 553 F.2d 860, 865 (3d Cir. 1977)); see also *United States v. Price*, 577 F. Supp. 1103, 1115–16 (D.N.J. 1983) (“[W]here a plaintiff must obtain a good deal of information from the opposing party, judgment should be withheld until the discovery process has been completed.” (citing *Nat’l Life Ins. v. Solomon*, 529 F.2d 59, 61 (2d Cir. 1975))); *Concord Labs., Inc. v. Concord Med. Ctr.*, 552 F. Supp. 549, 554 (N.D. Ill. 1982) (holding that if a case entails “knowledge and intent” issues, “material evidence is almost entirely in the hands of the defendants, and where plaintiff can establish a fair likelihood that it can obtain material evidence through discovery, we think it unfair to grant defendants summary judgment until plaintiff has had a full opportunity”).

194. FED. R. CIV. P. 56(f).

195. *Gary Plastic Packaging Corp. v. Merrill Lynch, Pierce, Fenner & Smith, Inc.*, 756 F.2d 230 (2d Cir. 1985).

exclusive possession of the material facts.¹⁹⁶

Consequently, there is good reason that “[m]ost courts are reluctant to grant summary judgment prior to the termination of discovery.”¹⁹⁷

But putting summary judgment after *all* discovery is just a commonsense convention, not a rule. “[T]here is no *requirement* in Rule 56 . . . that summary judgment not be entered until discovery is complete.”¹⁹⁸ In appropriate cases, courts entertain limited-scope summary judgment motions after only partial discovery; examples include motions for summary judgment limited to threshold questions like a governmental defendant’s claim of immunity from suit¹⁹⁹ or a libel defendant’s assertion that only limited evidence is necessary to undercut the plaintiff’s required showing that the allegedly libelous statement was false.²⁰⁰

In certain cases, some burdensome discovery could be allowed only after summary judgment. The main import of this suggestion is not that *more* discovery often should be delayed. Rather, it is that in a meritorious case, certain burdensome discovery is regularly denied—and *must* be denied because courts cannot tell whether the case is meritorious (and therefore is deserving of more discovery than usual) during the pooling equilibrium that exists before summary judgment.

196. *Id.* at 236 (citations omitted); *accord* *Weiss v. Reebok Int’l, Ltd.*, 91 F. App’x 683, 690 (Fed. Cir. 2004) (“Weiss has not had ample opportunity for discovery[, which] . . . was stayed pending resolution of Reebok’s summary judgment motion that was narrowly focused on the structural aspects [of the disputed shoes] . . . Weiss should be granted the time that all litigants receive to gather . . . evidence that the accused shoes can perform the claimed functions.”); *Conn. Bank of Commerce v. Congo*, 309 F.3d 240, 264 (5th Cir. 2002) (“[The] court should have allowed full discovery [to] . . . allow[] the Bank a fair opportunity to present all available material evidence pertinent to its opposition to . . . summary judgment.”).

197. *Price*, 577 F. Supp. at 1115.

198. *Pub. Serv. Co. of Colo. v. Cont’l Cas. Co.*, 26 F.3d 1508, 1518 (10th Cir. 1994) (emphasis added) (quoting *Weir v. Anaconda Co.*, 773 F.2d 1073, 1081 (10th Cir. 1985)); *see also* *Paul Kadair, Inc. v. Sony Corp. of Am.*, 694 F.2d 1017, 1029–30 (5th Cir. 1983) (“[A] plaintiff’s entitlement to discovery prior to a ruling on a motion for summary judgment is not unlimited, and may be cut off when the record shows that the requested discovery is not likely to produce the facts needed by plaintiff to withstand . . . summary judgment.”).

199. *See, e.g., Moore v. Busby*, 92 F. App’x 699, 702 (10th Cir. 2004) (holding that the district court was permitted to stay discovery pending disposition of the summary judgment motion by the defendant, a judge, on the threshold question of his immunity from suit as a judge).

200. *See, e.g., Living Will Ctr. v. NBC Subsidiary (KCNC-TV), Inc.*, 857 P.2d 514, 520 (Colo. Ct. App. 1993) (holding that “[l]imited discovery on the issue of falsity is therefore appropriate” before summary judgment motion because “discovery pertaining to defendants’ state of mind . . . is not pertinent to the issue of falsity . . . [and] the issue of falsity . . . [entails] production only of several hours of original unedited video and audio tapes and internal production memoranda and records”), *rev’d on other grounds*, 879 P.2d 6 (Colo. 1994).

In such a case, the summary judgment denial is a determination that the case is one in which more discovery is warranted than in the broader pool of all cases in discovery; it is a determination that relatively more discovery is warranted than the court could have assumed during discovery.

Notably, courts' existing broad case-management powers over discovery and summary judgment make a new rule technically unnecessary. There already is "a great deal of discretionary power in the trial court" as to discovery,²⁰¹ including as to "controlling and scheduling of discovery"²⁰² and "determining the appropriateness and timing of summary adjudication under Rule 56."²⁰³ Some courts already order that discovery occur in stages, such as by issuing a scheduling order at a Rule 16 pretrial conference (which occurs early in litigation) and requiring a certain order of discovery devices (like document disclosures before depositions) and discovery topics (like fact discovery before expert witness discovery).²⁰⁴ Sequencing any discovery after summary judgment, though, remains rare.²⁰⁵

More broadly, there has been an increasing trend—away from deeming all discovery to occur at once, in a single discovery phase of the lawsuit and toward timing discovery based on the outcome of certain motions or the outcome of initial limited discovery. For example, by statute (the Private Securities Litigation Reform Act²⁰⁶), in securities fraud cases courts presumptively stay discovery²⁰⁷ pending

201. Rex R. Perschbacher & Debra Lyn Bassett, *The End of Law*, 84 B.U. L. REV. 1, 34 (2004) (noting that judges at the trial court level "are permitted to make discretionary rulings with respect to discovery" and other matters); see also Richard L. Marcus, *Slouching Toward Discretion*, 78 NOTRE DAME L. REV. 1561, 1587–90 (2003) (discussing the procedural discretion of district courts in case-management activities, including discovery matters).

