THE TELLTALE SIGN OF DISCRIMINATION: PROBABILITIES, INFORMATION ASYMMETRIES, AND THE SYSTEMIC DISPARATE TREATMENT THEORY

Jason R. Bent*

The systemic disparate treatment theory of employment discrimination is in disarray. Originally formulated in United States v. International Brotherhood of Teamsters, the systemic disparate treatment theory provides plaintiffs with a method for creating an inference of unlawful discriminatory intent if plaintiffs can first present sufficient statistical evidence establishing that the employer was engaged in a "pattern or practice" of discrimination. While the Court and scholars have recently given substantial attention to the disparate impact theory, they have not adequately analyzed the contours of the systemic disparate treatment theory. For example, there are currently disputes about whether the systemic disparate treatment theory can be utilized by plaintiffs in (1) ADEA cases, (2) ADA cases, (3) hostile work environment cases, and (4) private individual (non-class) discrimination cases brought under any of the antidiscrimination statutes. The development of a coherent approach to systemic disparate treatment law is becoming increasingly important. Since 2006, the EEOC has made pursuit of its "Systemic Initiative" a top priority in its enforcement strategy. Thus, difficult questions about the application of the systemic disparate treatment theory will continue to surface.

This Article takes a first step toward bringing order to the systemic disparate treatment theory. Drawing upon economic models of optimal allocation of burdens of proof in civil litigation and an application of Bayesian probability analysis, this Article argues that the proper extent of the systemic disparate treatment theory depends on three critical considerations: (1) the background prevalence (or prior probability) of the type of discrimination alleged; (2) the relative strength of the evidentiary signal provided by the proffered statistical evidence; and (3) the parties' relative access to evidence on the element of discriminatory intent. Applying the principles developed here, this Article concludes that the Teamsters systemic disparate treatment theory should be available as a method of proof for plaintiffs in ADEA, ADA, and individual (non-class) cases, but not in hostile work environment cases.

* Attorney, Smith & Bent, P.C.; Assistant Professor of Law, Stetson University College of Law (Appointment commencing Fall 2011); J.D., University of Michigan (2000). An earlier version of this paper was presented at the Seton Hall Law School Fourth Annual Employment and Labor Law Scholars Forum, on January 22–23, 2010. I am grateful for the many helpful suggestions offered by the Forum participants, including Roberto Corrada, Timothy P. Glynn, Tristin K. Green, Charles A. Sullivan, Susan P. Sturm, Rebecca Hanner White, Michael J. Zimmer, Charlotte Garden, Stephen Rich, and Hina B. Shah, as well as valuable comments received from the Honorable Joan B. Gottschall, Shaina Bent, Ramona L. Paetzold, J.J. Prescott, Michael Shapiro, Michael D. Smith, Steven L. Willborn, and Robert Vischer.

> *"Statistics showing racial or ethnic imbalance are probative in a case such as this one only because such imbalance is often a telltale sign of purposeful discrimination; absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired."*[1]

The systemic disparate treatment theory may soon reclaim center stage in employment discrimination law. Until very recently, the disparate impact theory of discrimination received all the attention. The Supreme Court considered the legality of employer actions meant to avoid disparate impact liability in the high-profile *Ricci v. DeStefano* case.[2] Scholars and jurists have likewise noted and analyzed potential tension between the disparate impact theory and the Constitutional requirement of equal protection.[3] Because the disparate impact doctrine raises politically volatile issues including the use of racial quotas, testing biases, and affirmative action, it has been a source of controversy since its inception.[4]

---

1. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 340, n.20 (1977). A leading treatise on employment discrimination adds: "The converse of this is that a substantial departure from what is expected, absent discrimination, is so improbable that the trier of fact should conclude, at least prima facie, that discrimination explains the disparity." 1 CHARLES A. SULLIVAN ET AL., EMPLOYMENT DISCRIMINATION: LAW AND PRACTICE § 3.04[A], at 175 (3rd ed. 2002).

2. *See* Ricci v. DeStefano, 129 S. Ct. 2658 (2009) (considering whether a public employer could disregard test scores and decline to make promotions, out of concern for a potential disparate impact problem, without running afoul of Title VII's prohibition on disparate treatment); Charles A. Sullivan, Ricci v. DeStefano*: End of the Line or Just Another Turn on the Disparate Impact Road?*, 104 NW. U. L. REV. 201 (2009) (contending that Ricci represents only the latest salvo in an ongoing battle between Congress and the Supreme Court on the disparate impact theory).

3. *See Ricci*, 129 S. Ct. at 2681–82 (Scalia, J., concurring) ("I join the Court's opinion in full, but write separately to observe that its resolution of this dispute merely postpones the evil day on which the Court will have to confront the question: Whether, or to what extent, are the disparate-impact provisions of Title VII of the Civil Rights Act of 1964 consistent with the Constitution's guarantee of equal protection?"); Richard A. Primus, *The Future of Disparate Impact*, 108 MICH. L. REV. 1341 (2010); Richard A. Primus, *Equal Protection and Disparate Impact: Round Three*, 117 HARV. L. REV. 493 (2003). The Supreme Court has also recently heard a number of cases involving more peripheral issues relating to disparate impact liability. *See* Lewis v. City of Chicago, 130 S. Ct. 2191 (2010) (considering the timeliness of discrimination charges in a disparate impact case involving the repeated use of test scores in making municipal hiring decisions over a number of years); Meacham v. Knolls Atomic Power Lab., 128 S. Ct. 2395, 2401–02 (2008) (determining the appropriate allocation of the burdens of proof for the affirmative defenses to disparate impact claims under the Age Discrimination in Employment Act of 1967, 42 U.S.C. §§ 621–634 (2006) ("ADEA")); Smith v. City of Jackson, 544 U.S. 228, 243 (2005) (determining that the disparate impact theory is available to plaintiffs under the ADEA).

4. The disparate impact theory was first articulated by the Court in *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), and has been one of the most controversial aspects of

But the Supreme Court's decision in the *Wal-Mart v. Dukes* case in June 2011 has the potential to revive interest in the disparate impact theory's cousin—the systemic disparate treatment theory—which implicates many of the same issues in a different theoretical setting.[5] While attention was focused on the disparate impact theory, a number of disputes were brewing in the lower courts regarding the proper interpretation, extent, and application of the systemic disparate treatment theory. These issues demonstrate fundamental disagreements and misunderstandings at the core of systemic disparate treatment law. This Article will take a first step toward bringing order to the systemic disparate treatment theory.

In *International Brotherhood of Teamsters v. United States*, the Supreme Court articulated a theory for systemic disparate treatment.[6] In basic terms, the systemic disparate treatment theory provides that if the plaintiff(s) can establish that an employer engaged in a "pattern or practice" of disparate treatment discrimination in the first phase of a bifurcated case, then the burden of proof shifts onto the employer to prove, in the second phase of the case, that any particular employment action it took against any individual member of that group was not the result of unlawful discrimination.[7] Since *Teamsters*, the Equal Employment Opportunity Commission (EEOC) and private plaintiffs have pursued the systemic disparate treatment theory—sometimes called the pattern or practice theory

---

antidiscrimination law ever since. *See generally* Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 702–03 (2006).

5. *See* Wal-Mart Stores, Inc. v. Dukes, 131 S. Ct. 2541 (2011). *See also* Dukes v. Wal-Mart Stores, Inc., 603 F.3d 571, 635, 644–45 (9th Cir. 2010) (Ikuta, J., dissenting) *cert. granted in part*, 131 S. Ct. 795, 178 L. Ed. 2d 530 (U.S. 2010) and *rev'd*, 131 S. Ct. 2541 (2011) (highlighting the difficulty in applying the systemic disparate treatment proof framework, including bifurcated trial procedures, to the largest class action ever certified). Although the issues presented to the Supreme Court in *Dukes* center on procedural issues involving class certification under Rule 23 of the Federal Rules of Civil Procedure, substantive questions about the proper scope and application of the systemic disparate treatment theory lurk beneath the surface and were raised at the oral argument. *See* Transcript of Oral Argument at 46–47, Dukes v. Wal-Mart, No. 10-277, 2010 WL 3358931 (2010). Recognizing the potential of the *Dukes* case to act as a catalyst for change in substantive disparate treatment theory, legal scholars have only recently begun to turn their attention to systemic disparate treatment law. *See* Tristin K. Green, *The Future of Systemic Disparate Treatment Law*, 32 BERKELEY J. EMP. & LAB. L. (forthcoming 2011) (arguing that systemic disparate treatment liability has been under-theorized due to a misguided focus on individual-level wrongdoing, rather than organization-level wrongdoing); Michael Selmi, *Theorizing Systemic Disparate Treatment Law*, 32 BERKELEY J. EMP. & LAB. L. (forthcoming 2011).

6. 431 U.S. at 360–62; *see also* Tristin K. Green, *Insular Individualism: Employment Discrimination Law After* Ledbetter v. Goodyear, 43 HARV. C.R.-C.L. L. REV. 353, 357, n.19 (2008) (calling *Teamsters* the Court's "foundational systemic disparate treatment case"); Michael J. Zimmer, *A Chain of Inferences Proving Discrimination*, 79 U. COLO. L. REV. 1243, 1268 (2008) (referring to *Teamsters* as the "seminal systemic disparate treatment case").

7. *Teamsters*, 431 U.S. at 360–61.

or the pattern or practice method of proof[8]—in a variety of employment discrimination cases.[9]

Today, systemic disparate treatment doctrine is in disarray. Questions abound regarding what types of discrimination claims, brought under which statutes, are subject to the *Teamsters* two-part burden-shifting analysis because of the unique characteristics of certain types of discrimination claims and slight variances among the relevant antidiscrimination statutes. For example, it is not clear whether the *Teamsters* method of proof should be available to age discrimination plaintiffs under the ADEA, disability discrimination plaintiffs under the ADA, or hostile work environment plaintiffs under any of the antidiscrimination statutes. It is also uncertain whether the *Teamsters* method of proof should be available to individual private plaintiffs, or if it is only available to private plaintiffs, where a class has been certified under Rule 23 of the Federal Rules of Civil Procedure.

Although some commentators have addressed isolated issues in systemic disparate treatment law, there has not yet been an effort in the literature to formulate a consistent theoretical approach that can help resolve these doctrinal issues and guide the future development of systemic disparate treatment law.[10] Principled approaches to systemic disparate treatment cases and clearer thinking on the *Teamsters* burden-shifting doctrine are becoming increasingly important. One reason is the EEOC's renewed emphasis on enforcing systemic discrimination cases. In 2006, the Commission announced a "Systemic Initiative" as part of its enforcement strategy. Tackling, in part, systemic disparate treatment, the implementation of the Systemic Initiative is a top priority for the EEOC.[11] According to the EEOC, the launch of the Systemic

8. *See, e.g.*, Semsroth v. City of Wichita, 304 Fed. App'x 707, 715 (10th Cir. 2008) (referring to "the pattern-or-practice method of proof"); Bacon v. Honda of America Mfg., 370 F.3d 565, 575 (6th Cir. 2004) (same).

9. *Cf.* Timothy G. Healy, Comment, *Sexual Pattern: Why a Pattern or Practice Theory of Liability is not an Appropriate Framework for Claims of Sexual Harassment*, 10 ROGER WILLIAMS U. L. REV. 537, 541 (2005) ("The pattern or practice theory of liability has been an essential tool in combating widespread, institutional discrimination.").

10. Examples of attempts to address individual pattern or practice doctrinal problems include: Jason R. Bent, *Systemic Harassment*, 77 TENN. L. REV. 151 (2009); Healy, *supra* note 9; and David J. Bross, Note, *The Use of Pattern-And-Practice by Individuals in Non-Class Claims*, 28 NOVA L. REV. 795 (2004).

11. *See* EEOC, FISCAL YEAR 2008 PERFORMANCE AND ACCOUNTABILITY REPORT 9 (2008), *available at* http://www.eeoc.gov/eeoc/plan/archives/annualreports/par/2008/par2008.pdf; EEOC, FISCAL YEAR 2007 ANNUAL REPORT ON THE OPERATIONS AND ACCOMPLISHMENTS OF THE OFFICE OF THE GENERAL COUNSEL § II.A. (2007), *available at* http://eeoc.gov/eeoc/litigation/reports/07annrpt/index.html ("To accomplish the goals of the systemic initiative, the Office of General Counsel anticipates a shift in the composi-

Initiative has been successful and is reflected in recent litigation statistics.[12] As long as the EEOC continues to prioritize systemic discrimination cases, difficult questions about the contours of the systemic disparate treatment theory will continue to surface in the courts.

