BYU Law Review

Volume 1997 | Issue 1 Article 3

3-1-1997

Pleading and Proof: The Economics of Legal Burdens

Thomas R. Lee

Pleading and Proof: The Economics of Legal Burdens

Thomas R. Lee[*]

Courts have traditionally assigned burdens of pleading and burdens of proof by mechanically applying any of a number of meaningless "tests." Conventional doctrine assigns these burdens to the party to whose case the issue in question is "essential,"[1] or to the party who must establish the "affirmative proposition."[2] Neither of these tests provides a coherent methodology for making such allocations. The first is circular—an issue is "essential" by virtue of the fact that the party has been assigned the burden. The latter is unworkable; it depends on accidents of syntax and may be easily manipulated.[3]

---

* Shareholder—Kimball, Parr, Waddoups, Brown & Gee, Salt Lake City, Utah; J.D. 1991 with high honors, The University of Chicago; former law clerk to the Honorable Clarence Thomas, United States Supreme Court, 1994-95; former law clerk to the Honorable J. Harvie Wilkinson III, United States Court of Appeals for the Fourth Circuit, 1991-92.

This Article is dedicated to the memory of my father, Rex E. Lee, who was and is my inspiration in all things, and who had the patience to wade through this Article at its early stages. Thanks also to William M. Landes and Kevin J Worthen for their helpful comments on earlier drafts.

1. *See* MCCORMICK ON EVIDENCE § 337 (John William Strong ed., 4th ed. 1992); *see also* Fontenot v. Upjohn Co., 780 F.2d 1190, 1194 (5th Cir. 1986) (explaining that a party must establish all of the "essential elements" of a claim or defense); Carr v. Wisecup, 263 F.2d 157 (3d Cir. 1959) (explaining that failure of consideration is an affirmative defense because it should be pleaded by the party who would rely on it); Dyco Petroleum Corp. v. Rucker Co., 443 F. Supp. 685 (E.D. Okla. 1977) (discussing imposition of burden of proof on claimant with respect to its claims and theories of recovery, and on counterclaimant with respect to its claim).

2. *See* MCCORMICK ON EVIDENCE, *supra* note 1, § 337; *see also* Mashpee Tribe v. New Seabury Corp., 592 F.2d 575 (1st Cir. 1979); Levine v. Pascal, 236 N.E.2d 425 (Ill. App. Ct. 1968).

3. As one commentator has noted:

> Breach of a promise may be called nonfulfillment. Negligence is often described as the lack of or failure to exercise due care. An action in which plaintiff seeks a declaration of nonliability is just as truly one seeking a declaration that good defenses exist to the claim asserted by the defendant.

FLEMING JAMES, JR. & GEOFFREY C. HAZARD, JR., CIVIL PROCEDURE § 7.8, at 322 (3d ed. 1985).

A procedural controversy in the civil rights field illustrates the point. In *Wards Cove Packing Co. v. Atonio*,[4] the Supreme Court engaged in a debate (put to rest by the 1991 Civil Rights Act) over the proper allocation of the burden of proof on the issue of "business necessity" in Title VII disparate impact cases.[5] Both sides of the sharply divided *Wards Cove* Court called upon the conventional tests in support of their positions. Justice Stevens, dissenting in *Wards Cove*, based his argument on the fact that business necessity was the affirmative proposition. Stevens invoked the unassailable—and equally unhelpful—proposition that the burden of proof belongs to the party "'who seeks to move a court in his favor, whether as an original plaintiff whose facts are merely denied, or as a defendant, who, in admitting his adversary's contention and setting up an affirmative defense, takes the role of *actor*.'"[6] According to Stevens, the burden of proving business necessity belonged to defendant because it was an "excuse" asserted by defendant and not a "harm" alleged by plaintiff.[7] The presence of such an excuse is "essential" to the defendant's case once a statistical showing has been made, so the *Wards Cove* dissenters reasonably concluded that business necessity "is a classic example of an affirmative defense."[8]

The *Wards Cove* majority made equally plausible arguments on the other side, however. To the majority, the ultimate "affirmative proposition" in any Title VII case is discrimination on the basis of race, color, etc., so it is the plaintiff "who must prove that it was 'because of such individual's race, color,' etc., that he was denied a desired employment opportunity."[9] Business necessity, in other words, is merely a shorthand for the conclusion that the employer's decision was not based on discriminatory reasons. If the challenged practice serves the employer's "legiti-

---

4. 490 U.S. 642 (1989).

5. The question of business necessity arises at the second phase of a disparate impact claim, after the plaintiff has produced statistical evidence that a certain employment practice has a discriminatory impact on a minority group. Such statistical evidence may give rise to an inference that the practice unlawfully discriminates on the basis of race, sex, or religion, but that inference may be erased if the practice is justified by a "business necessity"—i.e., if it "serves, in a significant way, the legitimate employment goals of the employer." *Id.* at 659; *see also* Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971).

6. *Wards Cove*, 490 U.S. at 670 n.17 (Stevens, J., dissenting) (quoting J. THAYER, A PRELIMINARY TREATISE ON EVIDENCE 368-69 (1868)).

7. *Id.* at 669-70 (Stevens, J., dissenting).

8. *Id.* at 670 (Stevens, J., dissenting).

9. *Id.* at 660 (quoting 42 U.S.C. § 2000e-2(a) (1994)).

mate employment goals,"[10] then any adverse impact is not "because of [plaintiff's] race, color, etc." Thus, disproof of business necessity is the same as proof of discrimination, so the element of business necessity is "essential" to plaintiff's case.

The indeterminacy of the conventional doctrine has led both courts and commentators to throw up their hands and give up on deriving any sort of coherent analytical framework for assigning burdens of pleading and burdens of proof. The Minnesota Supreme Court concluded that "[w]here the burden of proof should rest 'is merely a question of policy and fairness, based on experience in the different situations.'"[11] A prominent commentator was even more candid, declaring that "[t]here is no satisfactory test for allocating the burden of proof."[12]

This Article attempts to fill this analytical void by employing economic analysis to provide a framework for allocating burdens of pleading and proof. The Article analyzes procedural rules on burdens of pleading and burdens of proof under a framework of cost minimization; the goal is to explain the law's allocations of burdens as attempts to minimize social costs. Part I introduces the economic framework to be employed throughout the Article. Part II utilizes the model to derive some economic justifications for the basic system of pleading under current law; it offers a rationale for the default rule that assigns the burden of pleading on most issues to plaintiffs and suggests two theories for the law's deviations from the default rule (i.e., affirmative defenses). Part III follows the same pattern in the context of burdens of proof—it first makes the economic case for the default rule and then outlines several economic justifications for affirmative defenses. Both sections illustrate the economic theory with specific applications under current law. Part IV sharpens the distinction between the different economic functions performed by burdens of pleading and burdens of proof by analyzing two areas in which the law has departed from the traditional rule that the burden of proof follows the burden of pleading. Finally, Part V attempts to synthesize the analysis developed throughout the Article by returning to the issue of "business necessity" and applying the eco-

---

10. *Id.* at 661.

11. Rustad v. Great Northern Ry., 142 N.W. 727, 728 (Minn. 1913) (quoting 9 WIGMORE ON EVIDENCE § 2486).

12. JAMES & HAZARD, *supra* note 3, § 7.8, at 322.

nomic model to the decision of where to allocate the burden of pleading and burden of proof in Title VII disparate impact claims.

I. THE ECONOMIC FRAMEWORK: PROCEDURAL RULES AND COST MINIMIZATION

William Landes and Richard Posner were the first to employ economics to analyze rules of procedure.[13] They posited that procedural rules were devices for minimizing social costs, and applied their hypothesis to provide a positive economic analysis of various aspects of our procedural system. Their analysis included evaluations of discovery rules, fee shifting, res judicata, stare decisis, the bail system, and pretrial detention.[14]

The basic theory of cost minimization may also be used to develop an economic model for evaluating the law's assignments of burdens of pleading and proof.[15] Under the framework devised here, such assignments are seen as minimizing a social loss function (L) composed of two kinds of costs: direct costs (DC) and error costs (EC), each of which is a function of other variables. The loss function can be expressed as follows:

$$L = DC + EC,$$

where:

$$EC = kQq_1EC_1 + (1 - k)Qq_2EC_2,$$

and where:

Q	=	quantity of litigation (number of cases)
k	=	fraction of cases in which defendant is truly liable
q_1	=	probability that undeserving defendant will win (Type I error)
q_2	=	probability that undeserving plaintiff will win (Type II error)
EC_1	=	error costs per Type I error
EC_2	=	error costs per Type II error

13. *See* William M. Landes, *An Economic Analysis of the Courts*, 14 J.L. & ECON. 61 (1971); Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399 (1973).

14. *See* Landes, *supra* note 13, at 62; Posner, *supra* note 13, at 400.

15. Posner also devoted several paragraphs to the burden of proof. *See* Posner, *supra* note 13, at 408-09. This Article responds to some of the questions that he raised.