202. FED. R. CIV. P. 16(c)(2)(F).

203. *Id.* 16(c)(2)(E).

204. See *id.* 16 advisory committee's note ("[T]he initial disclosures required by Rule 26(a)(1) will ordinarily have been made before entry of the scheduling order, [and] the timing and sequence for disclosure of expert testimony and of the witnesses and exhibits to be used at trial should be tailored to the circumstances of the case and is a matter that should be considered at the initial scheduling conference.").

205. See *supra* notes 192–97 and accompanying text.

206. Private Securities Litigation Reform Act of 1995 (PSLRA), Pub. L. No. 104-67, 109 Stat. 737 (codified as amended in scattered sections of 15 and 18 U.S.C.).

207. 15 U.S.C. § 78u-4(b)(3)(B) (2006) ("In any private action arising under this chapter, all discovery and other proceedings shall be stayed during the pendency of any motion to dismiss . . .").

dismissal motions that face a heightened pleading standard;²⁰⁸ by interpretation of Federal Rule of Civil Procedure 23, courts adjudicating class actions commonly grant limited discovery prior to the motion for class certification and the rest of discovery only when and if the motion is granted;²⁰⁹ and courts facing discovery motions about costly e-discovery sometimes order sampling of a limited amount of the data to help determine whether later discovery of the remainder is warranted. In the securities context, one commentator has observed that the statutory stay of discovery “credentials suits that survive pretrial motions,”²¹⁰ a concept similar to this Article’s broader point that surviving a dispositive motion serves to separate out of the pool those lawsuits most deserving of broader discovery.

In short, if a court denies costly discovery when a case is hard to distinguish from the pool of all cases in discovery (Stage II), it should reconsider that denial of discovery if the case survives summary judgment (that is, reaches Stage III). Surviving summary judgment separates a case from a broader pool (all cases in discovery) into a narrower one (cases reaching trial). More specifically, a summary judgment denial means a reasonable jury could decide either way. In other words, p is roughly 0.5—higher than in most cases, which means that more evidence is more valuable than in most cases (that is, Δp of additional evidence is high). The key problem courts face in deciding discovery disputes is that they would need minitrials to assess case merit sufficiently; summary judgment is the existing point in litigation when the court already undertakes that effort. In deciding summary judgment, the court is bearing the information costs necessary to switch from a pooling equilibrium (where p and Δp are hard to discern) to a separating equilibrium (where it is clearer which are

208. *Id.* § 78u-4(b)(1) (requiring that plaintiffs identify each false or misleading statement and specify why each was false or misleading).

209. *See, e.g.,* *Bouaphakeo v. Tyson Foods, Inc.*, 564 F. Supp. 2d 870, 878 (N.D. Iowa 2008) (“The court limited discovery to class certification issues, and set deadlines for the parties’ briefs related to class action and collective action certification.”); *Hoving v. Transnation Title Ins. Co.*, 545 F. Supp. 2d 662, 670–71 (E.D. Mich. 2008) (“[D]iscovery may commence immediately, but it shall be limited to class certification issues. Discovery must be relevant to the issues of class certification, including numerosity, typicality, commonality, adequacy of representation, and the definition of a proposed class.”); *In re Sonus Networks, Inc. Sec. Litig.*, 247 F.R.D. 244, 252 (D. Mass. 2007) (“This case, of course, is not yet at the summary judgment stage, and [the court] cannot determine the merits of the case based upon the limited discovery that has taken place for the purposes of class certification.”).

210. James D. Cox, *Making Securities Fraud Class Actions Virtuous*, 39 ARIZ. L. REV. 497, 520 (1997).

high- p , high- Δp cases). Based on this analysis, courts' efforts to assess case merit on summary judgment can serve double duty, helping courts decide discovery disputes that, earlier in litigation, they had trouble deciding because case merit and evidentiary value was unclear.

D. The Devil is in the Details: Making Workable the Proposal for Post-Summary Judgment Revisiting of Discovery the Court Had Denied Earlier

In proposing a different way for courts to handle nuts-and-bolts practical matters like discovery disputes, the details matter. This Section discusses five legitimate concerns about this Article's proposal and responds to those concerns by fleshing out how this proposal best could be implemented.

1. Concern Number One: Summary Judgment May Not Clarify Case Merits.

A denial of summary judgment does not always mean a case is a close call; weak cases can survive summary judgment when they are fact intensive or depend on debatable inferences from the facts.

This Article's premise is that cases surviving summary judgment have higher odds of success and are more likely to be the close calls, than the broader pool of all filed cases. Yet some summary judgment denials do not indicate such odds. Decisions denying summary judgment sometimes actually say that the case "barely" survives summary judgment,²¹¹ or that it does so despite "weak" evidence.²¹² Further, summary judgment "is not commonly interposed, and even less frequently granted," in certain areas of law. For example, in negligence lawsuits, "the judge and jury each have a specialized

211. *E.g.*, *Sylvester v. SOS Children's Vill. Ill., Inc.*, 453 F.3d 900, 904 (7th Cir. 2006) ("There is no rich mosaic of circumstantial evidence of retaliation in this case, but there is enough (though maybe barely enough) to preclude summary judgment."); *Smith v. Mattox*, 127 F.3d 1416, 1419–20 (11th Cir. 1997) (holding that the evidence of excessive use of force was "barely" enough because the "hazy border between permissible and forbidden force is marked by a multifaceted, case-by-case balancing test," precluding a ruling on the level of force "within the confines of summary judgment review").