This Article develops the missing theoretical foundation for systemic disparate treatment law and provides a roadmap for resolving questions about the proper application of *Teamsters*. It begins by examining the historical roots of the systemic disparate treatment theory. It demonstrates that, at its core, the *Teamsters* doctrine—like much of employment discrimination doctrine—is simply an evidentiary device for allocating and shifting burdens of proof among the parties. More specifically, it is a device for determining how the burdens of production and persuasion on the element of discriminatory intent should be allocated at an identifiable point during the course of a systemic discrimination case. Having established the operational nature of the systemic disparate treatment theory, the Article turns to the policy considerations that inform burden allocation and shifting in general civil litigation. It examines the considerations offered in traditional evidence literature as well as those suggested more recently in the law and eco-economics scholarship. By examining these policy considerations, this Article develops the proper foundation for analyzing, develop-

---

tion of its litigation docket over time. We expect fewer small, individual cases and more cases on behalf of larger groups of individuals.").

12. Statistics offered at a Commission meeting held on June 18, 2008 reflect the EEOC's new emphasis on systemic cases:

> I want to start with my conclusion which is that the EEOC has made substantial progress towards a revitalized and dynamic Systemic program following the Systemic Initiative adopted by the Commission in April 2006. The results support this conclusion. There is a striking increase in the number of Commissioner charges. As of May 31, 2008, 45 Commissioner charges were under investigation. Only 15 Commissioner charges were in investigation as of March 31, 2006. In fiscal year 2007, 24 Commissioner charges were developed and approved. Only 11 were signed in fiscal year 2006. In the first half of fiscal year 2008, 12 new Commissioner charges had been signed. The number of Systemic charges has also increased dramatically. From just over 200 Systemic cases in investigation prior to March 31, 2006 to over 1,000 as of May 31, 2008. Over 550 Systemic charges have been resolved since the adoption of the Systemic Initiative. Approximately 400 merit closures in which nearly $10,000,000 in monetary benefits has been obtained and approximately 1,200 persons were benefited through the administrative process.

Transcript of EEOC Commission Meeting of June 18, 2008 on Strategic Plan, Systemic Program and State and Local Program (Jun. 18, 2008) [hereinafter EEOC Transcript] (testimony of Katharine Kores), *available at* http://www.eeoc.gov/eeoc/meetings/archive/6-18-08/transcript.html.

ing, and applying the systemic disparate treatment theory in employment discrimination law.

The conclusion drawn here is that the proper application of the systemic disparate treatment theory depends on three critical factors:

(1) *Prior Probability (or Estimated Prevalence of the Discrimination Type)*: The estimated prior likelihood that the alleged unlawful discrimination actually occurred, before considering any evidence offered in the present case. This factor might also be thought of as the background prevalence of the particular type of discrimination alleged.

(2) *Evidentiary Signal*: The estimated probability that one would observe a particular evidentiary "signal" in Phase I of a bifurcated systemic disparate treatment case, assuming that the alleged discrimination did occur—and conversely, the estimated probability that one would observe that same evidentiary "signal" assuming that the alleged discrimination did not occur. In systemic disparate treatment cases, the evidentiary signal at issue will generally be statistical evidence proffered by the parties in Phase I.

(3) *Information Costs*: The estimated relative costs facing each party in searching for, finding, and presenting evidence establishing (or refuting) the alleged discriminatory intent.

The careful examination of these three factors will help decision-makers determine when and under what circumstances a systemic disparate treatment burden-shifting scheme is appropriate. Examination of these factors will allow a court to decide, in any particular systemic disparate treatment discrimination case, whether burden-shifting is justified after considering and weighing the proffered Phase I evidence.

The usefulness of this theoretical foundation extends beyond just informing courts' decisions in individual systemic discrimination cases. Generalizations might be made about *types* of employment discrimination cases based on what can be observed about those categories of cases. These generalizations can assist in determining the proper doctrinal extent of the systemic disparate treatment method of proof. Legislators and appellate courts can use an analysis of the three key factors identified herein to make principled decisions about what types of cases are appropriate for

systemic disparate treatment burden-shifting, and amend the anti-discrimination statutes and/or develop the *Teamsters* judicial doctrine accordingly. Thus, the theoretical foundation described above can assist in resolving the uncertainties in systemic disparate treatment theory.

This Article proceeds in four parts. Part I explores *Teamsters* and the origins of the systemic disparate treatment theory. The Part discusses systemic disparate treatment theory as a tool for allocating and shifting the burdens of production and persuasion with respect to the element of discriminatory intent. Part I also briefly considers some of the significant uncertainties that have recently developed in systemic disparate treatment law.

Part II begins the process of developing the proper foundation for application of the systemic disparate treatment theory and explores the policy roots of the systemic disparate treatment theory under both traditional and economic justifications for burden allocation and shifting in civil litigation. Such justifications suggest the factors of prior probability, evidentiary signal, and information cost should be determinative in the proper application of the systemic disparate treatment theory.

Part III applies the principles developed in Parts I and II to the specific context of systemic disparate treatment employment discrimination. Further, the Part explores how courts have misconstrued the evidentiary signal provided by statistical evidence, and overlooked the importance of prior probabilities and asymmetrical access to evidence in employment discrimination cases.

Finally, Part IV demonstrates how policymakers and appellate courts can use the principles this Article develops to make generalizations about categories of employment discrimination cases for which systemic disparate treatment burden-shifting might be appropriate.

I. ORIGINS OF THE SYSTEMIC DISPARATE TREATMENT THEORY

A. Title VII and Teamsters: Drawing Inferences of Discriminatory Intent in Systemic Disparate Treatment Cases

Title VII of the Civil Rights Act of 1964 prohibits discrimination in the terms or conditions of employment "because of . . . race,

color, religion, sex, or national origin."[13] Disparate treatment discrimination is often described as "the most easily understood type of discrimination. The employer simply treats some people less favorably than others because of their race, color, religion, sex, or national origin."[14] Discriminatory intent or motive is the key to a disparate treatment case, "although it can ... be inferred from the mere fact of differences in treatment."[15]

The *systemic* disparate treatment theory describes one method of creating an inference of the required discriminatory intent.[16] Systemic disparate treatment theory has been defined by some as encompassing two different types of cases: (1) cases in which an employer explicitly announces or admits a formal or express policy treating employees differently on the basis of a protected characteristic, and (2) cases in which no express or announced policy exists, but there is evidence that the employer consistently treats employees differently on the basis of such protected characteristics.[17] Others define systemic disparate treatment cases as only those falling into the second category and involving the use of statistical evidence to create an inference of discriminatory intent.[18] There is no need to use an inference to establish discriminatory intent in cases falling within the first category.[19]

Decided in 1977, *Teamsters* is an example of the second category and is the cornerstone of the systemic disparate treatment theory.[20]

---

13. Pub. L. No. 88-352, 78 Stat. 241 (codified as amended at 42 U.S.C. § 2000e-2 (2006)).

14. *Teamsters*, 431 U.S. at 335 n.15.

15. *Id.*; *see also* SULLIVAN ET AL., *supra* note 1, § 3.01, at 146–47. This is to be contrasted with disparate impact cases, where "liability does not depend upon a finding of intent to discriminate." *Id.* at § 4.01; *see also Teamsters*, 431 U.S. at 335–36 n.15 (similar).

16. SULLIVAN ET AL., *supra* note 1, § 3.03[C].

17. *See, e.g.*, *id.* § 3.01. Cases involving the explicit use of protected characteristics in announced policies have become increasingly rare. *See id.*

18. RAMONA L. PAETZOLD & STEVEN L. WILLBORN, THE STATISTICS OF DISCRIMINATION: USING STATISTICAL EVIDENCE IN DISCRIMINATION CASES § 1:8, at 1–26 (Thompson West 2006); *see also* SULLIVAN, ET AL., *supra* note 1, § 3.01, at 146 n.4. Cases falling within the second category are sometimes referred to in the case law and scholarship as "pattern or practice" cases, and I will accordingly use that terminology at times in this Article. Although I am primarily concerned with the second category of cases, the principles that I develop herein might also be used to analyze cases from the first category. *See infra* note 168.

19. *See* Bates v. United Parcel Serv., Inc., 511 F.3d 974, 988 (9th Cir. 2007) (the *Teamsters* burden-shifting framework is unnecessary where the case involves an employer's facially discriminatory policy, because the fact to be uncovered by the burden-shifting scheme has already been admitted). Where an employer maintains a formal policy that treats employees differently based on race or sex the discriminatory intent is established without the aid of an inference, and the defendant must then resort to defending its policy as a bona fide occupational qualification. *See* SULLIVAN, ET AL., *supra* note 1 § 3.02[B], at pp. 167–170.

20. *See* SULLIVAN, ET AL., *supra* note 1, § 3.01. Earlier cases established the concept of shifting burdens based on inferences of discrimination in the employment discrimination context. *See, e.g.* McDonnell Douglas Corp. v. Green, 411 U.S. 792, 802 (1973) (setting out

In *Teamsters*, the United States government, as plaintiff, alleged that the defendant employer engaged in a "pattern or practice" of discriminating against minorities in hiring for certain positions and in the pay, assignments, and promotions of those minorities that were hired.[21] The *Teamsters* Court held that statistical evidence of a longstanding and gross disparity between the composition of the defendant's workforce and the general population could serve as a "telltale sign" of discrimination and support an inference of discriminatory intent.[22]

The *Teamsters* Court drew upon language in Title VII that expressly authorizes the government to pursue "pattern or practice" enforcement actions.[23] The relevant text provides that whenever the government has cause to believe that a defendant "is engaged in a pattern or practice of resistance to the full enjoyment" of the rights secured by Title VII, and "that the pattern or practice is of such a nature and is intended to deny the full exercise" of those rights, the government may bring a civil action in federal district court.[24] Title VII was later amended to expressly permit private individuals to file charges alleging a pattern or practice of discrimination and the Supreme Court has interpreted Title VII as allowing private pattern or practice actions.[25]

---

the basic prima facie case for individual claims of discrimination). Burden-shifting under the heading "pattern or practice" had been recognized in *Franks v. Bowman*, 424 U.S. 747 (1976), involving a private class action. However, *Teamsters* is generally considered the seminal case for the modern systemic disparate treatment doctrine because of its approval of the use of statistical evidence in Phase I to establish the pattern or practice and to justify shifting the burden of proof onto the defendant. *See supra* notes 6–7 and accompanying text; *see also* Semsroth v. City of Wichita, 304 Fed. App'x 707, 715 (10th Cir. 2008) (referring to *Teamsters* as "the seminal case on the pattern-or-practice method of proof."); David G. Karro, *Common Sense About Common Claims*, 25 Hofstra Lab. & Emp. L.J. 33, 45 (2007) (referring to *Teamsters* as "the seminal pattern-or-practice case"); Healy, *supra* note 9, at 541 (similar).

21. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 329-30 (1977). The government alleged that defendant engaged in a pattern or practice of discriminating against minorities in hiring line drivers. Those minorities that were hired were allegedly given less desirable assignments and less pay, and were discriminated against in promotion decisions. *Id.* at 329, 335.

22. 431 U.S. at 339 n.20.

23. *See* 431 U.S. at 329–30 n.1 (quoting the "pattern or practice" language of 42 U.S.C. § 2000e-6(a) (1970) (amended 1972)).

24. 42 U.S.C. § 2000e-6(a) (2006). The power to bring pattern or practice suits under Section 707 was originally vested in the Attorney General. Title VII of the Civil Rights Act of 1964, Pub. L. No. 88-352, § 707, 78 Stat. 241, 261–62 (1964). In 1972, Congress amended Section 707 to transfer this authority to the EEOC. *See* Equal Employment Opportunity Act of 1972, Pub. L. No. 92-261, § 5, 86 Stat. 103, 107 (1972) (codified as amended at 42 U.S.C. §§ 2000e-6(c)–(e)).