Given these variables, other useful expressions can be derived and described as follows:

$(1 - q_1)$	=	probability that deserving plaintiff will win
$(1 - q_2)$	=	probability that deserving defendant will win
$(1 - k)$	=	fraction of cases in which defendant is truly not liable
kQ	=	number of defendants who are truly liable
$(1 - k)Q$	=	number of defendants who are truly not liable

Direct costs are those most commonly associated with litigation—the costs of deciding whether to pursue a claim, of undertaking discovery, drafting pleadings, conducting a trial, and fashioning and enforcing a final judgment. Direct costs are a function of the quantity of litigation (Q), and they include both private and public costs of litigation.

Error costs are the social costs associated with erroneous legal judgments and are a function of several variables. Erroneous judgments include decisions for undeserving defendants (Type I errors) and decisions for undeserving plaintiffs (Type II errors).[16] The expected cost of each individual error is the product of the probability of the error (q_1 or q_2) and the magnitude of the error (EC_1 or EC_2). Total error costs additionally depend on the fraction of defendants who are truly liable (k) and the total quantity of litigation (Q). In the loss function expressed above, total Type I error costs are kQq_1EC_1 and total Type II error costs are $(1\text{-}k)Qq_2EC_2$. The probability of error (q_1 or q_2), will depend on several variables: the standard of proof used by the court, the allocation of burdens, and the court's level of confidence in the accuracy of its decision.[17]

This Article demonstrates that the law's allocations of burdens can be understood as attempts to minimize the sum of both direct and error costs. The goal is to cut through the rhetoric of the prevailing "tests" and to utilize the simple economic assumption of cost minimization to generate a coherent analytical structure for the assignment of burdens. For both burdens of pleading

---

16. *See* A. Mitchell Polinsky & Steven Shavell, *Legal Error, Litigation, and the Incentive to Obey the Law*, 5 J.L. ECON. & ORG. 99 (1989) (analyzing the effect of these two types of legal errors on the decision to bring the suit and on the incentive to obey the law).

17. The court's level of confidence in the accuracy of its decision can depend, among other things, on direct costs (DC).

and burdens of proof, the analysis proceeds in two stages: first the economic justifications for a default rule that assigns burdens to plaintiffs are developed, and then some economic bases for deviations from the default rule are explained.

II. THE ECONOMICS OF BURDENS OF PLEADING

The chief function of pleading rules is to facilitate efficient "communication" between parties to a lawsuit. In other words, pleading rules set the parameters for the parties' exchange of "signals" as to what issues are to be disputed; they define the set of issues that each party must assert in order to put those issues into play. A plaintiff in a contract suit, for example, can allege that defendant failed to perform his contractual duties and thereby signal to defendant that he intends to raise that issue. Defendant can respond by admitting or denying plaintiff's allegations, or he can raise new issues such as the lapse of the statute of limitations period. Traditional burdens of pleading inform plaintiff that, in order to proceed with a breach of contract claim, he must allege the validity of the contract and defendant's failure to perform. They also inform defendant that he must raise the statute of limitations issue, since the period is presumed not to have run unless defendant alleges that it has.

In this sense, pleading rules economize on direct costs by virtue of their clarity—their role is to refine the scope of the litigation. The intersection between plaintiff's signals (averments in the complaint and denials of affirmative defenses) and defendant's signals (denials in the answer and averments as affirmative defenses) constitutes the set of matters at issue,[18] i.e., the matters to which direct costs will be devoted. So long as it is clear who has the burden of pleading on any given issue, direct costs are economized because the parties are able to signal which issues they wish to dispute.

Under this view of pleading rules, one might question the propriety of affirmative defenses. The parties' signals operate

---

18. The Federal Rules contain two provisions that, in tandem, advance simplified pleading while preserving the economizing function of narrowing the issues. Rule 10(b) requires that each paragraph of plaintiff's complaint contain a single averment. *See* FED. R. CIV. P. 10(b). Defendant's answer then narrows the scope of disputed issues by admitting or denying each averment and by raising affirmative defenses. A backup provision is Rule 8(b), which permits defendant to engage in an economical narrowing even if plaintiff fails to follow 10(b). Rule 8(b) allows defendant to "specify so much of [the averment] as is true and material . . . and . . . deny only the remainder." FED. R. CIV. P. 8(b).

independently of the allocation decision—they depend only on the clarity of the allocation. The clearest and simplest rule—the default rule—might be to assign the burden of pleading on all matters to the plaintiff. Plaintiff could plead affirmatively all issues now defined as part of the prima facie case and could also plead affirmatively the absence of those matters now categorized as affirmative defenses. Although the economics behind the default rule seem to dictate this result, two economic rationales exist for departing from the default rule in the pleading context. The next two sections explain these two rationales and then offer some examples of affirmative defenses that can be explained by these theories.

A. The Economic Cases for Affirmative Defenses

An ironclad default rule with no exceptions would have the virtue of simplicity, but it would be extremely costly. Consider plaintiff's breach of contract suit. Plaintiff would have to allege affirmatively that the statute of limitations has not run, that the contract was made in the absence of fraud or duress, and that there are no estoppel, statute of frauds, or res judicata problems, even though defendant may never dispute such allegations. Thus, under what we will call the "probability" theory, the law assigns the burden of pleading to the defendant on issues considered unlikely to arise. Although the resulting pleading rules are more complex than the simple default rule proposed above, the rules economize on direct costs by presuming the absence of certain issues and saving the direct costs that would otherwise be incurred in raising issues that likely will not be disputed. In this way, pleading rules do more than allow the parties to signal each other; in many cases, they obviate the need for costly signals.

Under this rationale, an "affirmative defense" might be defined as a matter not ordinarily expected to coincide with the elements entitling a plaintiff to prevail on a certain issue. Prima facie elements would be those issues that always arise, while affirmative defenses would be issues that arise less frequently. By assigning the burden of pleading the latter to defendant, pleading rules save the direct costs associated with raising these issues in most cases.

Affirmative defenses perform another function that can be understood only by identifying a second economizing function of burdens of pleading. In addition to setting clear ground rules, burdens of pleading may also economize on direct costs by nar-

rowing the *scope* of the issues in dispute. On some issues, assigning the burden of pleading to plaintiff would require him to make an allegation that would be much broader and more comprehensive than if the burden were assigned to defendant. In such cases, the pleading allocation economizes on direct costs by assigning the burden to the party whose version of the issue in question is more discrete. Consider the issue of illegality. If plaintiff has the burden of pleading, he must allege that no illegality of any kind is present in any provision of the contract. If the burden is assigned to the defendant, on the other hand, defendant can make the narrower allegation that a specific provision of the contract violates a specific provision of law.

The cost effects of this aspect of the burden of pleading are subtle, but not unimportant. First, assigning the burden of pleading to the party whose version of an issue is inherently more narrow allows the parties to more effectively fulfill the signaling function of pleading rules. A defendant's assertion of a discrete instance of illegality in a contract, for example, performs a narrowing function that is not adequately performed by a plaintiff's assertion as to the general absence of illegality of any kind. Second, allocating the burden of pleading to defendant when his allegation is more narrowly focused economizes on the direct costs associated with investigating the factual basis of the claim. In other words, a more narrowly focused claim is less costly to corroborate, and so the pleading rules economize on pleading costs by assigning the burden of pleading to defendants on issues like illegality. We will call this second theory the "relative-cost-of-pleading" theory.

The Federal Rules version of "notice pleading" makes the burden of prepleading investigation lighter than it would be under the "fact pleading" regime, but the Federal Rules regime preserves a duty to investigate the factual basis of any allegation. Under Rule 11, signed pleadings attest that to the best of the signer's knowledge, information, and belief—formed after reasonable inquiry—all matters asserted are "well grounded in fact." Assignment of the burden of pleading thus allocates initial "investigation" costs to the plaintiff on prima facie issues and to the defendant on matters designated as affirmative defenses. By assigning the burden of pleading for a given issue to the party whose version of that issue is narrower and less comprehensive, investigation costs are minimized and the scope of the litigation is more effectively narrowed.

Consider once again the issue of illegality. Traditional rhetoric used by the courts might assign to plaintiff the burden of pleading the legality of a contract because legality is "essential" to plaintiff's breach of contract claim. Yet this allocation would be economically unsound. Assigning the burden of pleading to plaintiff would result in unnecessary expenditures on investigating the basis of a claim that there was no illegality of any sort in the contract. The narrower claim is that there was a specific illegality inherent in a particular provision. By assigning the burden of pleading to defendant, we avoid the costs associated with making an unnecessarily comprehensive allegation.

This sort of economizing function seems to underlie the "test" that assigns burdens based on which party must establish the "affirmative proposition." It is often less costly to investigate the basis of an affirmative proposition, and the "affirmative" side of a coin is often the more discrete side. For example, an allegation of specific illegality is more discrete than an allegation of the general absence of illegality. Yet, although the affirmative proposition test will often yield economically justifiable results, the test is still an imperfect proxy for assigning the burden to the party whose allegation is less comprehensive and thus less costly. Such a test is vulnerable to accidents and manipulations of syntax—lack of illegality can be called legality.

Instead of concentrating on syntax, then, a more consistent test would look to ease of investigation and the adequacy of the narrowing function, assigning the burden to defendant where his allegation is more discrete and thus easier to investigate. Consider again the issue of business necessity in Title VII employment discrimination suits.[19] The test proposed here would ignore imponderable questions of syntax, and would assign the burden of pleading to defendant on the ground that defendant's allegation of a particular business justification is inherently more discrete than plaintiff's claim of the absence of any business justification. This allocation economizes on direct expenditures on investigation and more effectively narrows the scope of issues to which further direct costs must be devoted.