212. *E.g.*, *MetroNet Servs. Corp. v. U.S. W. Commc'ns*, 329 F.3d 986, 1008 (9th Cir. 2003) ("Although the evidence of the financial harm to MetroNet is weak, it is sufficient to withstand summary judgment."); *rev'd sub nom. MetroNet Servs. Corp. v. Qwest Corp.*, 383 F.3d 1124 (9th Cir. 2004); *Colburn v. Trs. of Ind. Univ.*, 739 F. Supp. 1268, 1293 (S.D. Ind. 1990) ("Plaintiffs' evidence . . . is weak, but it appears to be just enough to get them past summary judgment.").

function”;²¹³ such cases often turn on pure factual disputes (for example, drivers disputing who entered an intersection first) or reasonableness and due care inquiries fuzzy enough that even in cases that seem weak, it is hard for the court to say *no* reasonable jury could find for the plaintiff. The same may hold for other areas of law featuring similar reasonableness tests like unreasonable use of force by police.²¹⁴

But a party’s ability to avoid summary judgment by citing factual disputes is less than it once was. Since the 1980s, the Supreme Court has “signal[ed] to the lower courts that summary judgment can be relied upon more so than in the past to weed out frivolous lawsuits and avoid wasteful trials, and the lower courts have responded accordingly.”²¹⁵ As one much-cited case noted, courts “cannot resolve factual disputes that could go to a jury at trial, but weak factual claims can be weeded out through summary judgment motions,” because the mere “existence of a triable [fact] issue” is insufficient to avoid summary judgment; “the triable issue must be evaluated in its factual context, which suggests that the test for summary judgment is whether sufficient evidence exists in the pre-trial record.”²¹⁶ Similarly,

the fact that a summary judgment is difficult to obtain in actions in which the parties’ states of mind are relevant does not mean that it will never be granted . . . [S]ummary judgment has been granted to defendants in suits involving fraud, conspiracy, and other claims turning on state of mind when plaintiffs’ allegations were not sufficiently supported.²¹⁷

213. 10A CHARLES ALAN WRIGHT, ARTHUR R. MILLER & MARY KAY KANE, FEDERAL PRACTICE AND PROCEDURE § 2729, at 533 (3d ed. 1998).

214. See, e.g., *Smith*, 127 F.3d at 1419–20 (“[W]e cannot within the confines of summary judgment review hold the force not obviously unreasonable.”).

215. 10B WRIGHT ET AL., *supra* note 213, § 2727, at 468–69 (footnotes omitted).

216. *Collins v. Associated Pathologists, Ltd.*, 844 F.2d 473, 476 (7th Cir. 1988); see also *Thompson Everett, Inc. v. Nat’l Cable Adver., L.P.*, 57 F.3d 1317, 1323 (4th Cir. 1995) (deeming the “mere existence of some disputed facts” insufficient, because “the quality and quantity of the evidence offered to create a question of fact must be adequate to support a jury verdict [and] if the evidence is ‘merely colorable’ or ‘not significantly probative,’ it may not be adequate to oppose entry of summary judgment” (citation omitted) (quoting *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 249–50 (1986))).

217. 10B WRIGHT ET AL., *supra* note 213, § 2730, at 40–43 (footnotes omitted) (collecting cases); see also *Betkerur v. Aultman Hosp. Ass’n*, 78 F.3d 1079, 1087 (6th Cir. 1996) (“Cases involving state of mind issues are not necessarily inappropriate for summary judgment.”).

Thus, courts do meaningfully assess case merit on summary judgment even on claims that are quite fact specific or that turn on state of mind.

Still, with some weak cases surviving summary judgment, this Article's proposal will not be useful in *every* case surviving summary judgment. It is unsurprising that this Article's proposal is imperfect because its premise is that there is *no* perfect fix. Judges make suboptimal discovery decisions not because they are bad at their jobs or because the rules are badly written, but because of the nature of the information-timing problem: courts lack sufficient information on case merit and evidentiary value to make optimal discovery decisions.

When summary judgment denials do not indicate case merit, judges should not, and will not, view that denial as sufficiently informative to affect their prior discovery rulings. Such uninformative summary judgment denials mean that the pooling equilibrium, in which judges have too little information to make optimal discovery decisions, will persist until trial, because summary judgment does not move the case from a pooling equilibrium into a separating equilibrium consisting mainly of higher-than-average merit close-call cases warranting more discovery. But the judge will know this; after all, the judge, having sifted through each party's evidence and arguments to assess how a reasonable jury could rule, is well positioned to know whether the summary judgment denial was or was not based on the merits of the case.

Consequently, the fact that some summary judgment denials do not indicate case merit means this proposal will not be useful in all cases. Importantly, though, it does not create a risk of bad post-summary judgment discovery grants, because judges will know when their summary judgment denials indicate enough about case merit to warrant reconsideration of their denials of discovery.

2. Concern Number Two: Courts Should Use Alternatives Such as Sampling and Cost Shifting.

When courts hesitate to allow potentially relevant but costly discovery, they need not postpone it until summary judgment, because they have two alternatives more in conformity with existing discovery practice: ordering cost shifting that allows the discovery only if the requesting party is willing to pay some or all of the cost; or ordering a partial sampling of high-volume discovery.

Sampling and cost shifting are useful discovery tools but, as discussed in this Section, they are not a panacea and do not eliminate the information-timing problem that makes post-summary judgment discovery potentially useful.