25. 42 U.S.C. § 2000e-6(e). *See* Equal Employment Opportunity Act of 1972 § 5 (authorizing the EEOC to investigate and act on pattern or practice charges filed by individuals); Cooper v. Fed. Reserve Bank of Richmond, 467 U.S. 867, 876 n.9 (1984) ("Although *Teamsters* involved an action litigated on the merits by the Government as plaintiff

In *Teamsters*, the Court set out the two-phase litigation framework used in modern systemic disparate treatment cases. The plaintiff's burden, in Phase I, is to prove the existence of a pattern or practice of discrimination by the employer.[26] If the plaintiff successfully meets this burden, it creates a rebuttable presumption that all individual employment decisions made during the period of the pattern or practice were discriminatory.[27] The employer then bears the burden, in Phase II, of rebutting this presumption by proving that any individual employment decisions were not the product of discrimination.[28]

In setting out this pattern or practice framework, the *Teamsters* Court approved the use of statistical evidence in Phase I to demonstrate the existence of the employer's pattern or practice of discrimination.[29] The Court noted that statistical analyses "serve an important role" in cases where there is a dispute about the existence of discrimination. In footnote 20, the Court famously explained that: "[A]bsent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired."[30]

The Court explained that statistical evidence of an imbalance "is often a telltale sign of purposeful discrimination."[31] The rationale

---

under § 707(a) of the Act, it is plain that the elements of a prima facie pattern-or-practice case are the same in a private class action.") (citing *Teamsters*, 431 U.S. at 358–60).

26. *Teamsters*, 431 U.S. at 360 ("[The government's] initial burden is to demonstrate that unlawful discrimination has been a regular procedure or policy followed by an employer or group of employers.").

27. *Id.* at 362.

28. *Id.*

29. *Id.* at 340 n.20.

30. *Id.* The Court's reasoning in footnote 20 is largely responsible for the proliferation of statistical evidence in employment discrimination cases. However, some commentators have criticized the Court for overlooking factors other than discrimination—such as applicant self-selection—that would systematically prevent an observed work force from resembling the work force that would be expected if hiring decisions were made randomly. *See* Paul Meier et al., *What Happened in* Hazelwood*: Statistics, Employment Discrimination, and the 80% Rule*, 9 AM. B. FOUND. RES. J. 139 (1984). Meier contends that the quoted text from footnote 20 is "untenable, and it can only be given a reasonable interpretation as a statement of the impartiality with which the Court must approach a case of this kind." *Id.* at 158. *But see* Deborah A. Calloway, St. Mary's Honor Center v. Hicks*: Questioning the Basic Assumption*, 26 CONN. L. REV. 997, 997 n.2, 1025–36 (1994) (defending the Court's reasoning in footnote 20 as the "basic assumption," and contending that the assumption remains justified).

31. *Teamsters*, 431 U.S. at 340 n.20. The Court was careful to distinguish this role of statistics as a "telltale sign of purposeful discrimination" from the imposition of a quota or similar requirement that the workforce resemble the general population. *Id.* The Court stated: "Evidence of longlasting and gross disparity between the composition of a work force

for *Teamsters* is therefore an evidentiary one. Discriminatory intent remains a key element of any disparate treatment claim, but in systemic disparate treatment cases like *Teamsters* statistical evidence can, where appropriate, support an inference of that discriminatory intent and justify shifting the burden of proof onto the defend-defendant for that element.

The Court recognized that the burden-shifting framework it described was in accordance with general burden-shifting principles:

> The holding in *Franks* that proof of a discriminatory pattern and practice creates a rebuttable presumption in favor of individual relief is consistent with the manner in which presumptions are created generally. *Presumptions shifting the burden of proof are often created to reflect judicial evaluations of probabilities and to conform with a party's superior access to the proof.*[32]

In this passage, the *Teamsters* Court identified two key features of systemic disparate treatment cases that justify shifting the burden of proof on the element of discriminatory intent: (1) probability assessments and (2) relative access to proof.

B. Teamsters Applied to Systemic Disparate Treatment Law

Since *Teamsters*, the Supreme Court has rarely addressed systemic disparate treatment discrimination cases.[33] The drought of Supreme Court decisions has been especially noticeable in recent

---

and that of the general population thus may be significant even though § 703(j) makes clear that Title VII imposes no requirement that a work force mirror the general population." *Id.*

32. *Id.* at 359 n.45 (emphasis added) (citing C. MCCORMICK, LAW OF EVIDENCE §§ 337, 343 (2d ed. 1972)); Fleming James, Jr., *Burdens of Proof*, 47 VA. L. REV. 51, 61 (1961); Keyes v. School Dist. No. 1, 413 U.S. 189, 208–09 (1973)). The *Teamsters* Court appears to have envisioned a shift in both the burden of production and the burden of persuasion, although it did not use those terms. *See infra* note 70 and accompanying text. If plaintiff meets its burden in Phase I, then the defendant must "demonstrate that the individual applicant was denied an employment opportunity for lawful reasons." *Teamsters*, 431 U.S. at 362. The distinctions between the burdens of persuasion and production are further discussed below. *See infra* Part I.C.

33. *See, e.g.*, Bazemore v. Friday, 478 U.S. 385 (1986); Gen. Tel. Co. v. EEOC, 446 U.S. 318 (1980); Hazelwood Sch. Dist. v. United States, 433 U.S. 299 (1977). The Court has also recently acknowledged and reserved ruling on some difficult questions arising in the systemic disparate treatment context, including how the timely-filed charge requirement should be applied in pattern or practice cases. *See, e.g.*, Nat'l Passenger R.R. Corp. v. Morgan, 536 U.S. 101, 115 n.9 (2002).

years. In the interim, a number of fundamental uncertainties about the application of the doctrine have emerged.

1. Age Discrimination

Can the *Teamsters* doctrine be utilized in age discrimination cases under the Age Discrimination in Employment Act ("ADEA")[34]? This question stems from differences between the statutory text of the ADEA and Title VII. Title VII, which addresses discrimination based on race, color, religion, sex, and national origin, contains the language in Section 2000e-6(a), quoted above, that includes express references to the term "pattern or practice."[35] The ADEA contains no similar language and has no reference to "pattern or practice."[36]

Employers have argued that this textual difference is significant, but lower courts have not found it problematic.[37] Several circuit courts have allowed ADEA plaintiffs to use the *Teamsters* framework.[38] Their common reason is that the pattern or practice framework was developed by courts to assist in their decision-making and that despite the absence of an express statutory reference to "pattern or practice" in the ADEA, it is nonetheless appropriate to borrow the framework from Title VII cases and apply it in ADEA cases.[39]

Given the recent resurgent emphasis on rigid textualism, there is reason to question whether the current Supreme Court would allow the *Teamsters* framework developed in the Title VII context to be used in ADEA claims because the ADEA lacks any reference to "pattern or practice" liability. This possibility is illustrated by the Court's 2009 decision in *Gross v. FBL Financial Services, Inc.*[40] There, the Supreme Court—on its own initiative and before reaching the question certified by the parties—considered "whether the burden of persuasion ever shifts to the party defending an alleged mixed-

34. 29 U.S.C. §§ 621–634 (2006).

35. *See supra* note 24 and accompanying text.

36. 29 U.S.C. §§ 621–634; *see also* Thompson v. Weyerhaeuser Co., 582 F.3d 1125, 1127 (10th Cir. 2009) ("Unlike Title VII, the ADEA contains no express reference to pattern-or-practice claims."); King v. Gen. Elec. Co., 960 F.2d 617, 624 (7th Cir. 1992) (noting the ADEA "has no parallel provision" to Title VII's pattern or practice language).

37. *See, e.g.*, *Thompson*, 582 F.3d at 1129; *King*, 960 F.2d at 624.

38. *See Thompson*, 582 F.3d at 1129; Hipp v. Liberty Nat'l Life Ins. Co., 252 F.3d 1208, 1227 (11th Cir. 2001); *King*, 960 F.2d at 624; Haskell v. Kaman Corp., 743 F.2d 113, 119 (2d Cir. 1984); EEOC v. W. Elec. Co., 713 F.2d 1011, 1016 (4th Cir. 1983); Marshall v. Sun Oil Co., 605 F.2d 1331, 1336 n. 2 (5th Cir. 1979).

39. *See, e.g.*, *Thompson*, 582 F.3d at 1129; *King*, 960 F.2d at 624.

40. 129 S. Ct. 2343, 2350–52 (2009).

motives discrimination claim brought under the ADEA."[41] The Court held that it did not, focusing on textual differences between Title VII—specifically, the 1991 Amendments to Title VII—and the ADEA.[42] The Court emphasized that, "[w]hen conducting statutory interpretation, we 'must be careful not to apply rules applicable under one statute to a different statute without careful and critical examination.' "[43]

Because Title VII contains an express reference to "pattern or practice," (a textual reference relied upon by the *Teamsters* Court[44]) and the ADEA does not, textualist reasoning like that employed in the *Gross* opinion could lead the Court to preclude application of the *Teamsters* doctrine in ADEA cases.[45] The Supreme Court has never addressed the question, and the *Gross* opinion may have created uncertainty in this area.[46] Indeed, employers are already raising precisely this argument in the wake of *Gross*.[47]

This is typical of other areas in systemic disparate treatment law that are unsettled. This dispute highlights the need to develop a foundation, based upon sound policy considerations, for determining when and how the systemic disparate treatment theory should be applied.

---

41. *Id.* at 2348.

42. *Id.* at 2348–49 (noting that Congress is presumed to have acted intentionally when it amended Title VII without simultaneously amending the ADEA).

43. *Id.* at 2349 (quoting Federal Express Corp. v. Holowecki, 552 U.S. 389, 393 (2008)).

44. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 329 n.1 (1977).

45. The Tenth Circuit's opinion in *Thompson*, decided after *Gross*, rejected this line of argument advanced by the defendant. Thompson v. Weyerhaeuser Co., 582 F.3d 1125, 1131 (10th Cir. 2009). The *Thompson* court reasoned that the mixed-motive burden-shifting framework at issue in *Gross* was of a different kind and was founded on the text of the 1991 statutory amendments, while the pattern or practice burden-shifting doctrine is "mentioned in neither statute" and "has been established by the courts." *Id.* It remains to be seen whether other courts will agree with this distinction. In particular, it should be noted that the mixed-motive burden-shifting doctrine was originally a judicially created doctrine, and that Congress amended Title VII (but not the ADEA) in 1991 to override the Supreme Court's interpretation of the mixed-motive burden-shifting standards in the fractured opinion in *Price Waterhouse v. Hopkins*, 490 U.S. 228 (1989). *See* Desert Palace, Inc. v. Costa, 539 U.S. 90, 94–95 (2003). *See generally* Michael J. Zimmer, *The New Discrimination Law:* Price Waterhouse *is Dead, Whither* McDonnell Douglas?, 53 EMORY L.J. 1887, 1909–14 (2004). Prior to the *Gross* decision, lower courts frequently applied the same mixed-motive burden-shifting analysis to ADEA and Title VII cases. *See* Brown v. J. Kaz, Inc., 581 F.3d 175, 186 (3rd Cir. 2009) (Jordan, J., concurring) (collecting cases); Martin J. Katz, Gross *Disunity*, 114 PENN ST. L. REV. 857, 868–69 (2010).

46. *Brown*, 581 F.3d at 186 (Jordan, J., concurring) (noting that the Court's "fundamental instruction in *Gross* [was] that analytical constructs are not to be simply transposed from one statute to another without a thorough and thoughtful analysis").

47. *See Thompson*, 582 F.3d 1125, 1131.

2. Disability Discrimination

The Americans with Disabilities Act ("ADA")[48] presents a different type of problem than the one presented by the ADEA. Although the ADA's statutory text does not include the "pattern or practice" language found in Title VII, it does (unlike the ADEA) expressly incorporate the remedies and procedures set forth in Title VII, including 42 U.S.C. § 2000e-6 (Section 707 of Title VII), which contains the reference to "pattern or practice" liability.[49] Thus, the text at least facially supports application of *Teamsters* in ADA cases.[50] Still, there remains doubt. Although some courts of appeals have suggested that application of the *Teamsters* framework is appropriate in ADA cases, none have "addressed directly if and how this framework might apply to a private-plaintiff pattern-or-practice class action under the ADA."[51]

The primary argument against the application of the systemic disparate treatment theory in ADA cases is the uniquely individualized inquiry needed to determine whether a plaintiff is a "qualified individual with a disability" under the statute. The ADA prohibits covered employers from "discriminat[ing] against a qualified individual with a disability."[52] A "qualified individual with a disability," is defined as "an individual with a disability who, with or without reasonable accommodation, can perform the essential functions of the employment position that such individual holds or desires."[53] The definition requires courts to make an individualized inquiry. Although categorizations of race, sex, national origin, and religion can sometimes be imprecise, they do not require the same type of highly individualized inquiry that the ADA demands.