In sum, two economic theories explain affirmative defenses in the pleading context. Under the "probability" theory, some issues are unlikely to arise, and assigning the burden of pleading

---

19. For a full discussion of the economic implications of allocating business necessity in Title VII cases, see *infra* Part V.

to defendant operates as a presumption that saves the cost of raising the issue in most cases. Under the "relative-cost-of-pleading" theory, assigning the burden of pleading to defendant saves on investigation costs and more effectively narrows the scope of litigation. Either theory or a combination of the two can explain the law's departures from the default rule that assigns the burden of pleading to the plaintiff.

B. Application of the Theory to Federal Rule 8(c)

The probability and relative-cost-of-pleading theories may be employed to offer a positive analysis of the affirmative defenses listed in Federal Rule of Civil Procedure 8(c). Different economic purposes are served by different affirmative defenses, and some defenses perform a broader range of economizing functions than others.

Arguably, all of the Rule 8(c) affirmative defenses can be explained by the probability theory. Rule 8(c) assigns the burden of pleading on issues that are not likely to arise in connection with a plaintiff's accrued right. Accord and satisfaction,[20] arbitration and award, assumption of risk, contributory negligence, discharge in bankruptcy, duress, estoppel, failure of consideration,[21] fraud, illegality, injury by fellow servant,[22] laches,[23] license, payment, release, res judicata, statute of frauds, statute of limitations, and waiver are all "unusual" issues. All are less than 100% likely to arise in a given case. The law saves on direct costs by correctly presuming the absence of these issues in most cases, and avoiding any expenditure on raising the issue. Assigning the burden of pleading to the defendant thus saves the direct costs of raising these issues in most cases.

---

20. Accord and satisfaction is an agreement between the parties to settle a claim by substituting a new contract for an old contract or for a cause of action in tort. The agreement can be raised as a defense to a claim on the old contract or on the cause of action. *See* Holm v. Hansen, 248 N.W.2d 503, 506 (Iowa 1976).

21. Failure of consideration is a defense to a contract action. The defense arises when consideration, though initially valid, has since become worthless. *See* Royal Typewriter Co. v. Xerographic Supplies Corp., 719 F.2d 1092, 1107 (11th Cir. 1983).

22. Injury by fellow servant is a defense by an employer against an employee. The defense is essentially an allegation that the negligence of a fellow employee was responsible for that employee's injury.

23. Laches is a defense based on the plaintiff's unreasonable delay in bringing suit and on undue prejudice to defendant. *See* Kansas v. Colorado, 115 S. Ct. 1733, 1742 (1995); Ute Indian Tribe of Uintah & Ouray Reservation v. Probst, 428 F.2d 491, 496 (10th Cir. 1970).

Not all of the Rule 8(c) defenses can be explained by the relative-cost theory, however. Most of the defenses are more narrowly focused than their logical counterparts—an allegation of a discrete instance of accord and satisfaction, arbitration and award, assumption of risk, contributory negligence, discharge, duress, estoppel, failure of consideration, fraud, illegality, injury by fellow servant, license, payment, release, or waiver is "easier" to corroborate than the absence of such instance over a period of time. Other defenses listed in Rule 8(c) do not fit the paradigm, however. With laches, res judicata, statute of frauds, and statute of limitations issues, defendant's allegation is no more discrete than its counterpart; defendant has no advantage as to costs of pleading.

III. THE ECONOMICS OF BURDENS OF PROOF

Once the pleading rules perform their function of setting the dimensions of the ballpark—of assigning to plaintiff or defendant the burden of putting certain matters at issue—the outcome of the litigation often depends less on the allocation of the burden of proof than on whether it is plaintiff or defendant who presents a stronger case on key issues.[24] Regardless of whether plaintiff or defendant has the burden of proof on contributory negligence,[25] for example, plaintiff will present evidence and attempt to convince the factfinder that she was not negligent, and defendant will present his proof in an attempt to convince the factfinder that plaintiff was negligent. The burden of proof in this sense functions as a tiebreaker, dispositive only if the factfinder is equally compelled by the proof offered by both parties (i.e., when liability is said to be "indeterminate") or if the party with the burden of proof offers no proof at all. This aspect of the law of burdens has thus been aptly referred to as the "risk of nonpersuasion."[26]

The decision of where to allocate the risk of nonpersuasion is not unimportant in terms of its effects on the social loss function,

---

24. This analysis assumes the standard of proof is preponderance of the evidence, and concentrates on the economic effect of allocating burdens under that standard. The economics of choosing the appropriate standard of proof is taken up later. *See infra* text accompanying notes 76-84.

25. The states are split on this issue. *See* JAMES & HAZARD, *supra* note 3, § 3.16, at 166.

26. *See id.* § 7.6, at 314; *see also* O'Neal v. McAninch, 513 U.S. 432 (1995) (noting that the burden of proof operates only where "the record is so evenly balanced that a conscientious judge is in grave doubt" about an issue).

however. This section offers an economic rationale for the default rule that assigns the risk of nonpersuasion to the plaintiff, and then suggests a number of justifications for deviating from that default rule. The discussion reveals that the allocation of the burden of proof affects direct costs (*DC*), in the form of post-judgment costs and expenditures on proof, and has an impact on both the magnitude (EC_1 or EC_2) and frequency (q_1 or q_2) of error costs.

A. Theoretical Justifications for the Default Rule

Application of the cost-minimization model introduced above reveals two justifications for a default rule that assigns the burden of proof to the plaintiff. First, the rule economizes on direct costs incurred after the liability determination.[27] A judgment for the plaintiff involves at least three kinds of post-judgment direct costs that are not associated with a judgment for the defendant: remedy construction costs, enforcement costs, and transaction costs associated with payment. In many cases these costs are minimal—the court assigns a value to plaintiff's injury, orders defendant to pay it, and defendant pays plaintiff.

In other cases, however, these costs will be substantial. Constructing an equitable remedy may require a detailed inquiry into the possible effects on each party, and even damages calculations can be extraordinarily complex.[28] Enforcement of the judgment may also be costly—policing and monitoring compliance with an injunction and enforcement through subsequent contempt proceedings may occupy countless hours and span several years.[29] And even payment costs may be signifi-

---

27. This appears to be the explanation given by Judge Posner in his 1973 article. *See* Posner, *supra* note 13, at 408 (suggesting that the default rule is based on the notion that "no allocative purpose would be served by shifting a loss in a case where the defendant's liability was indeterminate"). Judge Posner leaves open the question why the burden of persuasion as to defenses is not also assigned to plaintiff. *Id.* at 408-09. This issue is taken up later in Part III.B.

28. *See, e.g.*, Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 545 (1983) ("[M]assive and complex damages litigation not only burdens the courts, but also undermines the effectiveness of treble-damages suits.").

29. Some remedies are so complex that courts employ special masters or experts to advise them in crafting a remedy. *See* Swann v. Charlotte-Mecklenburg Bd. of Educ., 306 F. Supp. 1291, 1313 (W.D.N.C.) (appointing expert to advise court in devising school desegregation remedy), *vacated on other grounds*, 431 F.2d 138 (4th Cir. 1969); Foster v. City of Detroit, 254 F. Supp. 655, 668 n.33 (E.D. Mich. 1966) (ordering master to supervise damages remedy after city urban renewal condemnation plan found unlawful), *aff'd*, 405 F.2d 138 (6th Cir. 1968). The expansive use of this power has

cant—defendant may have difficulty liquidating and assembling the damages and may incur devastating costs in attempting to comply with an injunction.[30]

Whatever the magnitude of these post-judgment direct costs, the default rule assures that none of them will be incurred when defendant's liability is indeterminate. Liability is indeterminate when a decision for plaintiff is correct just as often as it is erroneous, or when it is equally likely that plaintiff and defendant deserve to prevail. The default rule ensures that post-judgment direct costs are incurred only when they generate some savings in error costs. A court is required to craft a remedy and parties are required to comply with it only if it is more than equally likely that plaintiff deserves recovery. It is useful in this regard to keep in mind the production relationship between DC and EC. Direct expenditures make economic sense only to the extent they "produce" a reduction in error costs. So long as it is equally likely that plaintiff and defendant deserve to prevail, the marginal benefit of additional direct costs is zero (assuming symmetry in the magnitude of the errors, i.e., $EC_1 = EC_2$).[31]

Our model reflects this conclusion. For cases in which liability is indeterminate, assigning the burden of proof to plaintiffs would result in a consistent decision for defendants. Assuming that $k = 0.5$[32] and that $EC_1 = EC_2$, a consistent decision for defendants would make $q_1 = 0.5$ and $q_2 = 0$, so that total error costs would equal $0.25QEC_1$. On the other hand, assigning the burden of proof to defendants would result in a consistent decision for plaintiffs and would increase post-judgment direct costs (DC)

---

come under recent fire. *See* Missouri v. Jenkins, 115 S. Ct. 2038, 2066 (1995) (Thomas, J., concurring) (noting that the Supreme Court has sanctioned "unprecedented" expansions in the lower courts' equitable power to fashion remedies); *id.* at 2067 ("Judges have directed or managed the reconstruction of entire institutions and bureaucracies, with little regard for the inherent limitations on their authority."); *see also* John Choon Yoo, *Who Measures the Chancellor's Foot? The Inherent Remedial Authority of the Federal Courts*, 84 CAL. L. REV. 1121 (1996) (challenging the expansion of judicial remedial power as inconsistent with constitutional principles of separation of powers).