One tool courts do use is sampling, allowing discovery of a fraction of the data first, then discovery of the rest of the data if the sample proves to contain promising evidence.²¹⁸

A phased approach will allow the Court to engage in a more meaningful benefit-burden analysis before determining whether to require cost-shifting After Defendant restores a portion of the back-up tapes . . . Plaintiff will then have the opportunity . . . to determine if it contains relevant evidence and if additional restoration of back-up tapes is warranted. . . . [R]estoration of one-fourth . . . should be adequate to determine whether the tapes are likely to possess relevant evidence.²¹⁹

Yet sampling is useful only if two conditions both hold: first, the sample must be much cheaper than all the evidence (which is not true if the main cost is finding a way to read old data); second, the key evidence must be likely to be present in a small sample (which is not true if a plaintiff seeks just one key e-mail, because its absence from a sample will prove nothing). Sampling is only a limited fix because of these conditions and because it does not redress the main problem—courts' difficulty deciding whether to allow costly discovery before seeing much evidence.

Cost shifting, whether under the rules²²⁰ or to a greater degree,²²¹ gives courts a wider range of options for costly discovery than “yes, you can obtain it” and “no, you cannot”—but it is a limited fix that does not resolve the information-timing problem. To begin with, requiring requesting parties to pay for responding parties' production costs jeopardizes nonwealthy plaintiffs' ability to serve the important social function of suing to unearth and redress important violations of

218. For cases requiring sampling of high-volume deleted data, see *Zubulake v. UBS Warburg LLC (Zubulake III)*, 216 F.R.D. 280, 281–82 (S.D.N.Y. 2003); *McPeck v. Ashcroft*, 202 F.R.D. 31, 34–35 (D.D.C. 2001); *AAB Joint Venture v. United States*, 75 Fed. Cl. 432, 443–44 (2007).

219. *AAB Joint Venture*, 75 Fed. Cl. at 443–44 (footnote omitted) (citation omitted).

220. As to cost-shifting rules, see *supra* notes 69–73 and accompanying text. As to cost-shifting case law, see *infra* notes 238–53 and accompanying text.

221. See sources cited *supra* note 78.

law.²²² More fundamentally, allowing cost shifting in limited circumstances does not eliminate courts' need to make difficult-to-impossible decisions about the value of requested discovery.

That is, even with more cost shifting, courts still would face information-intensive decisions about which discovery is (1) sufficiently important that the requesting party should get it without paying production costs; (2) important enough that the requesting party could get it by paying production costs; or (3) sufficiently lacking in value that the requesting party cannot get it even if willing to pay for it. Such decisions remain intractable in many cases because (as is this Article's primary diagnosis about the problem of costly discovery) courts often lack sufficient information about case value and evidentiary value to undertake accurate cost-benefit analyses on discovery disputes.

In sum, both sampling and cost shifting have their place as important tools that, in some cases, can help courts expand their options for discovery rulings. Similarly, it is commendable that the 2000 e-discovery amendments to Rule 26 emphasized efforts to achieve cooperation among the parties in place of judicial resolution. Most notably, those amendments require that the parties' out-of-court discovery planning conference²²³ include "any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced."²²⁴ Given the limited potential of judicial rulings to manage discovery with perfect accuracy, it makes sense for the rules to encourage parties to obviate the need for such discovery rulings as much as possible. But although all of these tools—sampling, cost shifting, and precourt resolution—could reduce the frequency of the information-timing problem this Article diagnoses, none of them truly can eliminate it from all cases. Accordingly, alternative proposals are complements, not substitutes, for this Article's proposal.

222. See Hay, *supra* note 97, at 502 (discussing how discovery helps plaintiffs prove and redress illegality).

223. FED. R. CIV. P. 26(f)(1)–(3) (requiring parties, in advance of their in-court scheduling conference, to meet by themselves to try to reach agreement on a discovery plan before court intervention).

224. *Id.* 26(f)(3)(C).

3. *Concern Number Three: Judges Might Excessively Deny Discovery and Grant Summary Judgment*

Judges might respond to this proposal by denying more discovery (as a way out of difficult proportionality decisions) and by granting summary judgment more often (both to avoid cumbersome post-summary judgment discovery and because plaintiffs will be less able to obtain evidence they need to oppose summary judgment).

This concern is real; as has been argued about the Private Securities Litigation Reform Act provisions requiring a stay of discovery pending dismissal motions facing a heightened pleading standard, “[t]he lack of discovery will hobble the potentially meritorious suit from withstanding a test of its pleadings by denying the plaintiff access to information necessary to specifically plead a violation by the defendant.”²²⁵ For four reasons, however, this concern should not be overstated.

First, a court using this Article’s proposal to deny too much discovery risks reversal on appeal. Although courts cannot always allow *all the discovery* parties want before summary judgment, appellate courts do enforce the rule requiring *as full discovery as possible* before summary judgment, reversing courts that grant summary judgment after unduly denying discovery.²²⁶ Courts are aware of this presumption that pre-summary judgment discovery should be as full as possible; this awareness would not disappear if courts adopt this Article’s proposal.

Second, this Article does not suggest postponing *most* e-discovery. The media focus on the costliest cases,²²⁷ but much e-discovery is modest and should remain part of standard (pre-summary judgment) discovery. A simple, nontechnical search can respond to a request for all e-mails with certain text,²²⁸ some “deleted”

225. Cox, *supra* note 210, at 520.

226. See *supra* notes 194–97 and accompanying text.

227. Janet Novack, *Control/alt/discover*, FORBES, Jan. 13, 1997, at 60, 60 (telling how one e-discovery consultant charged over \$1 million for a court-ordered search of 50,000 tapes, which the consultant cast as “blackmail” (by the plaintiff, not the consultant), and a violation of some “gentleman’s agreement” not to go after each other’s electronic data” among lawyers generally (quoting John Jessen, President, Electronic Evidence Discovery, Inc.)).