---

48. Pub. L. No. 101-336, 104 Stat. 327 (codified as amended primarily in 42 U.S.C. §§ 12101–12231 (2006)).

49. 42 U.S.C. § 12117(a) (2006) ("The powers, remedies, and procedures set forth in sections 2000e-4, 2000e-5, 2000e-6, 2000e-8, and 2000e-9 of this title shall be the powers, remedies, and procedures this subchapter provides to the Commission, to the Attorney General, or to any person alleging discrimination on the basis of disability in violation of any provision of this chapter, or regulations promulgated under section 12116 of this title, concerning employment.").

50. *See* Hohider v. United Parcel Serv., 574 F.3d 169, 180, 182–83 (3d Cir. 2009) (assuming, without deciding, that 42 U.S.C. § 12117(a) and Title VII jurisprudence combine to make application of the *Teamsters* pattern or practice framework in ADA cases appropriate); Davoll v. Webb, 194 F.3d 1116, 1147–48 (10th Cir. 1999). Notably, the three judge appellate panel in *Hohider* included retired Justice Sandra Day O'Connor, sitting by designation.

51. *Hohider*, 574 F.3d at 179 n.11; *see also* EEOC v. Murray, Inc., 175 F. Supp. 2d 1053, 1060 (M.D. Tenn. 2001) (noting shortage of circuit court authorities). In *Hohider*, the Third Circuit assumed, without deciding, that the *Teamsters* framework could be applied to ADA cases. 574 F.3d at 182–83.

52. 42 U.S.C. § 12112(a) (2006).

53. *Id.* § 12111(8).

As the court in *Hohider* explained, "[t]he ADA does not define the scope of its protections and prohibitions as broadly as Title VII . . . . [I]t does not prohibit discrimination against *any individual* based on disability . . . ."[54] Rather, it prohibits discrimination only against "qualified individuals."[55] Therefore, the question of whether a defendant unlawfully discriminated against, or failed to provide reasonable accommodations to, any group of disabled persons in violation of the ADA will require an initial inquiry into whether any or all of the individuals were "qualified individual[s] with a disability."[56] Courts have split on whether this inquiry would need to be performed at Phase I of a *Teamsters* systemic disparate treatment case, or whether it could await individual liability determinations in Phase II.[57] Whether and how to apply the *Teamsters* systemic disparate treatment theory in ADA cases remains a difficult question for the courts.

3. Hostile Work Environment

No area of systemic disparate treatment law is more unsettled and contentious than its potential application to hostile work environment harassment claims. One district court recently referred to pattern or practice of harassment claims as a "still uncertain theory of liability."[58] The array of problems with applying *Teamsters* to hostile work environment cases include: the lack of statistical evidence in harassment cases, determining how to incorporate the subjective element of a harassment claim into the two phase framework of *Teamsters*, resolving questions about which individuals may properly be included in the class seeking relief, determining how many claimants is sufficient to constitute a "pattern or practice" justifying

54. *Hohider*, 574 F.3d at 191.

55. 42 U.S.C. § 12112(a).

56. *Id.*; *Hohider*, 574 F.3d at 192; *see also* EEOC v. J.B. Hunt Transp., Inc., 128 F. Supp. 2d 117, 124 (N.D.N.Y. 2001).

57. *Compare Hohider*, 574 F.3d at 192 (must be determined in Phase I), *and J.B. Hunt*, 128 F. Supp. 2d at 124 (same), *with* Davoll v. Webb, 194 F.3d 1116, 1147–48 (10th Cir. 1999) (affirming the district court's application of *Teamsters* such that the government would not need to prove the "qualified individuals" requirement until Phase II), *and* EEOC v. Murray, Inc., 175 F. Supp. 2d 1053, 1060 (M.D. Tenn. 2001) (can be determined in Phase II). *See also* Michael Waterstone, *A New Vision of Public Enforcement*, 92 MINN. L. REV. 434, 464 (2007) ("The same is true under the ADA, which has allowed the EEOC, in the few class-type cases it has brought, to escape thorny class issues of individualized inquiries that courts usually view as necessary to determine disability.").

58. EEOC v. Carrols Corp., No. 5:98 CV 1772, 2005 WL 928634, at *2 n.4 (N.D.N.Y. Apr. 20, 2005); *see also* Healy, *supra* note 9, at 579 (arguing that the pattern or practice burden-shifting doctrine should never apply in harassment cases).

application of the *Teamsters* doctrine, and answering questions about how the *Ellerth/Faragher* affirmative defense to a harassment claim properly can be applied in the context of a systemic case.[59]

Even where the application of *Teamsters* to systemic harassment cases has been allowed, courts and scholars have questioned whether a successful showing that the employer engaged in a pattern or practice of tolerating harassment in Phase I (however that might be accomplished without statistics) should be sufficient to shift the burden of proof onto the defendant (or create an inference of unlawful harassment in favor of the claimants) in Phase II.[60] Whether *Teamsters* burden-shifting should be permitted in hostile work environment cases remains an open question.[61]

4. Individual (Non-Class Action) Private Plaintiffs

Can individual private plaintiffs utilize the systemic disparate treatment theory to prove discrimination? Many individual discrimination plaintiffs have attempted to invoke the *Teamsters* burden-shifting mechanism.[62] Defendants contend that pattern or practice claims necessarily involve the existence of a group of victims, and therefore, the *Teamsters* burden-shifting framework should be available to private plaintiffs only where a class action is certified under Federal Rule 23.

This question has reached the appellate level in a number of circuits, and the overwhelming majority of the circuits have been persuaded by defendants' arguments. The rule in at least seven circuits is that a private individual plaintiff cannot pursue a *Team-*

59. *See* Bent, *supra* note 10.

60. *See* EEOC v. Int'l Profit Assocs., Inc., No. 01 C 4427, 2007 WL 3120069, at *14 (N.D. Ill. Oct. 23, 2007) (arguing that the burden of proof should not shift to defendant, even after a finding of liability in Phase I of a pattern or practice of harassment case); *see also* Bent, *supra* note 10, at 188–89; Healy, *supra* note 9, at 576–78. *But see* EEOC v. CRST Van Expedited, Inc., 611 F. Supp. 2d 918, 936–37 (N.D. Iowa 2009) (determining that a modified burden-shifting approach is the law of the Eighth Circuit); EEOC v. Dial, 156 F. Supp. 2d 926, 947 (N.D. Ill. 2001) (burden-shifting appropriate after a finding of pattern or practice liability in Phase I); EEOC v. Mitsubishi Motor Mfg. of Am., Inc., 990 F. Supp. 1059, 1077 (C.D. Ill. 1998) (same). The question of whether the burden of proof should shift in a pattern or practice of harassment case is complicated by the types of statistics available in harassment cases—usually, only the frequency and distribution of harassment complaints compared to the number and distribution of women in the workforce. *See* Bent, *supra* note 10, at 170. Additional complications are introduced by the subjective element required for a claim of hostile work environment and by the availability of the *Ellerth/Faragher* affirmative defense for employers. *See id.* at 168–69.

61. *See* Bent, *supra* note 10, at 157.

62. *See* cases cited *infra* note 63; Bross, *supra* note 10, at 804–05.

sters framework claim.[63] Only the D.C. circuit has even arguably indicated a willingness to apply *Teamsters* burden-shifting in favor of individual private plaintiffs.[64] Several of the courts prohibiting burden-shifting for individuals have nonetheless allowed the individual plaintiff to use available evidence of a pattern or practice of discrimination (such as statistical evidence of disparities) to help prove the pretext portion of an individual claim within the individual disparate treatment framework of *McDonnell Douglas*.[65] Again, the Supreme Court has not considered the issue.

5. Potential for Additional Confusion in Systemic Disparate Treatment Law

The foregoing uncertainties are not minor ones. They represent fundamental questions about the extent and application of the systemic disparate treatment theory. The issues come up repeatedly, as demonstrated by the number of federal district court and appellate court opinions addressing or referencing the topics. Further, any amendments to existing antidiscrimination statutes have the potential to add to this confusion.[66] The enactment of new antidiscrimination laws that fail to explicitly address the systemic dispardisparate treatment theory may also create new doctrinal issues.[67]

63. *See* Semsroth v. City of Wichita, 304 Fed. App'x 707, 716–17 (10th Cir. 2008); Davis v. Coca-Cola Bottling Co., 516 F.3d 955, 967 n.24, 969 n.30 (11th Cir. 2008) (disapproving of an argument that prior Eleventh Circuit authority permitted individuals to pursue pattern or practice burden-shifting); Williams v. Giant Food Inc., 370 F.3d 423, 429–30 (4th Cir. 2004); Bacon v. Honda of Am. Mfg., 370 F.3d 565, 575 (6th Cir. 2004); Celestine v. Petroleos de Venezuella SA, 266 F.3d 343, 355–56 (5th Cir. 2001); Thiessen v. Gen. Elec. Capital Corp., 267 F.3d 1095, 1106 (10th Cir. 2001); Lowery v. Circuit City Stores, Inc., 158 F.3d 742, 761 (4th Cir. 1998), *vacated on other grounds*, 527 U.S. 1031 (1999); Gilty v. Village of Oak Park, 919 F.2d 1247, 1252 (7th Cir. 1990); Craik v. Minn. State Univ. Bd., 731 F.2d 465, 469–70 (8th Cir. 1984); *see also* Bross, *supra* note 10, at 796–97.

64. *See* Davis v. Califano, 613 F.2d 957, 961–62 (D.C. Cir. 1979); *see also* Bross, *supra* note 10, at 807–08.

65. *See, e.g.*, Mendelsohn v. Sprint/United Mgmt. Co., 466 F.3d 1223, 1227 n.2 (10th Cir. 2006); *Lowery*, 158 F.3d at 761. This approach would allow the individual plaintiff to use statistical evidence of systemic discrimination in order to bolster a claim of pretext under the third step of the *McDonnell Douglas* framework, for which the plaintiff bears the burden of proof. *See* Bell v. E.P.A., 232 F.3d 546, 553 (7th Cir. 2000) (in non-systemic cases, statistical evidence can be relevant and probative on the issue of pretext); *see also* SULLIVAN ET AL., *supra* note 1, § 5.04[B].

66. For example, in the ADEA context the law appeared to be relatively settled that the pattern or practice theory was available to plaintiffs until the 1991 Amendments and the Supreme Court's decision in *Gross* opened the door for the textualist argument advanced by the defendants in *Thompson v. Weyerhaeuser Co.* 582 F.3d 1125, 1131 (10th Cir. 2009); *see also* Katz, *supra* note 45, at 867–71.

67. For example, the newest federal antidiscrimination law, the Genetic Information Nondiscrimination Act of 2008 (GINA), became effective on November 21, 2009. *See* Pub. L.

The uncertainty will continue to grow unless it is met directly—either by the legislature addressing the systemic disparate treatment theory explicitly in statutory text, or by courts adopting a coherent and principled approach to the application of *Teamsters* under the various antidiscrimination statutes.

C. A Note About Burdens of Production and Persuasion

The *Teamsters* Court refers to the "burden of proof" in setting out the systemic disparate treatment theory,[68] but that term glosses over an important complexity. The burden of proof can be divided into two separate components: the burden of production and the burden of persuasion. Although this distinction can be a slippery one,[69] it nonetheless plays an important role in systemic disparate treatment theory.

Evidence and procedure scholars have long recognized a conceptual distinction between the burden of production and the burden of persuasion.[70] The burden of production (sometimes

---

No. 110-233, Title II, 122 Stat. 905 (codified at 42 U.S.C. §§ 2000ff to 2000ff-11 (Supp. II 2007–2009)). GINA generally incorporates the enforcement provisions of Title VII, including Section 707 of Title VII, which contains the operative phrase "pattern or practice." *See* 42 U.S.C. § 2000ff-6(a)(1). GINA may add a new wrinkle to the analysis, however, because it provides that disparate impact is not a viable theory under GINA—at least not yet. *See* 42 U.S.C. § 2000ff-7(a) ("Notwithstanding any other provision of this Act, 'disparate impact', as that term is used in section 2000e-2(k) of this title, on the basis of genetic information does not establish a cause of action under this Act."). GINA creates a commission to review the development of genetic sciences and to make recommendations regarding whether the disparate impact theory should be made available to plaintiffs under GINA. *See* 42 U.S.C. § 2000ff-7(b). GINA contains no similar language expressly precluding a systemic disparate *treatment* theory. *See* 42 U.S.C. §§ 2000ff to 2000ff-11. However, the express disapproval of the disparate impact theory suggests that the legislature may not have intended for GINA to provide systemic disparate treatment claims based upon statistical evidence.

68. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 359 n.45 (1977).

69. *See* CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, EVIDENCE UNDER THE RULES: TEXT, CASES, AND PROBLEMS 761, 763–66 (3d ed. 1996). ("[T]he subject of burdens is among the most slippery in the larger areas of procedure and evidence, and attempts to refine what we know lead quickly to difficulty."); John T. McNaughton, *Burden of Production of Evidence: A Function of a Burden of Persuasion*, 68 HARV. L. REV. 1382, 1382 (1955) (observing that the two burdens are related); Ronald J. Allen, *Burdens of Proof, Uncertainty, and Ambiguity in Modern Legal Discourse*, 17 HARV. J.L. & PUB. POL'Y. 627, 634 (1994).

70. *See* FLEMING JAMES, JR. ET AL., CIVIL PROCEDURE § 7.12, at 414 (5th ed. 2001) ("The term 'burden of proof' is used to refer to two different concepts. Until the distinction was defined by James Bradley Thayer in 1898, the reasoning of the decisions on this point was hopelessly confused. The two distinct concepts may be referred to as (1) the risk of nonpersuasion, sometimes called the 'burden of persuasion,' and (2) the duty of producing evidence (or burden of production), sometimes called the 'burden of going forward with the evidence.'") (footnotes omitted); MUELLER & KIRKPATRICK, *supra* note 69, at 763–66; Chris William Sanchirico, *A Primary-Activity Approach to Proof Burdens*, 37 J. LEGAL STUD. 273, 274 n.1 (2008).

called the "risk of non-production" or the "burden of going forward with evidence") refers to the burden of coming forward with evidence on the point at issue.[71] There is not only a directional component to the burden of production, but also a sufficiency component. The party bearing the burden of production must come forward with something more than just a scintilla of evidence; it must come forward with evidence sufficient to justify a reasonable jury finding in its favor on the point.[72]

The burden of persuasion, in contrast, becomes relevant once all of the evidence has been submitted and considered by the trier of fact. The party bearing the burden of persuasion must affirmatively convince the trier of fact that it has proven, to the required level or standard of proof, the material facts necessary to prevail.[73] In most civil cases, the applicable standard is proof by a preponderance of the evidence.[74] The party bearing the burden of persuasion in a civil case will lose if, after considering all of the submitted evidence, the trier of fact's "mind is in equipoise."[75]

However, these two components of the "burden of proof" are not completely independent. Mueller and Kirkpatrick described their relationship as follows:

> The burdens of production and persuasion are related: The first requires a party to produce sufficient evidence to permit reasonable persons on the jury to find the point with the requisite measure of certainty, as defined by the burden of persuasion. The second means, in most civil actions, proof by a preponderance—that is, evidence which persuades the jury (acting as reasonable persons) that the points to be proved are more likely so than not.[76]

---

71. *See* JAMES ET AL., *supra* note 70, § 7.15, at 417 ("The second meaning that is commonly given to the term 'burden of proof' refers to the burden of coming forward with evidence. This is sometimes called 'the burden of production.' (A more accurate term might be 'the risk of nonproduction,' but this expression is not in general use.)").

72. *See* 1 CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, FEDERAL EVIDENCE § 3:4, at 434 (3d ed. 2007) ("A party carries the burden of production by introducing evidence sufficient to support the findings of fact that are necessary if she is to prevail.").

73. *See* JAMES ET AL., *supra* note 70, § 7.13.

74. *See id.* § 7.14.

75. *Id.* § 7.13. Professor Chris Sanchirico has argued that the term "equipoise" is not defined with sufficient precision. *See* Chris William Sanchirico, *The Burden of Proof in Civil Litigation: A Simple Model of Mechanism Design*, 17 INT'L REV. L. & ECON. 431 (1997).

76. MUELLER & KIRKPATRICK, *supra* note 69, at 766 (footnote omitted); *see also* Ronald J. Allen, *Burdens of Proof, Uncertainty, and Ambiguity in Modern Legal Discourse*, 17 HARV. J.L. & PUB. POL. 627, 634 (1994) ("Immediately, then, it becomes apparent that there are not two analytically distinct components to burdens of proof; as Professor McNaughton saw many decades ago, burdens of production are a function of burdens of persuasion. A burden of

Ordinarily, both the burden of production and the burden of persuasion are initially assigned to the plaintiff in civil cases.[77] In many areas of law, and particularly in employment discrimination law, courts use various presumptions to shift the burden of production back and forth. In the case that has largely defined individual disparate treatment law, *McDonnell Douglas Corp. v. Green*, the Supreme Court set out a framework for shifting the burden of production.[78] If the plaintiff makes out a prima facie case of discrimination, then the burden of production shifts to the defendant to offer a legitimate, nondiscriminatory reason for the adverse employment action. If the defendant does so, then the plaintiff must prove that the defendant's proffered reason was pretextual.[79] The *McDonnell Douglas* scheme shifts onto defendant the burden of production;[80] however, the burden of persuasion never shifts.[81]

It is not as obvious which burden is supposed to be shifting in a *Teamsters* systemic disparate treatment case. The terms "burden of persuasion" and "burden of production" are notably absent from the *Teamsters* opinion.[82] Courts and commentators have generally agreed, however, that the rebuttable presumption (or inference) of the required discriminatory intent prescribed by *Teamsters* amounts

---

production is satisfied when a reasonable person applying the relevant burden of persuasion could find in favor of the person bearing the burden of persuasion. Together, they operate to bring order to the evidentiary process at trial.") (footnote omitted); McNaughton, *supra* note 69.

77. *See infra* Part II.A.

78. *See* 411 U.S. 792, 802 (1973).

79. *See id.* at 804; *see also* Martin J. Katz, *Reclaiming* McDonnell Douglas, 83 NOTRE DAME L. REV. 109, 117 (2007); Zimmer, *supra* note 45, at 1894–95.

80. *See* Reeves v. Sanderson Plumbing Prods., 530 U.S. 133, 143 (2000) ("Although intermediate evidentiary burdens shift back and forth under this framework, '[t]he ultimate burden of persuading the trier of fact that the defendant intentionally discriminated against the plaintiff remains at all times with the plaintiff.'") (quoting Tex. Dep't. of Cmty. Affairs v. Burdine, 450 U.S. 248, 253 (1981)); *Id.* at 142 ("This burden is one of production, not persuasion; it 'can involve no credibility assessment.'")(quoting St. Mary's Honor Ctr. v. Hicks, 509 U.S. 502, 509 (1993)); *see also* Christopher R. Hedican et al., McDonnell Douglas*: Alive and Well*, 52 DRAKE L. REV. 383, 398 n.102 (2004).

81. *Reeves*, 530 U.S. at 143. That the burden of persuasion never shifts is due to the relatively weak showing required of a plaintiff in order to make out a prima facie case under *McDonnell Douglas*. *See* Price Waterhouse v. Hopkins, 490 U.S. 228, 267, 270–71 (1989) (O'Connor, J., concurring). Essentially, a prima facie case under the traditional *McDonnell Douglas* framework merely rules out the "two most common" and easily identified legitimate reasons for an adverse employment action: that the plaintiff was not qualified (or less qualified) for the position and that the position sought was not vacant. *See* Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 358 n.44 (1977). Where a higher initial showing is made by the plaintiff, the "strong medicine" of a shift in the burden of persuasion can be justified. *See Price Waterhouse*, 490 U.S. at 262 (O'Connor, J., concurring).

82. *See Teamsters*, 431 U.S. 324.

to a shift in the burden of persuasion, not just the burden of production.[83]

In her concurring opinion in *Price Waterhouse*, Justice O'Connor pointed to *Teamsters* as an example of when the law permits a shift of the burden of persuasion, rather than just a shift in the burden of production.[84] In explaining that a shift in the burden of persuasion can be justified in appropriate cases, Justice O'Connor stressed the importance of judicial estimates of probabilities and the parties' relative access to proof.[85] When a plaintiff's initial showing has only a weak effect on the court's estimate of the probability of unlawful discriminatory intent (as in the plaintiff's showing in a *McDonnell Douglas* prima facie case),[86] a shift in only the burden of production will be appropriate. However, where a plaintiff's initial showing has a stronger effect on the court's estimate of the probability of discrimination, a shift in the burden of persuasion might be appropriate.[87] This difference highlights the importance of a careful assessment of the impact of Phase I evidence on probability estimates in systemic disparate treatment cases and will be explored further in later parts.

II. ALLOCATING BURDENS OF PRODUCTION AND PERSUASION IN CIVIL LITIGATION

As shown above, the systemic disparate treatment theory is a method for plaintiffs to create an inference of discriminatory intent—one of the required elements of a disparate treatment claim.

83. *See, e.g.*, *Price Waterhouse*, 490 U.S. at 267 (O'Connor, J., concurring) ("[O]ur decisions in *Teamsters* and *Franks* do indicate a recognition that presumptions shifting the burden of persuasion based on evidentiary probabilities and the policies behind the statute are not alien to our Title VII jurisprudence."); Lujan v. Franklin Cnty. Bd. of Educ., 766 F.2d 917, 929 (6th Cir. 1985); Craik v. Minn. State Univ. Bd., 731 F.2d 465, 470 (8th Cir. 1984); Linda Hamilton Krieger, *The Burdens of Equality: Burdens of Proof and Presumptions in Indian and American Civil Rights Law*, 47 AM. J. COMP. L. 89, 110–11 (1999); Paul M. Secunda, *A Public Interest Model for Applying Lost Chance Theory to Probabilistic Injuries in Employment Discrimination Cases*, 2005 WIS. L. REV. 747, 768 n.123 (2005).

84. *See Price Waterhouse*, 490 U.S. at 267 (O'Connor, J., concurring).

85. "If, as we noted in *Teamsters*, '[p]resumptions shifting the burden of proof are often created to reflect judicial evaluations of probabilities and to conform with a party's superior access to the proof,' . . . one would be hard pressed to think of a situation where it would be more appropriate to require the defendant to show that its decision would have been justified by wholly legitimate concerns." *Price Waterhouse*, 490 U.S. at 273 (O'Connor, J., concurring) (citation omitted) (quoting *Teamsters*, 431 U.S. at 359 n.45).

86. *See supra* note 81 and accompanying text; *see also Price Waterhouse*, 490 U.S. at 270 (O'Connor, J., concurring).

87. *See Price Waterhouse*, 490 U.S. at 266–67 (O'Connor, J., concurring) (noting that in a *Teamsters* situation, plaintiffs receive the benefit of a burden shift "based on the *likelihood* that an illegitimate criterion was a factor in the individual employment decision").

Teamsters provides an alternative to the *McDonnell Douglas* burden-shifting framework for creating this inference of discriminatory intent.[88] Because the plaintiffs' required showing in a *Teamsters* prima facie case is stronger than the required showing in a *McDonnell Douglas* prima facie case, the result is a shift in the burden of persuasion rather than just the burden of production on the question of discriminatory intent.[89]

Recognizing this operational role of the systemic disparate treatment theory is critical to determining its proper scope and application. The question highlighted by the areas of confusion described in Part I.B. is: "In which cases (or categories of cases) should statistical evidence of a 'pattern or practice' of discrimination be permitted to create an inference of discriminatory intent and shift the burdens of production and persuasion on the element of discriminatory intent from the plaintiff(s) to the defendant?" This is fundamentally a question about allocating burdens of proof. More specifically, it is a question about how the burdens of production and persuasion on the question of discriminatory intent should be allocated at an identifiable point in the middle of the litigation: after Phase I evidence is presented and considered, but before Phase II begins.

In order to better understand whether permitting (or prohibiting) use of the systemic disparate treatment burden allocation device is valid in any particular application in employment discrimination law, it is logical to begin by considering how burdens of production and persuasion are generally allocated in civil litigation.

A. Traditional Considerations

Traditional evidence and civil procedure scholars identified a number of considerations for determining how to allocate the burdens of proof in civil litigation.[90] Of the four most commonly asserted traditional burden allocation considerations, the two that

88. *Teamsters*, 431 U.S. at 358–60.

89. *See supra* note 81 and accompanying text.