30. In mass tort product liability suits, for example, there is a significant risk that manufacturers will be bankrupted by punitive damages verdicts in multiple actions involving a single product. *See* Roginsky v. Richardson-Merrell, Inc., 378 F.2d 832 (2d Cir. 1967); Tetuan v. A.H. Robins Co., 738 P.2d 1210 (Kan. 1987).

31. Note that a perceived divergence between EC_1 and EC_2 is probably the economic explanation for standards of proof higher than preponderance. *See infra* text accompanying notes 76-84.

32. For a discussion of the economic implications of relaxing this assumption, see *infra* Part III.B.3.

without any change in error costs.[33] The default rule is therefore the cost-minimizing rule. Only when the court finds that liability is *not* indeterminate and that plaintiff is more likely than not to deserve recovery will there be an economic argument for incurring the marginal increase in *DC*. In such cases, regardless of who has the burden of proof, a consistent decision for defendants would make $q_1 > 0.5$ and $q_2 = 0$, so that total error costs would exceed $0.25QEC_1$. A consistent decision for plaintiffs would increase post-judgment direct costs, but would save on error costs in the amount of $q_1EC_1 - q_2EC_2$.[34]

A second and related justification for the default rule is that it economizes on direct costs by deterring claims based on indeterminate issues of liability. Intuitively, the idea behind the default rule is to encourage only those suits that serve an allocative function (i.e., that decrease error costs). In cases in which the burden of proof is dispositive (where there is a probability of exactly 0.5 that plaintiff is entitled to damages), there is no benefit in allowing the plaintiff to prevail—in 50% of the cases plaintiffs will recover when defendants should be liable; in 50% of the cases plaintiffs will recover when defendants should not be liable.[35] Assuming symmetry in the magnitude of each type of error ($EC_1 = EC_2$), error costs are constant regardless of whether plaintiffs or defendants win a given suit. In sum, the default rule is rooted in the notion that a plaintiff should prevail only if such a result saves on error costs. This condition is satisfied only if the probability that plaintiff is entitled to recovery is greater than 0.5.

Consider the following numerical illustration. If the factfinder in a given case determines that the probability that plaintiff is entitled to recover is 0.75, a decision in plaintiff's favor saves on error costs (again, assuming symmetry in errors): by deciding in plaintiff's favor, the factfinder is three times less likely to commit error as he is if he finds in defendant's favor. If the factfinder finds liability indeterminate, however, a decision in plaintiff's favor is just as likely to result in error as a decision

---

33. A consistent decision for plaintiffs would make $q_1 = 0$ and $q_2 = 0.5$, so that error costs would be $0.25QEC_2$.

34. A consistent decision for plaintiffs would make $q_1 = 0$ and $q_2 < 0.5$, so that error costs would be $< 0.25QEC_2$.

35. It is important to note that this assumes that there has been no a priori determination with regard to who is more likely right in the absence of any proof. In other words, the assumption is that $k = 0.5$. Where this assumption breaks down, justifiably there is a deviation from the default rule. *See infra* Part III.B.3.

for defendant (assuming $k = 0.5$), and error costs are constant regardless of the judgment. Given the symmetry of error costs, the default rule prefers to avoid the indeterminate suit and its attendant direct costs. Consistent decisions for defendants in the case of "ties" may discourage some suits at the margin and save on direct costs because a given plaintiff, as the party seeking to upset the status quo, knows he will succeed only if he persuades the factfinder that his position is *more* convincing than his defendant's. The default rule can thus be justified as economizing on direct costs by deterring some socially ambiguous claims at the margin.[36]

B. Deviations from the Default Rule: Affirmative Defenses

Economic analysis is also useful in explaining the law's departures from the default rule. Given the foregoing arguments, how can assignment of the burden of proof to defendant ever be justified? Why not assign the risk of nonpersuasion to the plaintiff on all issues?[37] This section offers three justifications for affirmative defenses.

1. Relative-cost-of-proof theory

The justifications advanced above for the default rule assumed symmetry in the relative cost of producing evidence. In such a world, the burden of proof functions only as a tiebreaker in the unlikely event that the factfinder is equally compelled by plaintiff's and defendant's versions of the issue (or in the absence of any proof on the issue). When defendant can produce core evidence on a relevant issue at a lower cost than plaintiff, however, assigning the burden of proof to the defendant can minimize the social loss function even when there is no "tie."[38] This

---

36. In most suits, this explanation will be unsatisfying. Litigation is driven by mutual optimism. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF THE LAW §§ 20.2, 21.5 (4th ed. 1992); Steven Shavell, *Suit, Settlement, and Trial: A Theoretical Analysis Under Alternative Methods for the Allocation of Legal Costs*, 11 J. LEGAL STUD. 55 (1982). An optimistic plaintiff is unlikely to be deterred from suit by the possibility that the factfinder may end up equally compelled by the defendant's case as by his own. This second rationale therefore seems convincing only where the burden of proof is likely to be dispositive. See, for example, the discussion of burdens of proof in employment discrimination cases, *infra* Part V.

37. Judge Posner poses this question in his 1973 article but does not offer an answer. *See* Posner, *supra* note 13, at 408-09.

38. To see why this is possible, but not inevitable, see the third theory justifying affirmative defenses in the burden of proof context, *infra* Part III.B.3.

section first analyzes the theoretical justifications for this function of the burden of proof, and then examines some burden allocations that can be explained by this theory.

a. Analytical foundations for the relative-cost theory. In the pleading context, relative costs depend on the nature of the parties' claims and not on the parties' "access" to relevant proof. The opposite is true in the burden of proof context—relative costs depend on access to proof, not on the nature of the parties' claims. Pleading rules usually eliminate relative cost differentials resulting from differences in the nature of the two parties' claims. Recall the example of the issue of illegality. Placing the burden of pleading on the plaintiff to show the general absence of any illegality in any provision of the contract would impose unnecessary costs on plaintiffs. As explained, pleading rules economize on investigation costs by assigning the burden of pleading illegality to defendants. Once defendant has met his burden of pleading illegality, however, there is no further economic basis for shifting the burden of proof. Assigning to defendant the burden of proving illegality cannot be justified by the fact that his version of the proof is more discrete or "easier" than plaintiff's.[39]

Relative access costs are thus the relevant variable in assigning the burden of proof under the relative-cost theory. There are at least two categories of issues where defendant's access to proof is likely to be superior. First are issues involving conduct of the defendant where the plaintiff may not have been involved. Estoppel is one example. A successful defense of estoppel requires a showing of detrimental reliance, i.e., that the party changed his position or did something detrimental to his position because of the other party's conduct.[40] Since defendants are likely to have superior access to proof of their own detrimental reliance, assigning to the defendant the burden of proving estoppel economizes on costs of proof.[41] The second category of issues for which defendant has greater access to proof involves interactions between the two parties for which the defendant has supe-

---

39. It is important to note, however, that this conclusion depends on the level of specificity required by the pleading rules. Liberal pleading rules may introduce the "ease of proof" concept into the burden of proof context. If pleading rules require only a very general allegation of illegality by defendant, for example, it would make sense to assign the burden of proving illegality to defendant because the pleading rules have failed to eliminate the cost differential resulting from the asymmetry in ease of proof.

40. *See* United States v. Hemmen, 51 F.3d 883, 892 (9th Cir. 1995).

41. Injury by fellow servant and duress fall in this category as well.

rior incentives to keep records of the transaction. Discharge in bankruptcy, license, payment, and release would fall in this category. Defendants are likely to have superior access to proof of these issues because they are more likely to have kept accurate records.[42]

Because of the differentials in access to proof, the allocation of burdens performs important economizing functions. Assigning the burden of proof to the party with lower relative costs of proof economizes on both direct costs and error costs. First, it economizes on direct costs by providing incentives to the "least-cost producer" of proof to produce the central "core" of evidence to be shared by both parties. This relative-cost theory depends on two assumptions. The first is that the party not assigned the burden of proof will do nothing (or at least spend relatively less on proof) until the party assigned the burden produces at least some proof. Thus, while in most cases plaintiff and defendant will each eventually seek to offer their own versions of the relevant proof, the party not assigned the burden of proof will have no incentive to invest in proof on that issue until the other party makes an initial expenditure on proof. The second assumption is that each party will be able to share at least some fraction of the proof produced by the other. In other words, there are some essential data and evidence that form the "core" of proof and are shared by the parties. This core must be generated by one of the parties, but need not be duplicated by the other. Where defendant's direct costs of producing "core" evidence are lower than plaintiff's, assigning the risk of nonpersuasion to the defendant decreases the direct cost of producing that evidence.

A simple numerical example helps to illustrate this feature of the burden of proof. Consider a contract action in which the key issue is whether defendant is entitled to a defense of duress. Assume that in the absence of other proof, plaintiff can convince the factfinder that there was no duress at a cost of $100, whereas defendant can convince the factfinder that there was duress at a cost of $10. In addition, either party can reduce his own costs by 50% if he relies on the evidence already advanced by the other party. Plaintiff's injury is worth $200 in damages. Also assume that if both plaintiff and defendant advance their

42. With other affirmative defenses, it seems clear that defendant's investigation costs will actually be higher than plaintiff's. The best examples are contributory negligence and assumption of risk. For an explanation of the law's treatment of these two issues as affirmative defenses, see *infra* Part IV.

versions of the proof, the court will find that there is a 75% likelihood that defendant acted under duress.