228. See, e.g., *Zubulake v. UBS Warburg LLC (Zubulake I)*, 217 F.R.D. 309, 315 (S.D.N.Y. 2003) (“[S]imply . . . create a plain language search. . . [for] ‘header’ information, such as the date or the name of the sender or . . . the text of the e-mail . . . UBS personnel could easily run a search for e-mails containing the words ‘Laura’ or ‘Zubulake.’”).

e-mails remain easily accessible on company servers;²²⁹ and some backup file restoration is affordable.²³⁰ Even costly discovery like many deleted files, or “metadata” in fraud cases addressing when a document was created or altered,²³¹ need not always be delayed because courts can initially allow a partial sample.²³²

Third, the exact impact of this Article’s proposal—more discovery or less, greater or lesser discovery cost—is hard to predict because postponing some discovery will create multiple, sometimes conflicting strategic effects. With some discovery postponed until after summary judgment, there could be (a) less discovery overall (because courts could eliminate certain costly discovery for some cases losing on summary judgment) or more discovery overall (because cases surviving summary judgment would enjoy especially intensive discovery), and (b) more discovery disputes (because there would be arguments over what discovery gets postponed), the same number of discovery disputes (because arguments over what gets postponed already reach the courts, as arguments over whether to allow or disallow the discovery), or fewer discovery disputes (because postponing certain discovery postpones any follow-up disputes, such as disputes about the extent of the data production). If there is a net decrease in discovery cost (the cost of discovery as well as of discovery disputes), that would decrease settlement incentives, which would increase the number of cases undergoing costly discovery—partially countering the decrease in discovery cost. In contrast, if there is a net increase in discovery cost, that too would have competing incentive effects: increasing parties’ incentives to settle early, before much discovery (which would decrease the number of cases with costly discovery) but also increasing the incentive to file frivolous lawsuits that defendants would settle to avoid discovery costs. In short, this (or any other) discovery reform proposal could

229. For example, this author has no technical skills but once recovered many “deleted” e-mails that remained accessible from university servers in an e-mail account subfolder.

230. See, e.g., *Semsroth v. City of Wichita*, 239 F.R.D. 630, 638, 640 (D. Kan. 2006) (rejecting an argument that e-mails on backup tapes were not readily accessible when the estimated cost was \$3,374.95).

231. See Panel Discussion, *supra* note 20, at 22 (comments of James C. Francis IV, J., United States District Court for the Southern District of New York) (noting that “metadata” includes “changes to the document over time [and] who the author of the document is,” as well as when the computer was used on the document, which may help assess “the authenticity of documents” and a party’s “intent . . . in drafting” them).

232. See *supra* notes 218–19 and accompanying text; see also *infra* notes 240, 249 and accompanying text.

have varied and dynamic effects—some effects foreseeable but others not, and some effects in conflict with others. As with any change to a complex system, there is reason to be cautious and humble in making predictions as to the ultimate mix of effects.

Fourth, and perhaps most importantly, the sort of costly discovery that courts likely would postpone under this Article's proposal is the sort that courts *already deny* in many cases. Consequently, this Article's main impact would be to give parties a *better* chance at costly discovery—just *later*, after summary judgment. In addition to cases denying relevant discovery due to cost,²³³ e-discovery decisions in employment cases (which are 12–14 percent of federal civil cases)²³⁴ show that the *best-case* scenario for a plaintiff in even a *high-value* case may be a court order allowing costly e-discovery only with cost shifting, that is, only if the plaintiff pays an often prohibitively high share of the defendant's production costs.

Consider *Zubulake v. UBS Warburg LLC (Zubulake III)*.²³⁵ Title VII (employment discrimination) monetary relief is only lost pay plus capped emotional distress and punitive damages,²³⁶ but *Zubulake's* potential damages “undoubtedly” were “higher than [those in] the vast majority of Title VII” cases: with millions in lost pay, the plaintiff claimed damages of roughly \$15 to \$19 million, and the defendant counterestimated approximately \$1.2 million.²³⁷ Though the more than \$165,000 e-discovery cost was “surely not ‘significantly disproportionate’” to the case value and “weigh[ed] against cost-shifting,” the court still shifted 25 percent of that cost to the plaintiff,²³⁸ even though the evidence was relevant and the plaintiff made a “limited and targeted request” for e-mails about her sent to or from five individuals.²³⁹ The defendant initially produced a small sample, five of ninety-four backup tapes, and “a review of these e-mails reveal[ed] that they [were] relevant” to *Zubulake's* claim of

233. For a collection of cases denying seemingly relevant discovery due to cost, see *supra* notes 84–85, 88.

234. See Ann C. Hodges, *Mediation and the Transformation of American Labor Unions*, 69 MO. L. REV. 365, 369 & n.27 (2004) (noting that such cases “also substantially increased in many state courts”).

235. *Zubulake v. UBS Warburg LLC (Zubulake III)*, 216 F.R.D. 280 (S.D.N.Y. 2003).

236. 42 U.S.C. § 1981a(b)(3) (2006) (capping emotional distress and punitive damages at \$50,000 to \$300,000).

237. *Zubulake III*, 216 F.R.D. at 281, 288 (recounting that *Zubulake* had earned a \$650,000 annual salary as an equities trader at a New York securities firm).

238. *Id.* at 287–88, 291.