90. Unlike later law and economics scholars, the traditional doctrinal scholars did not purport to set out a formula or equation for determining how the various considerations should be weighted. *See, e.g.*, JAMES ET AL., *supra* note 70, § 7.16, at 343 ("There is no a priori test for allocating the burden of persuasion or the burden of producing evidence."); MUELLER & KIRKPATRICK, *supra* note 72, § 3:3, at 430 ("[T]rial burdens are often allocated by specific custom . . . . Beyond these customs, there are some broad notions that account for the allocation of trial burdens."); *cf.* Bruce L. Hay & Kathryn E. Spier, *Burdens of Proof in Civil Litigation: An Economic Perspective*, 26 J. LEGAL STUD. 413 (1997).

prove most important in determining when discriminatory intent should be inferred in systemic disparate treatment cases are the same two identified in *Teamsters* and by Justice O'Connor in *Price Waterhouse*: (1) the parties' relative access to proof and (2) judicial estimates of probabilities.[91]

1. The Party Seeking to Change the Status Quo

The first consideration proffered in the traditional literature is that the party seeking to change the status quo should bear the evidentiary burdens. This explains why the burden of persuasion is normally assigned to plaintiffs in civil litigation: "Perhaps the broadest and most accepted idea is that the person who seeks court action should justify the request, which means that the plaintiffs bear the burdens on the elements of their claims."[92]

This is akin to a default rule in the traditional analysis. In the absence of some overriding consideration, the traditional analysis provides that the burdens of persuasion and production ought to be placed on the plaintiff, who is asking the court to disrupt the status quo allocation of rights or property.

A key limitation of this consideration is that it informs only how *initial* burdens should be allocated. Plaintiffs are always the ones who set the machinery of the courts in motion, and are therefore always the party asking for change from the status quo. This consideration invariably points to placing the initial burdens on plaintiffs; it offers no insight on what showings by plaintiff (made during the litigation) might be sufficient to *shift* these initial burdens onto defendant.[93] This first consideration offers no useful

91. *See, e.g.*, MCCORMICK, *supra* note 32, §§ 337, 343 (Edward W. Cleary, ed., 2d ed. 1972); JAMES ET AL., *supra* note 70, § 7.8; 9 JOHN HENRY WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 2486 (James H. Chadbourn, ed., rev. 1981); *see also* 31A C.J.S. EVIDENCE §§ 103–110, 112–113 (2010).

92. MUELLER & KIRKPATRICK, *supra* note 72, § 3:3, at 430; *see also* MCCORMICK, *supra* note 32, § 337, at 786 ("The burdens of pleading and proof with regard to most facts have been and should be assigned to the plaintiff who generally seeks to change the present state of affairs and who therefore naturally should be expected to bear the risk of failure of proof or persuasion.").

93. This consideration explains why the initial burden in a systemic disparate treatment case is on the plaintiffs to establish that the defendant employer engaged in a pattern or practice of discrimination. Because the plaintiffs are seeking court intervention, they should be required to provide some evidence to convince the court that action may be required. However, once the plaintiff has given the court some evidence that a change in the status quo may be justified, then other burden allocation considerations should be considered. *See generally* Hay & Spier, *supra* note 90, at 424–28 (noting that the default rule is to place the burdens of proof on plaintiffs, and that certain recognized exceptions to the

prescription for determining what types of showings, in what types of systemic disparate treatment cases, should be sufficient to raise an inference of discriminatory intent. For insight on that question, other considerations must be examined.

2. Relative Access to Proof

A second consideration discussed in the traditional scholarship is the parties' relative access to proof on the point in question. Professors Mueller and Kirkpatrick write: "[B]urdens are allocated to put them on the party most likely to be able to carry them, meaning the party most likely to have access to the proof."[94] Likewise, Professor James writes: "[t]he burden of proof traditionally is placed on the party having the readier access to knowledge about the fact in question."[95]

Access to proof is specifically mentioned by the *Teamsters* Court as a justification for shifting the burden of proof in appropriate employment discrimination cases.[96] Access to proof appears to be a particularly important consideration in disparate treatment employment discrimination cases. The key element in disparate

---

default rule can be explained by examining considerations other than the costs of initiating litigation).

94. MUELLER & KIRKPATRICK, *supra* note 72, § 3:3, at 432. For example, a debtor may more readily prove payment of a debt than a creditor proving nonpayment of the debt. *Id.* Policy reasons can sometimes override this consideration. Thus, a plaintiff ordinarily bears the burden of proving defendant's negligence, but a defendant must prove plaintiff's contributory negligence. "The explanation is that the party seeking relief must support the request (so the plaintiff has a central burden), but we prefer to provide full recovery if negligence is shown unless the defendant provides good reasons to deny or reduce recovery by showing that the plaintiff is all or partly at fault." *Id.* at 433. *See also* MCCORMICK, *supra* note 32, § 337, at 787 ("A doctrine often repeated by the courts is that where the facts with regard to an issue lie peculiarly in the knowledge of a party, that party has the burden of proving the issue . . . . This consideration should not be overemphasized.").

95. JAMES ET AL., *supra* note 70, § 7.16, at 344. Professor James notes that "[t]his consideration, however, has never been controlling." *Id.* As contrary examples, Professor James notes that that the plaintiffs usually have the burden of proving that defendant's conduct was negligent in a negligence claim, proving that defendant committed breach in a contract action, and proving the falsity of defendant's representations in a fraud claim. *Id. See also id.* § 7.17, at 349 ("Access to evidence is often the basis for creating a presumption. When goods are damaged in a bailee's possession, for instance, the bailee can more easily find out what happened to them than the bailor, so it is fair to presume the bailee's negligence as an initial matter and put the bailee to the production of exculpatory evidence, if any exists.").

96. *See* Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 359 n.45 (1977) (citing MCCORMICK, *supra* note 32, §§ 337, 343); *see also* Price Waterhouse v. Hopkins, 490 U.S. 228, 273 (1988) (O'Connor, J., concurring); Keyes v. Sch. Dist. No. 1, 413 U.S. 189, 208–09 (1973) (discussing burden shifting in the context of school discrimination cases); James, *supra* note 32, at 61 (describing burdens of proof as "handicaps" used against "disfavored contentions").

treatment cases is a showing of discriminatory intent—the element on which *Teamsters* provides a mechanism for drawing an inference in plaintiffs' favor. Proving that an employer's decision-maker acted with an intentionally discriminatory motive is a particularly difficult task for plaintiffs, given their lack of access to information on that point.[97]

Commentators have sometimes minimized the importance of this consideration in light of modern discovery rules, which provide for liberal exchange of information.[98] Access to relevant evidence, the argument goes, can be obtained through document requests, depositions, and other discovery techniques.[99] However, this is less likely to be true in disparate treatment discrimination cases, where the central issue is usually what was in the decision-maker's mind when he took an adverse employment action against the plaintiff. While modern discovery rules can complicate the question of the parties' relative access to proof, it seems unlikely that they can completely equalize the parties' ability to offer proof on all factual questions—especially in disparate treatment cases. The impact of modern discovery rules on information asymmetries is discussed in detail, *infra*, Part II.B.

3. Estimates of Probabilities

Traditional evidence and procedure scholarship also recognizes the importance of judicial estimates of which party's assertion was most probably true or correct. Professor McCormick stated: "Perhaps a more frequently significant consideration in the fixing of the burdens of proof is the judicial estimate of the probabilities of the situation. The risk of failure of proof may be placed upon the party who contends that the more unusual event has occurred."[100]

---

97. *See* Katz, *supra* note 45, at 881.

98. *See* JAMES ET AL., *supra* note 70, § 7.16, at 344 ("Access to evidence is a much diminished basis for allocating burden of proof in modern liberal discovery."); *see also* Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1503 (1999) ("But if, as a result of modern pretrial procedures for discovering evidence in the possession of the opposing party, the costs of searching for evidence are symmetrical, the burden of production should indeed be on the party that bears the burden of persuasion—that, is [sic] the plaintiff in the case of the main claim but the defendant in the case of affirmative defenses . . . .").

99. *See* JAMES ET AL., *supra* note 70, § 7.16.

100. MCCORMICK, *supra* note 32, § 337, at 787. Professor McCormick offers the following example: "[W]here a business relationship exists, it is unlikely that services will be performed gratuitously. The burden of proving a gift is therefore placed upon the one who claims it. Where services are performed for a member of the family, a gift is much more likely and the burden of proof is placed on the party claiming the right to be paid." *Id.*; *see*

Mueller and Kirkpatrick agree, writing: "[B]urdens are allocated to recognize what is probably true."[101] The judicial estimate of probabilities is specifically mentioned as a consideration in shifting the burden of proof in the pattern or practice context in *Teamsters*,[102] and by Justice O'Connor in her concurrence in *Price Waterhouse*.[103]

Judicial estimates of probabilities play a key role not only in initial burden allocations, but also in the operation and strength of many different types of legal presumptions and inferences that are employed after an initial showing by one party. "What is *likely*, for instance, is often presumed."[104] Because "most letters, properly sent, reach their destination" upon a showing that a letter is properly mailed, the law will presume that it reached its destination.[105] Likewise, in the *McDonnell Douglas* framework, a plaintiff's prima facie case is sufficient to rule out the two most common reasons for an adverse employment action—that plaintiff did not apply for an open position or was not qualified for it. This showing by the plaintiff is sufficient to show an increase in the probability that the adverse employment action was the product of discrimination, but it is not a strong enough showing to establish a *very high* probability of discrimination. As a result, only the burden of production is shifted under *McDonnell Douglas*, and the employer can easily meet that burden by providing some evidence of a legitimate, non-discriminatory reason for the adverse action.[106] Where the plaintiff's initial showing is stronger and makes the probability of discrimination significantly higher, however, a shift in the burden of persuasion may be justified.[107]

---

also Edward W. Cleary, *Presuming and Pleading: An Essay on Juristic Immaturity*, 12 STAN. L. REV. 5, 13 (1959) (asserting that the judicial estimate of probabilities should consider litigated cases as the baseline). *But see* V.C. Ball, *The Moment of Truth: Probability Theory and Standards of Proof*, 14 VAND. L. REV. 807, 817–18 (1961) (arguing against the use of probabilities in assigning the burden of persuasion).

101. MUELLER & KIRKPATRICK, *supra* note 72, § 3:3, at 433.

102. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 359 n.45 (1977) (citing MCCORMICK, *supra* note 32, §§ 337, 343; James, *supra* note 32, at 61; and Keyes v. Sch. Dist. No. 1, 413 U.S. 189, 208–09 (1973)).

103. Price Waterhouse v. Hopkins, 490 U.S. 228, 273 (1988) (O'Connor, J., concurring).

104. JAMES ET AL., *supra* note 70, § 7.17, at 348.

105. *Id.* at 348 n.19 (citing JOHN MACARTHUR MAGUIRE, EVIDENCE 183 (1947) and WIGMORE, *supra* note 91, § 2491).

106. *See supra* note 81 and accompanying text.

107. *See Price Waterhouse*, 490 U.S. at 261–62 (O'Connor, J., concurring). If, for example, a plaintiff can establish that race was a "motivating factor," the probability of unlawful discrimination is much higher (liability is established) and the burden of persuasion shifts to the defendants to prove that there was an independently sufficient reason that motivated

4. Substantive Policy and Disfavored Contentions

A final consideration discussed in the traditional evidence and procedure literature is whether the law favors or disfavors one of the parties' positions as a substantive policy matter.[108] For example, the law generally disfavors claims of libel or slander involving matters of public interest in order to adequately protect freedom of speech.[109] Therefore, a public figure plaintiff bears the burden of persuasion on the matter of whether the alleged defamatory statement was true or false, and must prove that the defendant acted with "actual malice" regarding the truth or falsity of the statement.[110] In this way, the burden of proof is "used as a handicap against the disfavored contention."[111]

Another example is fraud, for which courts often require a higher level of proof from plaintiff in order to satisfy the burden of persuasion.[112] Likewise, many affirmative defenses may generally be disfavored for substantive policy reasons, explaining why the burden of persuasion is often placed on the defendant rather than the plaintiff for those defenses.[113] Substantive policies may also explain courts' use of some legal presumptions to allocate or shift burdens.[114]

"Substantive policy" is a broad label that can cover different types of judicial considerations. It may be that there is a consensus in the law that Type II errors (false negatives) should be favored over Type I errors (false positives), or vice versa, for certain categories of cases.[115] This is likely the case in the context of the

---

them to take the adverse action in order to avoid damages. *See id.*; Civil Rights Act of 1991, 42 U.S.C. §§ 2000e-2(m), 2000e-5(g)(2)(B) (2006).