In this case, placing the burden on the defendant yields the same result (error costs are constant) but saves on direct costs. With the burden on the plaintiff, direct costs will total $105 and the court will issue a judgment for the defendant. Plaintiff will incur the $100 cost of advancing his proof because he knows that he will lose if he does not,[43] since in the absence of any proof, the burden of proof decides the issue in defendant's favor. Defendant has an incentive to incur the $5 cost of producing his proof only after plaintiff advances his proof.[44] Assigning the burden of proof to defendant yields the same result at a cost savings of $45. Defendant will incur the $10 of proving he acted under duress, and plaintiff will incur the $50 of advancing his proof.[45] The court will rule in defendant's favor.

Assigning the burden of proof to the party with lower direct costs may also decrease the frequency of legal error (q_1 or q_2). Consider the above scenario with two modifications: damages are worth $90, and the court will find it 75% likely that plaintiff deserves to recover if both parties advance their versions of the proof. With the burden on plaintiff, there will be no litigation. Plaintiff will prefer to forego the $90 recovery than to incur the $100 required to prove that defendant did not act under duress. On the other hand, if the burden of proof is assigned to defendant, defendant will advance his proof at a cost of $10, and plaintiff will advance his proof at a cost of $50. The court will decide in plaintiff's favor, and the probability of legal error will be reduced by two-thirds.[46]

---

43. This example, like all cases pursued to trial, depends on uncertainty. If plaintiff *knew* the strength of defendant's proof, he would never incur the $100 cost of proof. This scenario therefore depends on imperfect information and optimism. *See* POSNER, *supra* note 36, at §§ 20.2, 21.5; Shavell, *supra* note 36, at 65.

44. Of course, proof does not proceed entirely in two separate waves. In actual litigation, defendant does not wait for plaintiff to advance his entire case before investing in his version of the proof. To a certain extent, then, the sequential story told by this example is fictitious. But a large amount of sequential and responsive proof does occur throughout the course of litigation. In the above hypothetical, for example, plaintiff may postpone any substantial investment in disproving duress until defendant has offered some indication (through interrogatory responses or deposition testimony) of the basis for his claim of duress.

45. Or, in a world where the parties have perfect information about their opponent's proof *after* it is offered to the court, plaintiff will spend nothing on proof because he knows that the court will find in defendant's favor.

46. If the court were presented with 100 identical cases, a consistent decision for for defendants would result in 75 errors for every 100 decisions ($q_1 = 0.75$ and $q_2 = 0$),

b. Burden allocations based on relative-cost theory. The law is replete with examples of burden allocations based on asymmetries in relative costs of proof. Two illustrations of the theory are offered here: the doctrine of res ipsa loquitur in the law of torts and proof of copying in the law of copyright.

(1) Res ipsa loquitur. One straightforward example of the relative-cost theory is the doctrine of res ipsa loquitur in the law of torts. Under this doctrine, a plaintiff may pass his initial burden of proof to the defendant if three conditions are met: (1) the accident is of a kind that ordinarily does not occur in the absence of negligence; (2) the accident was caused by an instrumentality within the exclusive control of the defendant; and (3) the accident was not due to any voluntary action or contribution on the part of the plaintiff.[47]

Res ipsa loquitur is a conscious effort to economize on direct and error costs by allocating the burden to defendants in cases where their costs of proof are particularly low relative to plaintiffs. In cases where plaintiff makes no voluntary contribution and defendant is in exclusive control, plaintiffs' access costs will be enormous, even prohibitive. In medical malpractice cases where the accident takes place in the operating room, for example, plaintiffs are unconscious during the relevant period while defendants are recording relevant evidence about the procedure.[48] Assigning the burden of proving negligence to plaintiffs in such cases would increase direct costs and error costs. Error costs would increase because courts would be faced with inferior information to resolve the issue of negligence, and many justifiable suits would be deterred. Direct costs would increase because plaintiffs who are not deterred will have no personal information of the accident, and will have to mount their proof against the "professional conspiracy of silence"[49] by defendants.

(2) Proof of copying. A second example of the relative-cost theory comes from copyright law. The federal Copyright Act gives the owner of a copyright the exclusive right to "reproduce the copyrighted work."[50] The Act also gives the copyright owner a cause of action for any infringement of the exclusive right to

---

whereas a consistent decision for plaintiffs would produce only 25 errors per 100 decisions ($q_1 = 0$ and $q_2 = 0.25$).

47. *See, e.g.*, Ybarra v. Spangard, 154 P.2d 687 (Cal. 1944).

48. *See id.*

49. WILLIAM LLOYD PROSSER, SELECTED TOPICS ON THE LAW OF TORTS 346 (reprint ed. 1982) (1954).

50. Copyright Act of 1976, 17 U.S.C. § 106 (1994).

make copies.[51] Traditional "tests" might assign to the plaintiff the burden of proving copying since defendant's reproduction of the copyrighted work is "essential" to plaintiff's case, and "copying" is the "affirmative proposition" at issue.[52]

In order to prove infringement, however, a plaintiff need not offer direct proof of copying; copyright plaintiffs can shift the burden of proof to defendants upon a showing of "access" and "similarity" to the copyrighted work.[53] Proof of access and similarity gives rise to an inference of copying, and defendant then has the burden of proving that he did not copy the copyrighted work. This shift is an effort to economize on direct and error costs because of an asymmetry in costs of proof. "Because copiers are rarely caught red-handed,"[54] plaintiffs' costs of obtaining direct evidence of copying are high relative to defendants' costs of obtaining evidence that they did not copy. Defendants' access costs will often be lower because they may have evidence that they produced the work independently[55] or that they took portions of the work from the public domain.[56] Assigning the burden of proving copying to defendants thus economizes on direct and error costs.

At the same time, however, it would be inefficient to allow a plaintiff to hale anyone into court to require him to prove he did not copy plaintiff's copyrighted work. Such a rule would encourage frivolous litigation and would give plaintiffs an unduly favorable bargaining position in settlement negotiations. The economics behind the default rule require that the plaintiff carry some initial burden to justify initiating costly litigation. Proof of "access" and "similarity" attempt to serve that screening function while balancing the need to economize on direct costs of proof.

2. *Economizing on error costs: a priori judgments as to the magnitude of errors*

A second justification for departing from the default rule is that such departure reduces error costs, not by minimizing the frequency of legal error (q_1 or q_2), but by favoring legal errors

---

51. *See* 17 U.S.C. § 501(b).

52. *See* discussion *supra* accompanying notes 1-3.

53. *See* Gaste v. Kaiserman, 863 F.2d 1061, 1066 (2d Cir. 1988).

54. *Id.*

55. For example, defendants might be able to present handwritten notes of prior "drafts" of their work or evidence of time spent researching and preparing the work.

56. *See* Sheldon v. Metro-Goldwyn Pictures Corp., 81 F.2d 49 (2d Cir. 1936).

considered to have a lesser magnitude (EC_1 or EC_2). Placing the burden of proof on a defendant may be justified on the ground that an erroneous decision in defendant's favor is more costly than an erroneous decision for plaintiff.

a. Analytical foundations for the error-cost theory. Where liability is indeterminate, a decision in defendant's favor is just as likely to be erroneous as it is likely to be accurate (assuming $k = 0.5$). As discussed, one justification for the default rule was that a decision in plaintiff's favor involves direct costs not associated with a decision for defendant.[57] The added *direct* costs presented a basis for preferring a decision for defendant in cases of indeterminacy. Departures from the default rule can be justified on similar grounds: a decision for defendant may involve *error* costs not associated with a decision for plaintiff. Although a decision for defendant may result in an equal number of errors as a decision for plaintiff, an error in plaintiff's favor might be of a lesser magnitude, i.e., $EC_1 > EC_2$. In such cases, the loss function can be minimized by shifting the risk of nonpersuasion to the defendant: since $EC_1 > EC_2$, L is minimized by favoring plaintiffs at the margin.

It is important to note that shifting the burden of proof to defendant will encourage suits at the margin. While one of the purposes of the default rule is to discourage indeterminate suits, here the socially optimal goal is to encourage such suits. Shifting the burden of proof to defendant lowers plaintiff's direct costs of pursuing litigation, thereby promoting litigation.[58]

b. Applications of the error-cost theory. The error-cost theory corresponds to the notion expressed by some courts that certain burden allocations stem from substantive or policy determinations to discourage "disfavored" claims.[59] Some courts expressly acknowledge that policy concerns impel their burden allocation, taking their cue from legislative expressions of policy[60]

---

57. *See supra* text accompanying notes 27-31.

58. Plaintiff will sue if $P^A D - DC^A > 0$, where P^A = probability plaintiff wins, D = plaintiff's damages and other benefits from suit, and DC^A is plaintiff's direct costs of suit. Shifting the burden to defendant increases P^A or lowers DC^A (given P^A) and thus makes it more likely that $P^A D - DC^A > 0$.