239. *Id.* at 281–82, 285.

termination not for performance but due to discrimination by her supervisor Chapin and others:

[T]hey tell a compelling story of the dysfunctional atmosphere . . . [and] Chapin’s behavior. . . [T]he e-mails contradict testimony given by UBS employees . . . An e-mail from Chapin . . . acknowledg[ed] that Zubulake’s “ability to do a good job . . . is clear,” and that she is “quite capable.” . . . [E]-mail contains the precise words used by the author. . . a particularly powerful form of proof at trial . . . as an admission.²⁴⁰

The “marginal utility” of the evidence “may be quite high,” but just “*potentially*,” because the sample lacked “direct evidence of discrimination”²⁴¹—an oddly high threshold, given the Supreme Court holding that direct evidence is *not* necessary to prove discrimination.²⁴² Faulting the plaintiff for the inability to prove with certainty that it would find a smoking gun in as-yet-unseen discovery, the court held that despite the “powerful” admissions in e-mails that tell a “compelling story,” marginal utility analysis weighed only “*slightly* against cost-shifting.”²⁴³

In sum, *Zubulake 6t tTJ/F4 1 Tf3.616 0 TD0.0002 Tc0.0011 Tw(aimposd oost- -5.5(thif) chif st cost- hif twoweir laythe saeen h*

more—75 percent of the \$249,000 cost to search backup tapes for high-relevance evidence “relating to CBRE’s workplace environment,” including “pornograph[y] . . . distributed electronically (i.e., via e-mail) and displayed on [office] computers.”²⁴⁸ A “test search” (sampling) “result[ed] in relevant documents that had not been produced” earlier—between 1.64 and 6.5 percent of sampled e-mails were relevant, depending on which party one asks—so the evidence was “only available through restoring and searching the backup tapes.”²⁴⁹ Yet the court viewed those statistics negatively: “marginal utility” was low because the sample “revealed a significant number of *unresponsive* documents.”²⁵⁰ To say that finding hundreds of e-mails required searching thousands, however, is a criticism not of *utility* but of *cost*. Further, *Wiginton* deemed the case stakes of a class action under a major remedial federal statute (Title VII) insufficient to justify the discovery:

Plaintiffs claim that should a class be certified, their class recovery could extend into the tens of millions of dollars. While the Court cannot completely accept Plaintiffs’ speculative estimate . . . neither can it accept that their claims are worthless . . . Nevertheless, several hundred thousand dollars for one limited part of discovery is a substantial amount . . . Therefore, this factor weighs in favor of cost-shifting.²⁵¹

Even “[t]he importance of the issues at stake . . . [did] not weigh in favor of or against cost-shifting” in *Wiginton*.²⁵² Despite the plaintiffs’ allegations of mass sexual harassment, the Court noted the parties’ argument that “this factor ‘will only rarely come into play . . . [and that] discrimination in the workplace . . . is hardly unique.’”²⁵³

“Publication bias”—the fact that published decisions are just the tip of the iceberg, and unpublished or unwritten orders may be very different²⁵⁴—is especially salient for discovery rulings, which often

248. *Wiginton*, 229 F.R.D. at 569–70, 577.

249. *Id.* at 571, 574.

250. *Id.* at 575 (emphasis added).

251. *Id.*

252. *Id.* at 576.

253. *Id.* (quoting *Zubulake v. UBS Warburg LLC (Zubulake III)*, 216 F.R.D. 280, 289 (S.D.N.Y. 2003)).

254. See Margo Schlanger, *Civil Rights Injunctions over Time: A Case Study of Jail and Prison Court Orders*, 81 N.Y.U. L. REV. 550, 599 n.163 (2006) (“[E]ven ‘unpublished’ opinions in the federal courts of appeals are available via Westlaw, whereas the problem of non-

deny discovery in unwritten, unappealable²⁵⁵ oral orders at court conferences.²⁵⁶ In one unrecorded court conference in a typical Title VII case,²⁵⁷ the judge said, “That is insane, insane!” when the plaintiff’s attorney²⁵⁸ stated plans to depose ten employees (the number the Federal Rules of Civil Procedure deem permissible without court permission²⁵⁹); the judge also said Title VII plaintiffs “never” get to see the personnel files of “comparators” (those who got the disputed job), which commonly are part of discovery in Title VII cases.²⁶⁰

With courts often grudging about even run-of-the-mill discovery—ten depositions and relevant personnel files—and with courts refusing to allow much e-discovery without cost shifting in even high-value, high-import cases like *Zubulake III* and *Wiginton*, there is little hope for plaintiffs in most cases to obtain costly e-discovery—unless, under this Article’s proposal, the case proves its merit to the judge by surviving summary judgment. The effect of this Article’s proposal on most cases would be to allow plaintiffs a second chance, post-summary judgment, to seek the sort of discovery courts rarely allow.

Nevertheless, valid concern remains that courts may misuse this proposal to deny too much discovery. To address that concern, a new rule, though not required, would be advisable. A rule could make the intent of this proposal as clear as possible to district courts making discovery decisions and appellate courts reviewing summary judgment. A new accompanying Advisory Committee’s note, whether to the rules on case management,²⁶¹ discovery,²⁶² or summary judgment,²⁶³ could clarify similarly²⁶⁴ with phrasing like the following:

publication creates a bias of unknown direction and strength in *district court* opinion analysis.” (emphases added)).

255. See *supra* note 57 and accompanying text.

256. See FED. R. CIV. P. 16 (providing for court conferences on discovery, trial scheduling, etc.).

257. *Wright v. Sports Auth., Inc.*, 01 Civ. 2326 (LAP) (S.D.N.Y. Mar. 29, 2001).

258. The author of this Article was that unfortunate plaintiff’s lawyer.

259. See FED. R. CIV. P. 30(a)(2)(A)(i).

260. For cases allowing discovery of full personnel files, see, for example, *Gatewood v. Stone Container Corp.*, 170 F.R.D. 455, 458 (S.D. Iowa 1996); *Ladson v. Ulltra East Parking Corp.*, 164 F.R.D. 376, 378 (S.D.N.Y. 1996).