108. *See* MUELLER & KIRKPATRICK, *supra* note 72, § 3:3, at 431 ("[B]urdens are allocated to serve substantive policy, making it easier or harder for plaintiffs to recover or defendants to avoid liability.").

109. JAMES ET AL., *supra* note 70, § 7.16, at 345 ("Thus, the defendant in libel or slander formerly was required to prove the truth of the objectionable words to escape liability. The evolved interpretation of the First Amendment, however, requires plaintiff to prove the falsity of a statement in news media about a subject of public interest.").

110. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 285–86 (1964); *see also* JAMES ET AL., *supra* note 70, § 7.16.

111. JAMES ET AL., *supra* note 70, § 7.16, at 345.

112. *See* MUELLER & KIRKPATRICK, *supra* note 72, § 3:3, 431 ("Sometimes the policy is more negative than positive, as is true of some disfavored or handicapped claims (such as fraud) or defenses (such as statute of limitations), which are allocated as burdens on the party who claims them and also sometimes made subject to heavier pleading and persuasion burdens.").

113. *See id.* § 3.3.

114. *See id.*

115. Criminal law is an example of a situation in which the law so heavily prefers Type II errors (false acquittals) over Type I errors (false convictions) that we not only place the

defamation and fraud examples described above. It may be better that some defendants guilty of libel go unpunished in order to provide breathing space for First Amendment freedoms. Likewise, it may be preferable to err on the side of not finding fraud, absent a high degree of confidence that fraud was actually committed, because of the especially severe consequences to such defendants.[116] In order to favor Type II errors over Type I errors, the law uses burden allocation as an encumbrance against the disfavored position. Unlike defamation and fraud, there is no apparent reason to favor either Type I errors or Type II errors in employment discrimination cases.[117] There is no consensus that there should be a thumb on either side of the scale in discrimination suits.[118]

Alternatively, the label "substantive policy" may also describe situations in which the considerations at issue in altering the ordinary burden allocations are really the parties' relative access to information and judicial estimates of probabilities. The doctrine of *res ipsa loquitur* and the alternative liability rule of *Summers v. Tice*[119] provide examples. In each of these situations, the law alters burden allocations in some way, ostensibly for "policy" reasons, after plaintiff makes a certain evidentiary showing.[120]

The doctrine of *res ipsa loquitur* can relieve the plaintiff from the burden of proving the defendant's negligence if the plaintiff can first show: (1) the defendant was in exclusive control of the instrument that led to plaintiff's injury, (2) the plaintiff's conduct did not contribute to the injury, and (3) the injury could not ordinarily occur without defendant's negligence.[121] Many jurisdictions characterize this rule as permitting, but not requiring, the jury to

---

burden of persuasion on the prosecution, but we also require an especially high level of proof: proof beyond a reasonable doubt. *See generally* Posner, *supra* note 98, at 1504–05.

116. *See* Bruce L. Hay, *Allocating the Burden of Proof*, 72 IND. L.J. 651, 678 (1997).

117. *Cf.* Posner, *supra* note 98, at 1504–05 ("In the typical civil trial, there is no basis for supposing that Type I errors (false positives, such as convicting an innocent person) on average impose higher costs than Type II errors (false negatives, such as an erroneous acquittal). So it is enough to justify a verdict for the plaintiff that the probability that his claim is meritorious exceeds, however slightly, the probability that it is not.").

118. *See* SULLIVAN ET AL., *supra* note 1, § 3.04[B], at 178 ("While this outcome [favoring Type II errors over Type I errors] may be appropriate in criminal law, where the primary value is to not convict the innocent, there is no similar policy in civil litigation. The setting of the standard of proof at the preponderance of evidence in civil litigation means that there is no important social value difference between finding an innocent defendant liable and a guilty defendant not liable.").

119. 199 P.2d 1 (Cal. 1948).

120. *See* JAMES ET AL., *supra* note 70, at § 7.16; *see also Price Waterhouse*, 490 U.S. at 263–64 (O'Connor, J., concurring) (citing *Summers* as an example of shifting the burden of persuasion onto defendants).

121. *See* JAMES ET AL., *supra* note 70, at § 7.16 (referring to the *res ipsa loquitur* doctrine as a "substantively influenced rule of allocation.").

presume or infer negligence once the plaintiff has made such a showing.[122]

Likewise, the rule of alternative liability in cases like *Summers* can alter the ordinary burden allocations in civil cases. Where the plaintiff proves that his injury was caused by the negligence of one of several tortfeasors, but cannot prove that the injury was more likely than not the result of the negligence of any one particular tortfeasor, the burden of persuasion shifts onto the defendants to prove that they were not responsible for plaintiff's injuries.[123]

The *res ipsa loquitur* and *Summers* alternative liability doctrines can be explained by the defendants' superior access to proof of negligence and the high probability that one of the defendants' negligence caused the injury (given the showings made by plaintiff to trigger the doctrines). The "substantive policy" reasons for disfavoring one party's position or the other's can often be collapsed down into an analysis of access to information and probabilities.

Of the four burden allocation considerations discussed in the traditional evidence and civil procedure scholarship, the two that can best help determine the appropriate use of the systemic disparate treatment theory to create an inference of discriminatory intent are: (1) relative access to proof and (2) judicial estimates of probabilities.

B. Law and Economics Analysis

Law and economics scholars approach burden allocation questions from a different perspective, and their insights can assist in determining the efficient application of the systemic disparate treatment theory. Economists have formally modeled the efficient allocation of burdens of proof in civil litigation and offer a more rigorous and detailed analysis of the key considerations. These economic models, like the traditional analysis, point to the importance of the parties' relative access to information and estimates of probabilities.

122. *See* DAN B. DOBBS, THE LAW OF TORTS, § 154, at 370–71 (2000); *see, e.g.*, Khan v. Singh, 975 A.2d 389, 394 (N.J. Sup. Ct. 2009) (the doctrine of *res ipsa loquitur* "permits the jury to infer negligence in certain circumstances, effectively reducing the plaintiff's burden of persuasion, but not shifting the burden of proof.").

123. *See* DOBBS, *supra* note 122, § 175, at 426–27.

1. The Hay and Spier Model: Information Costs and Bayesian Probability Analysis

Professors Bruce Hay and Kathryn Spier, in a 1997 article entitled *Burdens of Proof in Civil Litigation: An Economic Perspective*, developed a formal model that is well-suited to analyzing systemic disparate treatment law.[124] Hay and Spier begin with a relatively simple game theory model to evaluate litigants' dominant strategies for presenting evidence to the court,[125] and determine that the party bearing the burden of proof will present evidence on a disputed issue of fact if and only if the evidence supports his position. The party not bearing the burden of proof will refrain from presenting evidence regardless of whether the evidence supports his position.[126]

Having modeled the parties' equilibrium behavior under alternate burden assignments using game theory, Hay and Spier then proceed to determine the optimal allocation of the burden of proof given such predicted litigant behavior.[127] The stated goal of

124. *See* Hay & Spier, *supra* note 90.

125. The three simplifying assumptions of Hay & Spier's game theory model are: (1) both parties have access to a body of evidence that, if presented to the court, indicates whether the defendant acted unlawfully (the issue in the case), and both parties know what that body of evidence contains; (2) the body of evidence is unitary in nature, such that the court either sees all of it or none of it; and (3) a party's cost of presenting the evidence to the court is low enough that he would choose to present the evidence on the issue if it is necessary to do so in order for that party to obtain a favorable ruling. *Id.* at 416–17.

126. *Id.* at 418. This is demonstrated by the following analysis: If the plaintiff bears the burden of proof of showing that *X* occurred (the defendant's unlawful activity), and the parties are aware of the burden assignment, then the defendant's dominant strategy is to not produce evidence. If the evidence shows that *X* occurred, then defendant will not offer the evidence because it hurts his case and he would needlessly incur presentation costs by offering the evidence. If the evidence shows that *X* did not occur, then the defendant will also not offer the evidence, because he will prevail even if no evidence is offered, and save the costs of presenting the evidence. If the evidence shows that *X* occurred, then plaintiff will present the evidence in order to prevail (assuming that presentation costs are sufficiently low, per assumption number 3). If the evidence shows that *X* did not occur, then plaintiff will not present the evidence because it hurts his case and he would pointlessly incur presentation costs. *Id.* at 417. The converse holds if the burden is assigned to the defendant. *Id.* at 417–18.

127. Hay & Spier assert that their model of the "burden of proof" allocation corresponds to the burden of production, rather than the burden of persuasion, although they acknowledge that the distinction between the production and persuasion burdens "is of course blurry." Hay & Spier, *supra* note 90, at 415 nn.3–4. However, as Professor Sanchirico correctly points out, Hay & Spier are actually modeling an amalgam of both burdens. Sanchirico *supra* note 70, at 281 ("The fact that the burdened party loses [in the Hay & Spier model] when neither party presents any evidence resembles the effect of the production burden. The fact that the burdened party loses when both parties present evidence resembles the effect of the persuasion burden."). This simplified analysis of an amalgam of both trial burdens at once is somewhat crude, but is also "standard in the economic literature on legal process," and has also been used by Professor Sanchirico, *see id.* at 301 n.32, and Hyun

the model is that "the court wants to assign the burden of proof to minimize the expected costs of presenting evidence on whether *X* occurred," where *X* represents the defendant's unlawful behavior.[128]

Having this stated goal, Hay and Spier posit that the court should give the plaintiff the burden of proof if, and only if, the following expression holds:

> (probability that *X* occurred) x (plaintiff's costs of showing *X* occurred) < (probability that *X* did not occur) x (defendant's costs of showing *X* did not occur)[129]

Because of the parties' equilibrium behavior, the left side of this equation represents the expected costs of giving the burden to the plaintiff, while the right side of the equation represents the expected costs of imposing the burden on defendant.[130] The factors Hay and Spier identify as the key determinants of proper burden of proof allocation are: (1) the probability of unlawful behavior and (2) the costs to each party of accessing and presenting evidence. These are the same two factors highlighted in the traditional scholarship, and the same two factors that the Supreme Court in *Teamsters* identified as justifying a shift in the burden of proof.[131]

a. The Parties' Relative Information Costs

The parties' relative information costs are the "costs of gathering and presenting evidence on the contested issue."[132] If one party has better access to the evidence on the question at issue in the case, or a better ability to assemble and present the evidence at a lower cost, then that party should (all other things being held equal) be

---

Song Shin, *Adversarial and Inquisitorial Procedures in Arbitration*, 29 RAND J. ECON. 378, 380–81 (1998). The formal modeling performed by Professor Hay in his related article, *Allocating the Burden of Proof*, likewise analyzes an amalgam of the burdens. *See* Hay, *supra* note 116, at 653–55.

128. Hay & Spier, *supra* note 90, at 418.

129. *Id.*

130. *See id.* at 418–19. Because the burdened party will incur the presentation costs—if any—only if the evidence supports that party's position, each side of the equation represents the expected costs of placing the burden on that party. Thus, if the expected costs of placing the burden on the plaintiff are less than the expected costs of placing the burden on defendant, then the burden should be placed on plaintiff.

131. *See* Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 359 n.45 (1977). *See also* Hay & Spier, *supra* note 90, at 419 n.10 ("These factors are widely recognized as the essential ones in assigning the burden of proof on an issue.") (citing JAMES ET AL., *supra* note 70, at 344–49 (4th ed. 1992)).