59. This phrase originated with CHARLES E. CLARK, HANDBOOK OF THE LAW OF CODE PLEADING § 96, at 609 (2d ed. 1947) (discussing "disfavored allegations" and "disfavored defenses").

60. *See In re* Regional Rail Reorganization Proceedings, 421 F. Supp. 1061, 1073 (Regional Rail Reorg. Ct. 1976) (assigning burden of proof to railroad because railroad's position with regard to pension plan was "disfavored" under the congressional purpose of protecting employees of bankrupt railroads).

or from their own views of the equities at stake. Because reasonable minds can differ as to the relative merits of these kinds of policy concerns, an inherent analytical difficulty with the error-cost theory is its subjectivity. Indeed, where the law diverges from the default rule based on policy judgments about error costs, that allocation is unlikely to be stable. Allocations based on the error-cost theory are therefore those most likely to be disputed and most likely to change over time.[61]

(1) Defamation and disparagement. One important example of the error-cost theory comes from the law of defamation. At common law, all defamatory statements were presumed false—defendants had the burden of proving the truth of all allegedly defamatory statements.[62] This common-law allocation was rooted in an implicit judgment that an error in defendant's favor was of greater magnitude than an error in plaintiff's favor. As one commentator put it, the burden allocation was based on "a tender regard for reputations."[63] Under the common-law conception of the relative error costs, EC_1 was $> EC_2$ in our loss function because an erroneous decision in plaintiff's favor involved only a finding of defamation and a transfer payment for damages, whereas an erroneous decision in defendant's favor resulted in the wrongful erosion of plaintiff's good reputation. In short, the common law valued the individual's interest in reputation more highly than it did the defendant's interest in his money.

In 1986, however, the United States Supreme Court voided the common-law allocation of the burden of proving fault in defamation cases where the plaintiff is a public figure or where the speech is of public concern. In *Philadelphia Newspapers, Inc. v. Hepps*,[64] the Court acknowledged the common-law interest in "'judicial redress of libelous utterances,'"[65] but introduced a more compelling policy concern, located in the First Amendment and first articulated by the Court in *New York Times Co. v.*

---

61. The most familiar example in recent memory is the swift legislative response to the Supreme Court's allocation of the burden of proving business necessity in *Wards Cove*. *See* 42 U.S.C. § 2000e-2(k)(1)(A)(i) (1994).

62. *See* Atwater v. Morning News Co., 34 A. 865, 868 (Conn. 1896); Bingham v. Gaynor, 96 N.E. 84 (N.Y. 1911).

63. W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 116, at 841 (5th ed. 1984).

64. 475 U.S. 767 (1986).

65. *Id.* at 774 (quoting Time, Inc. v. Firestone, 424 U.S. 448, 456 (1976)).

Sullivan,[66] in providing "'"breathing space"' for true speech on matters of public concern."[67] Thus, the Court held that the constitutional policy of "ensur[ing] that true speech on matters of public concern is not deterred"[68] outweighed the common-law policy favoring defamation plaintiffs, and assigned the burden of proving fault to the plaintiff where the speech is of public concern or where the plaintiff is a public figure.[69] To the majority of the Court, $EC_2 > EC_1$ because an erroneous decision in plaintiff's favor would deter true speech on matters of public concern.

Justice Stevens, joined by three other justices, dissented in *Philadelphia Newspapers*. The dissent's disagreement with the majority cut to the heart of the policy dispute underlying the differing burden allocations at common law and established by the majority of the Court. Justice Stevens criticized the majority for "attaching no weight to the State's interest in protecting the private individual's good name"[70] and preferred to defer to the state's decisions to adhere to the "common-law presumption."[71] Where the First Amendment already required the plaintiff to prove either negligence or "actual malice,"[72] Justice Stevens saw no constitutional basis for assigning plaintiff the additional burden of proving falsity. To Justice Stevens and the dissenters, EC_2 was *not* $> EC_1$ so long as the plaintiff was required to prove fault; the *New York Times* fault requirement provided the necessary "breathing space" to fulfill the constitutional policy of enabling speech on matters of public concern. As noted above, the subjective nature of the substantive policy judgment at issue explains the lack of agreement among members of the Court.

---

66. 376 U.S. 254 (1964).

67. *Philadelphia Newspapers*, 475 U.S. at 778 (quoting *New York Times*, 376 U.S. at 272 (quoting NAACP v. Button, 371 U.S. 415, 433 (1963))).

68. *Id.* at 776.

69. *See id.* at 778.

70. *Id.* at 781 (Stevens, J., dissenting).

71. *Id.* at 780 (Stevens, J., dissenting) (citing Elliott v. Roach, 409 N.E.2d 661, 681 (Ind. Ct. App. 1980)); *see, e.g.*, Trahan v. Ritterman, 368 So. 2d 181, 184 (La. Ct. App. 1979); Madison v. Yunker, 589 P.2d 126, 129-30 (Mont. 1978); Rogozinski v. Airstream by Angell, 377 A.2d 807, 814 (N.J. 1977).

72. *New York Times* requires a showing of "actual malice," i.e., knowledge or recklessness as to falsity, wherever the plaintiff is a public figure. *See* 376 U.S. at 279-80. *Gertz v. Robert Welch, Inc.* requires a showing of fault where the plaintiff is not a public figure but the speech is on a matter of public concern. *See* 418 U.S. 323, 347 (1974). Thus, constitutional policy concerns are implicated, and common-law burden allocations affected, wherever the speech at issue is about a public figure or about a matter of public concern. Otherwise, common-law policy considerations favoring defendants are left intact.

An interesting comparison to defamation is provided by the law of product disparagement. Unlike defamation, the common-law rule for disparagement assigned the burden of proving falsity to the plaintiff.[73] This difference between allocations in defamation and disparagement can be explained by the error-cost theory. Consider a purely private defamation suit where the plaintiff is not a public figure and the speech is on a matter of private concern, and where the Constitution leaves intact the common-law rule assigning the burden of proving falsity to defendant. Compare that suit to a product disparagement suit, where the plaintiff has the burden of proving falsity. It is difficult to explain the difference between these two common-law allocations on the basis of relative cost of proof. There does not appear to be any reason why the disparagement plaintiff might have access superior to the defamation plaintiff.[74]

The error-cost theory, however, does provide an explanation of the differing allocations of burdens of proof. Type II errors (judgments for undeserving defendants) in disparagement suits are more costly than Type II errors in private defamation suits because of the third-party interests involved in disparagement suits. This difference explains the common-law rules favoring defendants in disparagement suits while favoring the plaintiffs in defamation suits. With the private defamation suit, $EC_1 > EC_2$ because of the interest in preserving reputations, as explained above. There is no reason to favor a decision for defendants where the underlying dispute implicates no third-party interests. With the disparagement suit, however, a decision for plaintiff discourages the production of consumer information at the margin,[75] so that EC_2 may be $> EC_1$. The common-law rule assigning the burden to plaintiffs in disparagement cases thus favors defendants for reasons similar to the reasons advanced by the ma-

---

73. *See* Systems Operations, Inc. v. Scientific Games Dev. Corp., 555 F.2d 1131, 1141 (3d Cir. 1977); KEETON ET AL., *supra* note 63, § 128, at 967.

74. Indeed, in both cases it would seem that the plaintiff would have superior access to general information about himself, whereas the defendant might have superior access to information that would specifically corroborate the defamatory or disparaging statement made. One possible explanation under the relative-cost theory would depend on the relative "ease" of the proof. Perhaps it could be established that statements in disparagement actions are relatively specific, and thus relatively "easy" to prove false, whereas statements in defamation actions are relatively general, and thus more costly to prove false.

75. *See* Testing Sys., Inc. v. Magnaflux Corp., 251 F. Supp. 286, 288-89 (D. Pa. 1966) (explaining that the disparagement rule is based on concern for consumer information).

jority in *Philadelphia Newspapers*. Assigning the burden of proving falsity to plaintiffs provides the necessary "breathing space" for the production of consumer information in the same way it protects speech on matters of public concern.

(2) Different standards of proof. The error-cost theory can also explain the existence of standards of proof other than the preponderance standard. The Supreme Court has explained that the standard of proof functions to "'instruct the factfinder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for a particular type of adjudication.'"[76] This Article's economic model suggests that the proper "degree of confidence" should depend on the relative magnitudes of legal errors. The basic three-tiered menu of standards available in most jurisdictions—"preponderance" in civil cases, "beyond a reasonable doubt" in criminal cases, and "clear and convincing evidence" in some special cases—is consistent with this model.

The preponderance standard is proper where errors are symmetrical. Where $EC_1 = EC_2$, error costs are minimized by reducing the total number of errors. The preponderance standard achieves that goal by telling the court to decide in favor of the party who convinces the court that his version of the facts is "more likely than not." So long as the court is more than 50% confident that plaintiff should prevail, it will decide in plaintiff's favor, thus minimizing the number of legal errors. Civil cases are the paradigm for symmetrical error costs. Error costs are symmetrical in civil cases because society values equally an improperly coerced payment of money and an improper windfall.[77] In other words, the preponderance standard is proper in civil cases because an error in plaintiff's favor is no more costly than an error in defendant's favor—"society has a minimal concern with the outcome of such private suits."[78]

Criminal cases, on the other hand, are the paradigm for asymmetrical error costs. In criminal cases, $EC_2 > EC_1$ because criminal remedies differ in kind from civil remedies. Civil remedies are largely financial and involve no more than a transfer

---

76. Addington v. Texas, 441 U.S. 418, 423 (1978) (quoting *In re* Winship, 397 U.S. 358, 370 (1970) (Harlan, J., concurring)).