261. FED. R. CIV. P. 16.

262. *Id.* 26.

263. *Id.* 56.

“When a court denies discovery it might have allowed were it clearer that additional evidence would prove helpful to the factfinder at trial, the court may reconsider that discovery denial if the case survives summary judgment.” Whatever the wording of a new rule, the message is that post–summary judgment discovery should be a vehicle not mainly for restricting, but primarily for expanding, discovery when a case proves worthy of more evidence gathering by surviving summary judgment.

4. Concern Number Four: Judicial Reluctance to Delay Trial to Reargue a Discovery Dispute.

Judges might be reluctant to allow redundant rearguments of already-decided discovery disputes after summary judgment; relitigating discovery disputes might undesirably delay trial.

To obviate this concern, the procedure for post–summary judgment reconsideration of discovery can be simpler than a full motion for reconsideration. One method is an expedited, streamlined motion: the court could entertain a short reconsideration motion on any discovery previously denied within one week of denying summary judgment. An even more streamlined method would be to allow parties to add to their summary judgment briefings a short discussion of possible post–summary judgment discovery:

- (1) the party opposing summary judgment (typically the plaintiff) could submit, with its summary judgment opposition filing, a short (say, three-page) supplement to its summary judgment brief stating what additional discovery it wants if the case survives summary judgment;
- (2) the party moving for summary judgment (typically the defendant), in its reply papers, also could submit a concise supplement arguing against that additional discovery; and
- (3) the court, if it denies summary judgment, could include in its decision an order stating what, if any, additional discovery is being granted and by when (within so many weeks of the summary judgment ruling, for example) that discovery must occur.

264. Although the Judicial Conference has not issued any Advisory Committee’s notes without a new rule, that policy is not mandated by any law or rule of civil procedure.

More broadly, courts already are creative in managing discovery. They share work between trial and magistrate judges,²⁶⁵ intersperse limited-scope dispositive motions with partial discovery,²⁶⁶ and allow class actions partial discovery limited to discerning the presence or absence of a true class before adjudicating the question of whether a class should be certified.²⁶⁷ There is no reason to think courts could not use these two procedures, or quite likely better ones, to minimize any possible disruption or redundancy that might result from reexamining a discovery dispute after summary judgment.

5. Concern Number Five: Loss of Evidence While Discovery is Delayed.

Evidence might be lost or destroyed between a discovery dispute and a summary judgment denial: summary judgment might not occur until weeks or months after a discovery dispute; it can take months just to brief and argue summary judgment; and it can take months or over a year for courts to decide summary judgment motions.

In an order denying burdensome discovery, the court should issue a preservation order stating that the evidence requested should be preserved until the court decides any summary judgment motions. The extent of parties' duties to preserve evidence is a key e-discovery battleground²⁶⁸ but not a new issue; preservation has been a high-

265. See Richard A. Posner, *Coping with the Caseload: A Comment on Magistrates and Masters*, 137 U. PA. L. REV. 2215, 2216 (1989) (discussing magistrate judges' discovery expertise and arguing that "[a]buse . . . is more likely to occur in a case supervised by a district judge, whose primary responsibilities lie in trying cases and managing . . . docket[s], than in a case supervised by a magistrate, whose most challenging and responsible task is, precisely, to manage discovery in big civil cases").

266. For a discussion of decisions limiting discovery based on anticipated merits, see *supra* note 150 and accompanying text.

267. See 7B WRIGHT ET AL., *supra* note 213, § 1796.1, at 57 ("[I]nitially . . . discovery should be limited to what is necessary for determining whether a proper class action exists."); see also, e.g., *Parker v. Time Warner Entm't Co.*, 331 F.3d 13, 21 (2d Cir. 2003) ("[I]t is likely that at least minimal class discovery must be conducted in order to provide the court with the factual information necessary to decide whether or not to certify a Rule 23(b)(2) class.").

268. See, e.g., *Kronisch v. United States*, 150 F.3d 112, 126 (2d Cir. 1998) (requiring preservation of data upon notice of relevance to ongoing or impending litigation). *Zubulake IV* imposes a broad but not unlimited duty to "suspend [a] routine document retention/destruction policy" to preserve data:

Must a corporation, upon recognizing the threat of litigation, preserve . . . every e-mail or electronic document, and every backup tape? . . . Such a rule would cripple large corporations . . .

. . . Once a party reasonably anticipates litigation, it must suspend its routine document retention/destruction policy . . . [with] a "litigation hold" . . . As a general

stakes bone of contention among parties for decades in disputes about destroying evidence ranging from body parts²⁶⁹ to records of Cold War–era CIA programs.²⁷⁰ Sometimes, courts actually subject a party who destroyed necessary evidence to an adverse inference that the evidence would have been favorable to the other side.²⁷¹

Notably, any preservation controversies arising out of this Article’s proposal would be more limited than the usual preservation disputes. Preservation disputes typically occur early in litigation, when a party demands preservation of *all* data on *every* computer system or data device²⁷² because it does not yet know what data or devices will prove relevant, and it does not want to lose data day by day while, over the first few weeks and months of litigation, it figures out exactly which data or devices actually are most relevant. Courts hesitate to make prediscovery preservation orders unlimited in scope

rule, that litigation hold does not apply to inaccessible backup tapes . . . maintained solely for the purpose of disaster recovery[], which may continue to be recycled . . . [per] company[] policy. On the other hand, if backup tapes are accessible (*i.e.*, actively used for information retrieval), then such tapes *would* likely be subject to the litigation hold.

. . . [However, i]f a company can identify where particular employee documents are stored on backup tapes, then the tapes storing the documents of “key players” to the existing or threatened litigation should be preserved if the information contained on those tapes is not otherwise available. This exception applies to *all* backup tapes.