77. *See Winship*, 397 U.S. at 371 (Harlan, J., concurring) ("In a civil suit between two private parties for money damages, for example, we view it as no more serious in general for there to be an erroneous verdict in the defendant's favor than for there to be an erroneous verdict in the plaintiff's favor.").

78. *See Addington*, 441 U.S. at 423.

payment between private parties. Criminal remedies, on the other hand, involve penalties such as imprisonment that result in social loss. Criminal remedies also differ from civil remedies in that unsuccessful criminal defendants suffer a deadweight loss that exceeds any "gain" to the prosecution. While the unsuccessful civil defendant's losses mirror the plaintiff's gains, unsuccessful criminal defendants lose most of their freedoms and liberties without any corresponding gain to the prosecution.

These error cost asymmetries require a standard of proof greater than a preponderance standard. In criminal cases, the cost of an erroneous decision in the prosecutor's favor (EC_2) greatly exceeds the cost of an error in defendant's favor (EC_1), so that the social loss function is minimized by requiring plaintiff to prove his case at > 50% confidence level, even though total errors will not be minimized. Consider a murder prosecution where $EC_2 = 100$ and $EC_1 = 10$. If the prosecutor convinces the court that it is 51% likely that defendant committed the murder, a conviction will minimize the number of legal errors—51% of the time the decision for the prosecution will be correct. But this conviction is not economically justifiable. The conviction imposes error costs of 49 ($L = 0.51(0)(1) + 0.49(1)(100)$). An acquittal would increase the total number of errors by two percent, but would decrease error costs to 5.1 ($L = 0.51(1)(10) + 0.49(0)(100)$). Raising the burden of proof in this case thus economizes on error costs.[79] In fact, with these hypothetical numbers, the court should convict defendant only after it is about 91% convinced of guilt.[80]

Consistent with this theory, the Supreme Court, in *In re Winship*,[81] cited the asymmetry in legal errors in criminal cases in support of its holding that the Due Process Clause requires that every factor necessary to convict a criminal defendant be proven beyond a reasonable doubt. Justice Harlan's concurring

---

79. One might be tempted to conclude from this analysis that the proper standard should increase with the seriousness of the offense. A more serious offense, the argument would go, will result in a longer sentence, and hence a higher cost of erroneously imprisoning a defendant. But this argument is short-sighted. It ignores the fact that for more serious offenses, the cost of erroneous acquittals also increases, in the form of general and specific deterrence costs.

80. Where the factfinder is 91% certain, the expected error costs associated with an erroneous conviction (0.91(0)(10) + 0.09(1)(100) = 9) are substantially equal to the expected error costs of an erroneous acquittal (0.91(1)(10) + 0.09(0)(100) = 9.1).

81. 397 U.S. 358 (1970).

opinion puts the court's argument in terms that fit an economic model:

> In a criminal case . . . we do not view the social disutility of convicting an innocent man as equivalent to the disutility of acquitting someone who is guilty. . . .
>
> In this context, I view the requirement of proof beyond a reasonable doubt in a criminal case as bottomed on a fundamental value determination of our society that it is far worse to convict an innocent man than to let a guilty man go free.[82]

In other words, Justice Harlan recognizes that $EC_2 > EC_1$ in criminal cases. The Court's imposition of the "beyond a reasonable doubt" standard can thus be seen as an attempt to minimize error costs by favoring the less costly error of deciding for the defendant.

Perceived asymmetries in noncriminal cases sometimes give rise to the imposition of the third basic standard, the "clear and convincing evidence" standard. The clear and convincing standard is an intermediate standard that generally applies where EC_2 is greater than EC_1 because "[t]he interests at stake . . . are deemed to be more substantial than mere loss of money."[83] In areas like libel, deportation, and denaturalization,[84] for example, a standard greater than preponderance is deemed proper because of the perceived hardships resulting from an error in plaintiff's favor. The asymmetry is less pronounced than in criminal cases, but nevertheless requires a greater confidence level than the "more likely than not" standard appropriate with perfect symmetry.

3. *Economizing on error costs: a priori judgments as to the fraction of cases where defendant is truly liable*

A third justification for departing from the default rule is that it reduces error costs by favoring plaintiffs in cases in which defendants are statistically likely to be liable (i.e., $k > 0.5$). In terms of the economic model, assigning the burden of proof to defendants will minimize the loss function, assuming $EC_1 = EC_2$ and symmetrical access costs. If we knew that a given class of

---

82. *Id.* at 372 (Harlan, J., concurring).

83. *Addington*, 441 U.S. at 424.

84. *See* California *ex rel.* Cooper v. Mitchell Brothers' Santa Ana Theater, 454 U.S. 90, 93 (1981) (per curiam) (discussing instances where the clear and convincing evidence standard applies).

defendants were statistically likely to be liable 75% of the time, for example, then assigning the burden of proof to those defendants would minimize our loss function. Assuming indeterminate liability, a consistent decision for plaintiffs would result in three times fewer errors than a consistent decision for defendants.[85]

Although the theoretical bases for this justification are strong, it is unlikely to have any useful application in practice. There are two practical problems with this theory. First, it requires a reliable estimate of k, and there is unlikely to be any degree of consensus on that estimate. Second, and more importantly, k seems likely to equal 0.5 over time because of the inevitable uncertainty involved in suits brought to litigation.[86] Defendants will settle when their liability is clear, and plaintiffs will not bring suit when they have little chance of prevailing. Thus, suits brought to trial will generally be hotly disputed. In the face of this uncertainty, there is no reason to predict that litigated suits would result in more decisions for defendants than for plaintiffs—k will approach 0.5 over time.[87]

IV. EXPLAINING THE IMPERFECT CONGRUENCE BETWEEN BURDENS OF PLEADING AND BURDENS OF PROOF

In most cases, the law assigns the burden of proof to the same party that has the burden of pleading. Plaintiffs generally have the burden of proof on those elements they must plead, as do defendants.[88] Commentators have noted two principal exceptions to this rule. First, the Federal Rules of Civil Procedure require defendants to plead payment,[89] yet the burden of proving nonpayment is often assigned to the plaintiff.[90] Second, in federal court, defendants must plead contributory negligence and assumption of risk, but state law often places the burden of proof on plaintiff.[91]

---

85. A consistent decision for plaintiffs would result in total error costs per case of $(0.75)(0)(EC_1) + (0.25)(0.5)(EC_2) = 0.125EC_2$. A consistent decision for defendants would result in total error costs per case of $(0.75)(0.5)(EC_1) + (0.25)(0)(EC_2) = 0.375EC_1$.

86. *See* Shavell, *supra* note 36, at 55.

87. One caveat is the case where liability is consistently indeterminate or unprovable. In such instances, defendants may know they are truly liable yet have no incentive to settle, and k may exceed 0.5 over time. For a discussion of why this might be the case in Title VII suits, sea *infra* note 105.

88. *See* CLARK, *supra* note 59, at 607; JAMES & HAZARD, *supra* note 3, § 7.8, at 322.

89. *See* FED. R. CIV. P. 8(c).

90. *See* Thomas v. Tygart, 6 S.W.2d 827 (Ark. 1928).

91. *See* Palmer v. Hoffman, 318 U.S. 109, 117 (1943); MCCORMICK ON EVIDENCE,

Each of these examples can be explained by distinguishing the economic functions of pleading rules from the economic functions of burdens of proof. Payment, contributory negligence, and assumption of risk are all issues that fulfill the justifications for departing from the default rule in the pleading context, but not in the burden-of-proof context. With each of these issues, it makes economic sense to assign the burden of pleading to defendant because these issues are both unlikely to arise and are more narrowly focused than their logical counterparts. It is less costly for defendant to allege a discrete instance of payment or specific conduct amounting to contributory negligence or assumption of risk than it is for plaintiff to plead the absence of payment, contributory negligence, or assumption of risk.

At the same time, however, the cost-minimization model argues in favor of preserving the default rule assignment of the burden of proof on these issues. A sufficiently particularized allegation by defendant of payment, contributory negligence, or assumption of risk fully performs the "focusing" function envisioned by the probability and ease of proof theories—defendant's pleadings focus the litigation on a particular instance of payment or on specific conduct allegedly amounting to contributory negligence or assumption of risk.[92] From this point, each litigant will offer his version of the facts of the particular instance of payment (or of the specific conduct at issue). Shifting the burden of proof is justifiable only if (1) defendant has superior access to proof on these issues,[93] (2) Type I errors on these issues are more costly than Type II errors (i.e., $EC_1 > EC_2$),[94] or (3) there is some reason to believe that defendants are statistically unlikely to prevail on such issues ($k > 0.5$).[95] None of these justifications seems applicable in the case of payment, contributory negligence, or assumption of risk. There appears to be no reason to favor Type II errors or to conclude that defendants are unlikely

---

supra note 1, § 337.