Zubulake v. UBS Warburg LLC (*Zubulake IV*), 220 F.R.D. 212, 217–18 (S.D.N.Y. 2003).

269. *See, e.g.*, *Welsh v. United States*, 844 F.2d 1239, 1244 (6th Cir. 1988) (“[Defendant’s] act of discarding the skull flap was, if not intentional, at least seriously negligent.”).

270. *See, e.g.*, *Kronisch*, 150 F.3d at 116–18, 126 (allowing an adverse inference against a government defendant when “records were destroyed” by CIA personnel in a case concerning a CIA program of “surreptitious administration of LSD to unwitting nonvolunteer subjects”).

271. *See Welsh*, 844 F.2d at 1244, 1246, 1249 (upholding an adverse inference as to liability, because although the “[defendant’s] negligent destruction of the skull flap does not lead to a conclusion that the medical care of Mr. Welsh was negligent. . . . [t]he destruction did . . . foreseeably prejudice his legal rights”).

272. *See, e.g.*, *AAB Joint Venture v. United States*, 75 Fed. Cl. 432, 443 (2007) (holding that the defendant had “a duty to preserve e-mails from July 2002 to the present, and that Defendant’s decision to transfer the e-mails to back-up tapes does not exempt Defendant from its responsibility to produce relevant e-mails”). As one federal judge explained,

the costs of preservation can be exorbitant, not just . . . not recycling back-up tapes, but . . . implementing a litigation hold, just contacting everybody, finding out where the information resides [T]here is inevitably uncertainty about the scope Are you going to have to preserve back-up data? How far back are you going to have to preserve it? What are your employees going to be able to do in terms of deleting their e-mails?

Panel Discussion, *supra* note 20, at 17 (comments of James C. Francis IV, J., United States District Court for the Southern District of New York).

but do issue quite broad orders because of the uncertainty about what eventually will be discoverable.²⁷³

The sort of preservation order most likely under this Article's proposal, however, would be narrow, extending not to all data and all devices, but just to specific devices with the specific data on which discovery was denied. In contrast, early in litigation a plaintiff might seek a preservation order against deleting any files or e-mails, disposing of computers or personal data devices, or disposing of paper files with personnel matters. The preservation this proposal would require would be more limited, covering only the particular evidence denied in discovery (preserving only e-mails, for example) and only a limited duration—from the discovery dispute to the time summary judgment is decided (when the court would either allow the discovery or end the preservation order).

Further limiting the burden of the necessary preservation is that this Article does not envision preserving all disputed evidence from the time of a discovery dispute to the time summary judgment is decided; preservation is necessary only if the court both (1) sees a real risk of evidence destruction and (2) sees the particular discovery as the sort of close call that it should deny to the requesting party in most cases but perhaps should grant if the case proves its worth by surviving summary judgment.

CONCLUSION

The prevalence of cost-benefit analyses of discovery presents a mismatch of problem and solution. The main problem with discovery decisionmaking is not that judges lack the skill to decide what discovery is insufficiently beneficial or is beneficial but too costly. The problem is that courts face discovery decisions before they have cost-effective access to information needed to make those decisions—before cases separate from a pooling equilibrium in which there are many low-merit cases and in which individual case merit is hard to discern.

273. One judge explained the need for broad preservation orders as follows:

In the paper realm, I can pretty well say, "And thou shalt not destroy any documents of this type" In the electronic arena, I am probably going to have to know which servers the data is likely to reside on, and perhaps who the individuals are whose e-mails have to be preserved. . . . [Y]ou may well have to preserve inaccessible data even though you will make an argument later on that you do not have to produce it.

Panel Discussion, *supra* note 20, at 19 (comments of James C. Francis IV, J., United States District Court for the Southern District of New York).

A timing problem requires a timing solution, such as this Article's proposal for when a court finds a discovery dispute a close call, typically because the requested evidence is likely to be useful but too costly to be warranted in the mass of low-merit cases. In such situations, a court could postpone the discovery until after summary judgment, the point at which the judge can evaluate the case in depth and determine if it likely is the sort of close call in which more evidence gathering than usual is warranted. The proposal aims not to deny or postpone more discovery, but rather to allow, after summary judgment, the sort of helpful but costly discovery that courts deny on a regular basis.

This proposal is not a perfect solution, but that is the point: there is *no* perfect solution to the information timing problem. Optimal discovery depends on case merit, which cannot be assessed until after discovery. Shifting from a pooling to a separating equilibrium for discovery rulings requires considering as much evidence as possible—a prohibitively costly endeavor during discovery, but exactly what courts do on summary judgment. Postponing certain costly discovery until summary judgment is an imperfect solution but better than the prevailing impractical alternative—insisting that rulemakers and judges make accurate discovery cost-benefit decisions without the information necessary to do so.

More broadly, economics is getting better at recognizing one of the defining aspects of litigation—how information emerges over time. For economics to provide accurate diagnoses and useful proposals, it must do more than just prescribe a cost-benefit analysis. Instead, economics must consider critical information-timing matters, whether by modeling cases with options theory,²⁷⁴ by noting differences in information disclosure by litigation stage,²⁷⁵ or—as this Article attempts with discovery—by modeling litigation disputes based on how information costs and merit signals change as litigation progresses.

274. See, e.g., Joseph A. Grundfest & Peter H. Huang, *The Unexpected Value of Litigation: A Real Options Perspective*, 58 STAN. L. REV. 1267, 1270–71 (2006) (modeling litigation with “real options theory,” “[a] tool[] applied to the economic analysis of research and development projects”).

275. See, e.g., Moss, *supra* note 179, at 877 (analyzing settlement confidentiality based on information distinctions between settlements reached before and after litigation commences).