92. The probability theory introduced in the pleading context plays no role in the burden of proof context. Probabilities are relevant in the pleading context because of the signaling function served by pleading rules. By assigning the burden of pleading unlikely issues to defendant, pleading rules obviate the need to even raise these issues in the vast number of cases. Absent other cost considerations (such as a judgment that $EC_1 > EC_2$), there is no reason to handicap a claim on the basis that it is statistically unlikely to be raised by the parties.

93. *See supra* Part III.B.1.

94. *See supra* Part III.B.2.

95. *See supra* Part III.B.3.

to prevail. Furthermore, access arguments are plausible only on the issue of payment, where it is difficult to make any a priori judgment about which party would have superior access to proof. Presumably, both parties would keep records of such transactions, although defendants might have better incentives to keep records. In any event, the comparison is far from clear, and deference to the considerations underlying the default rule may explain the lack of agreement on this issue.[96] In the case of contributory negligence and assumption of risk, it would be difficult to argue that defendant has superior access to proof of plaintiff's conduct. Accordingly, many jurisdictions assign the burden of pleading these issues to defendants, but leave the burden of proof to the plaintiff.

V. PUTTING THE THEORY TOGETHER: ALLOCATING THE BURDENS ON THE ISSUE OF BUSINESS NECESSITY IN TITLE VII CASES

Up to this point, this Article has analyzed separately the different cost considerations involved in allocating burdens of pleading and burdens of proof, offering illustrations of allocations that seem to be based on different theories of cost minimization. In practice, however, the various cost considerations cannot be analyzed in isolation, but must be considered together. This Part attempts to bring together each of the relevant cost considerations in the context of the decision of where to allocate the burdens on the issue of "business necessity" in Title VII cases. This application is appropriate both because of the currency of the debate on this issue,[97] and because the role that burdens play in deciding often unprovable questions of intent in the employment context tends to sharpen an otherwise muted debate.[98]

---

96. Some jurisdictions require plaintiff to plead nonpayment of an obligation, but assign the burden of proof to defendant. *See* Thomas v. Tygart, 6 S.W.2d 827, 827-28 (Ark. 1928).

97. In *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 660 (1989), the Supreme Court assigned the burden of proving the absence of business necessity to plaintiffs. Many courts and commentators had interpreted previous Supreme Court decisions as assigning the burden to defendants. *See* Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975). The 1991 Civil Rights Act restored the pre-*Wards Cove* assignment. *See* 42 U.S.C. § 2000e-2(k)(1)(A)(i) (1994).

98. *See* Candace S. Kovacic-Fleischer, *Proving Discrimination After* Price Waterhouse *and* Wards Cove: *Semantics as Substance*, 39 AM. U. L. REV. 615, 649 (1990) (explaining how "semantic disagreements" among members of the Court have affected burden of proof issues in discrimination cases).

The economic model developed in Part I provides a concrete and consistent framework for evaluating the decision of where to allocate burdens of pleading and proof. By focusing on the various cost considerations at issue, the model permits an analysis of burdens that is less susceptible to the manipulation and mutability of the conventional rhetoric used by the courts. For the pleading decision, probability and relative costs of pleading are central. For the burden of proof decision, we must consider the economic functions served by the default rule and attempt to balance these considerations against any potential savings (due to relative costs of proof and relative error costs) that could arise by shifting the burden to defendants.

Before addressing the controversial question of where to allocate the burden of proof of business necessity, we will examine the cost considerations at stake at the pleading stage. Assigning the burden of pleading business necessity is not a controversial decision. Although the probability of the issue arising in a disparate impact case is high, relative pleading costs counsel in favor of assigning the burden to the defendant. Pleading costs are minimized by assigning the burden of pleading to defendant because verifying the existence of a particular business justification is easier than verifying the absence of any conceivable business necessity. Thus, it is not surprising that the courts consistently have assigned the burden of pleading business necessity to defendants.[99]

The cost considerations in the decision of where to allocate the burden of proof are less clear. Before evaluating the cost factors that might counsel in favor of treating business necessity as an affirmative defense, it is important to recall the economics of the default rule. Sound economic bases favor requiring plaintiffs to bear the burden of proving all issues necessary to prevail in a Title VII case. The default rule conserves the direct litigation costs incurred after the liability determination, including remedy construction costs and enforcement costs. In the Title VII context, such costs are often substantial; a judgment for plaintiff requires an elaborate inquiry into whether and to what extent an award of "front pay" is appropriate or whether the plaintiff should be reinstated to his former position.[100] A finding of liability also produces significant external costs by creating incentives

---

99. *See, e.g.*, James v. Stockham Valves & Fittings Co., 559 F.2d 310, 330 (5th Cir. 1977).

100. *See* 42 U.S.C. § 2000e-5(g).

for employers to restructure their hiring practices and even to impose numerical quotas. Assigning the burden of proof to plaintiffs avoids these costs in cases of indeterminate liability.

The economic loss function favors retaining the default rule unless shifting the allocation to defendants would produce cost savings that outweigh the above savings in direct costs. Initially, cost of access considerations seem to favor assigning the burden to defendants: employers clearly have superior access to records of their own employment practices. Relative cost considerations thus seem to favor the position of the *Wards Cove* dissenters. On the other hand, existing asymmetries in direct costs are minimized in disparate impact cases for two reasons. First, as the *Wards Cove* majority noted, liberal discovery rules give plaintiffs wide access to employer's records, and the Uniform Guidelines on Employee Selection Procedures require qualifying employers to maintain detailed records of their employment practices.[101] Second, the law assigns the burden of pleading as well as a burden of "production" to defendants,[102] which eliminates any cost asymmetries resulting from the "relative-cost-of-proof" considerations mentioned above. With these mechanisms in place, defendants may no longer have any significant advantage in the relative cost of proof; therefore, there does not appear to be a cost-of-proof basis for departing from the default rule.

Error cost considerations also favor retaining the default rule. Error costs are likely to be high in employment discrimination cases because "discrimination" is difficult to prove with certainty. Uncertainty means that the probability of error (q_1 or q_2) will be high, so that error cost considerations play an especially important role in this context.

The magnitude of the cost of Type II errors (EC_2) in this context is substantial for employers. First, charges of discrimination, if proved, "carry an enormously stigmatizing effect."[103] Second, such charges allow defendants to be "haled into court and forced to engage in the expensive and time-consuming task of defending the 'business necessity' of the methods used to select the other members of his work force," and leave defendants with the adoption of racial quotas as "the only practicable option."[104]

---

101. *See Wards Cove*, 490 U.S. at 657 (citing 29 C.F.R. § 1607.1 (1988)).

102. *See id.* at 660. The burden of production requires the defendant to "articulate" a valid business justification.

103. Harris v. Marsh, 679 F. Supp. 1204, 1212 (E.D.N.C. 1987).

104. *Wards Cove*, 490 U.S. at 652.

In light of these costs, it is difficult to argue that Type I errors are more costly than Type II errors (i.e., that $EC_1 > EC_2$), and even more difficult to establish that the difference in error costs is sufficient to justify foregoing the cost savings inherent in the default rule. Thus, with all due respect to the 1991 Congress and to the *Wards Cove* dissenters, the economic model developed herein suggests that the burden of proof as to business necessity should remain with the plaintiffs.[105]

VI. CONCLUSION

The economic model developed in this Article cannot provide an easy answer as to how to balance the cost considerations identified above. Because of the varied and sometimes conflicting nature of the variables, this Article does not provide any universal maxim for the assignment of legal burdens in all cases. What it does offer, more modestly, is a compendium of the various factors that enter into the economic calculus in allocating burdens of pleading and burdens of proof, thereby shifting the focus away from a hollow, semantic debate about "affirmative" propositions and "essential" elements towards a concrete balance of social costs.

According to this model, the burden of pleading should ordinarily be assigned to plaintiff to preserve the clarity of the signaling function. Departures from this default rule are economically justifiable where the issue is unlikely to arise or where defendant's version of the issue is narrower and thus easier to corroborate. Various direct and error cost considerations also support the default rule assigning the burden of proof to plaintiffs. Affirmative defenses may make economic sense where defen-

---

105. As discussed in Part III.B.3, another economic consideration that focuses on error costs is the fraction of cases where defendants are truly liable—k in our loss function. If the majority of defendants are liable, then assigning the burden of proving business justification to defendants would reduce error costs (all else being equal). Those who advocate assigning the burden to defendants would argue that discrimination is common, and that $k > 0.5$. Because discrimination is difficult to prove, the employment discrimination context is one area where k may not equal 0.5 over time. If plaintiffs are unable to prove discrimination, defendants will have no incentive to settle, and the fraction of defendants who are truly liable may consistently exceed 0.5. If so, assigning the burden of proof to defendants will minimize error costs. The problem with this argument is that it relies on an empirical proposition (about the pervasiveness of discrimination) that is at best controversial. Discrimination may well be a statistical rarity. *See* RICHARD A. EPSTEIN, FORBIDDEN GROUNDS (1992) (arguing that discrimination is an economically unsound practice and is thus eliminated by market forces).

dant's relative cost of proof is lower, where Type I errors are more costly than Type II errors ($EC_1 > EC_2$), or where defendants are likely to be truly liable the majority of the time ($k > 0.5$). By cataloguing the various cost considerations that may affect the allocation decision, the model provides a coherent framework superior to any doctrine currently used by the courts.