132. Hay & Spier, *supra* note 90, at 419.

given the burden of proof. In other words, "[o]ther things being equal, the lower one party's relative costs, the stronger the argument for giving him the burden of proof."[133]

Hay and Spier note that modern discovery rules can complicate this question. Under the Federal Rules of Civil Procedure ("Fed. R. Civ. P."), each party can request the production of evidence in the other party's possession and depose the other parties' witnesses.[134] Judge Richard Posner, in his economic analysis of evidence law, posits that modern pretrial discovery procedures can make the costs of searching for evidence symmetrical.[135]

Hay and Spier are less optimistic on this point. They note that if the defendant has exclusive possession of the relevant evidence, giving plaintiff the burden of proof may result in the defendant complying with discovery rules by sending too much information, forcing the plaintiff to sift through the piles to find relevant evidence, thus incurring a real information cost.[136] However, this is only one way in which the Fed. R. Civ. P. might come up short in equalizing asymmetrical information costs. An even more obvious way, in the context of intentional employment discrimination, is that reliable evidence of the decision-maker's true intent is essentially impossible for a plaintiff to obtain using any of the discovery tools. For example, the plaintiff in a sex discrimination suit is highly unlikely to get a "guilty" decision-maker, even under oath in a deposition, to admit that the actual reason he fired plaintiff is that she is a woman. In the typical discrimination case, not only is direct evidence of discriminatory intent solely within the defendant's possession, it is also particularly elusive even under modern discovery rules.

b. Probabilities and Signals

The second determinative factor is the probability that *X* (defendant's unlawful behavior) occurred. In a discrimination case, for example, *X* would represent the probability that the employer actually discriminated against the plaintiff based on age, sex, race, or other protected characteristic. Hay and Spier further refine the probabilities factor by applying a Bayesian probability analysis. They posit that the court begins with some information, or signal,

133. *Id.* at 419.
134. *See* FED. R. CIV. P. 26–37.
135. *See* Posner, *supra* note 98, at 1503.
136. *See* Hay & Spier, *supra* note 90, at 419.

about the case (Y).[137] For example, in a discriminatory failure to hire case where the plaintiff has successfully established a prima facie case under *McDonnell Douglas*, the court would have the following information (or signal Y): that the plaintiff belonged to a protected class, plaintiff was qualified for and applied for the open position, plaintiff was not hired, and that a person from outside the protected class was hired. The observation of such a signal allows the court to refine or update its understanding of the probability factor using Bayes' Rule.[138]

The probability that X occurred (after taking into account the observation of signal Y) can be represented as:

$$\frac{\text{prob}(Y|X) \text{ x prob}(X)}{\text{prob}(Y)}$$

In this expression, prob($Y|X$) represents the probability that the court would observe signal Y given that X occurred, prob(X) represents the unconditional prior probability that X would occur, and prob(Y) represents the unconditional prior probability of observing Y.

Likewise, the probability that X did not occur (after taking into account the observation of signal Y) can be represented as:

$$\frac{\text{prob}(Y|\sim X) \text{ x prob}(\sim X)}{\text{prob}(Y)}$$

In this expression, prob($Y|\sim X$) represents the probability that the court would observe signal Y given that X *did not* occur, prob($\sim X$) represents the unconditional prior probability that X did not occur, and prob(Y) represents the unconditional prior probability of observing Y.[139]

Returning to Hay and Spier's determinative expression for placing the burden, substituting the foregoing refinements for the probability that X occurred (or did not occur) and simplifying, the following expression obtains:

---

137. For purposes of analyzing pattern or practice burden-shifting, I will assume that signal Y is the observation of statistical evidence suggesting that a prohibited independent variable (i.e., race, sex, or other protected classifications) was a statistically significant predictor of the adverse employment action. *See* discussion *infra* Part III–IV

138. For a further discussion and defense of Bayesian probability analysis in the analysis of evidence, see Richard D. Friedman, *"E" is for Eclectic: Multiple Perspectives on Evidence*, 87 VA. L. REV. 2029, 2041–45 (2001). For a recent argument in favor of application of Bayesian probability analysis to the interpretation of contracts, see Yair Listokin, *Bayesian Contractual Interpretation*, 39 J. LEGAL STUD. 359 (2010).

139. *See* Hay & Spier, *supra* note 90, at 419–20 (setting out these expressions).

prob($Y|X$) x prob(X) x (plaintiff's costs of showing that X occurred) <
prob($Y|\sim X$) x prob($\sim X$) x (defendant's costs of showing that X did not occur)

Thus, the burden of proof should be placed on the plaintiff if and only if the foregoing expression holds.[140] Simplifying from the Hay and Spier expression, there are three key factors that determine how to allocate the burdens of proof:

(1) *Prior Probability*: (prob(X)) as compared to (prob($\sim X$)). The prior, unconditional likelihood that X occurred, as compared to the prior, unconditional likelihood that X did not occur, before considering any evidence in the case. All else held equal, the higher (prob(X)), the stronger the argument for placing the burden on defendant. Conversely, all else held equal, the higher (prob($\sim X$)) the stronger the argument for placing the burden on plaintiff.

(2) *Evidentiary Signal*: (prob($Y|X$)) as compared to (prob($Y|\sim X$)). The probability that the court would observe signal Y if X did occur, as compared to the probability that the court would observe that same signal Y if X did not occur. All else held equal, the higher (prob($Y|X$)), the stronger the argument for placing the burden on defendant. Conversely, all else held equal, the higher (prob($Y|\sim X$)), the stronger the argument for placing the burden on plaintiff.

(3) *Information Costs*: The relative costs facing each party in searching for, finding, and presenting evidence that X did (or did not) occur. All else held equal, the higher plaintiff's relative information costs, compared to the defendant's, the stronger the argument for placing the burden on defendant, and vice-versa.

The results of the Hay and Spier model reinforce the importance of relative access to proof and estimates of probabilities in allocating burdens. The key difference from the traditional analysis is that the probabilities factor is further refined, using

140. *See* Hay & Spier, *supra* note 90, at 423.

Bayes' theorem, into two components: a prior probability and an update to the prior.

The Hay and Spier model, by utilizing Bayesian probability analysis, is particularly well suited to addressing systemic disparate treatment cases. In a systemic disparate treatment case, the court must decide, at an identifiable point in the middle of the litigation and after the initial burden assignment, whether to permit an inference of discriminatory intent and shift the burden of persuasion to the defendant. In other words, the court will have an opportunity to consider some evidence (the Phase I evidence, usually statistical evidence) and update its estimate of probabilities based on this evidentiary signal. By incorporating the Bayesian concept of probability updating, the Hay and Spier model provides an excellent framework for analyzing systemic disparate treatment cases.

2. Economic Analysis Confirms the Importance of Traditional Considerations

The Hay and Spier model is a theoretical tool meant to sharpen analysis and provide a framework for thinking about the costs and benefits of allocating burdens of proof to one party or the other in given situations. Certainly, not every assumption underlying the model will hold true in every case. Moreover, it is essentially impossible for a court or policy-maker to assign specific values to any of the variables in the equation with any degree of mathematical precision. Nobody can attach specific values, with absolute certainty, to the variables (prob (X)), (prob ($\sim X$)), or identify with precision the difference in the information presentation costs of plaintiff and defendant.

Nevertheless, the Hay and Spier model provides useful guidance in allocating burdens of proof in systemic disparate treatment discrimination cases. The model reinforces the importance of the parties' relative access to information and estimates of probabilities. The model identifies the three key factors that should be essential in determining whether a shift in the burden of persuasion, or permitting an inference of discriminatory intent, would tend to reduce the overall costs of accurately resolving the dispute. The economic model can serve as a theoretical guide in systemic disparate treatment cases, despite its inherent assumptions and the mathematical imprecision in its application.[141]

141. In a separate paper, Professor Hay further addresses many of the underlying assumptions and simplifications in the basic model. *See* Hay, *supra* note 116, at 665–73. Hay

As the *Teamsters* Court stated, "[p]resumptions shifting the burden of proof are often created to reflect judicial evaluations of probabilities and to conform with a party's superior access to the proof."[142] Economic modeling reinforces the importance of probabilities and information asymmetries.[143]

III. Application in Individual Cases: Analyzing Probabilities and Information Costs in the Systemic Disparate Treatment Discrimination Setting

Having identified the three key factors for proper allocation and shifting of the burdens of production and persuasion, Part III considers how each of the key factors is likely to arise in a systemic disparate treatment case. Case characteristics that courts should look for when determining whether to draw an inference of discriminatory intent in any particular case are identified.

A. Prior Probabilities in Disparate Treatment Cases

The first factor for analysis is the prior probability that discrimination occurred against any particular individual claimant (prob(X)) as compared to the prior probability of nondiscrimination (prob($\sim X$)), before considering the statistical evidence in the case. In the discrimination context, the question is: How prevalent is the type of discrimination alleged by the plaintiff—knowing nothing else about the employer, the plaintiff, or the circumstances? Critics of the use of statistical evidence in systemic employment discrimination cases charge that this factor is too often completely overlooked in courts' interpretation of proffered statistical evi-

models the effect of introducing the possibility of settlement between the parties, the signals sent by the parties' litigation decisions, and the possibility of litigant uncertainty about what the evidence shows. *See id.* Because of these modifications, Hay concludes that three other variables might factor into the efficient allocation of burdens of proof: (1) the amount at stake for the party, (2) the social cost of error against the party, and (3) the party's "optimism," or level of confidence that the evidence will support her position. *See id.* However, none of these three factors change the basic conclusions from the Hay & Spier model.

142. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 359 n. 45 (1977)(citing MCCORMICK, *supra* note 32, §§ 337, 343 and James, *supra* note 32, at 61).

143. This may be an example of what Professor Richard Lempert described as "common sense on stilts." *See* Richard Lempert, *The Economic Analysis of Evidence Law: Common Sense on Stilts*, 87 VA. L. REV. 1619, 1623 (2001). It may be that common sense arrived at the "just" and "fair" result of focusing on probabilities and access to information when drawing inferences, and only later did economic analysis demonstrate that this result was also efficient. However, the model's emphasis on Bayesian probability updating can provide a new perspective in systemic disparate treatment cases.

dence.[144] Courts in discrimination cases sometimes fall victim to what some commentators describe as a "statistical fallacy"—interpreting the results of binomial distribution analysis or regression analysis incorrectly by failing to explicitly take into account the prior, unconditional probability of *X*.[145]

These critics are correct, to the extent that they argue prior probabilities cannot simply be ignored in the interpretation of statistical evidence. In order to make informed estimates of probabilities, the prior probability must be considered. But determining with a high degree of confidence the prior probability of (*X*) in discrimination cases is unquestionably difficult.[146] Courts, scholars, and social scientists would likely disagree about the incidence of intentional discrimination in today's workplaces.[147] It would be unsurprising if some would estimate (prob(*X*)) as high as .9 for certain types of discrimination, while others might put it as low as .001. The possibility of discrimination resulting from purely

144. *See* Kingsley R. Browne, *Statistical Proof of Discrimination: Beyond "Damned Lies"*, 68 WASH. L. REV. 477, 488 (1993) ("In the discrimination context, the probability that an employer's work-force disparities are a consequence of chance is completely dependent upon a statistic which the courts never have: the likelihood of discrimination prior to making the employment decision. Although one might attempt some estimate of the percentage of employers who engage in systematic discrimination, the estimate could be no more than the crudest approximation.") While Browne is correct that an estimate of priors will necessarily be a crude approximation, that does not render statistical evidence useless. As shown below, arriving at a crude approximation of the prior is possible. In any event, estimating a prior is a problem for the consideration of any type of circumstantial evidence; not just for statistical analyses.

145. *See id.* at 490–92. Courts will sometimes incorrectly interpret statistical significance at the .05 level to mean that the observed imbalanced distribution in the workplace (of, for example, minority hires) would be due to chance only 5% of the time, and that there is therefore a 95% chance the observed distribution was caused by some other "suspicious" factor—namely, discrimination. *See id.*; *see also* Meier et al., *supra* note 30, at 148 (noting the same statistical fallacy). This interpretation fails to adequately consider the prior, unconditional probability of discrimination. If, by hypothesis, the alleged type of discrimination (in this example, discrimination against minorities in hiring) occurs only .01% of the time overall, then the probability that the observed distribution of minority hires for this employer was due to chance is actually much higher than 5%. An accurate assessment of probabilities must take into account both the statistical evidence (whether that be in the form of binomial distributions or regression analysis) and the relevant prior, unconditional probability.

146. *See* Browne, *supra* note 144, at 488 (calling the prior, unconditional probability "a statistic which the courts never have").

147. *See generally* Calloway, *supra* note 30 (arguing that courts are overly skeptical about the prevalence of race discrimination); Tristin K. Green, *A Structural Approach As Antidiscrimination Mandate: Locating Employer Wrong*, 60 VAND. L. REV. 849, 854–57 (2007) (discussing empirical evidence that shows discriminatory bias in the workplace remains a widespread problem, although discrimination is now more implicit and "behind the scenes" than in decades past); Michael Selmi, *Why Are Employment Discrimination Cases So Hard to Win?*, 61 LA. L. REV. 555, 561–69 (2001) (courts' biases may cause them to be overly skeptical about the prevalence of discrimination).

structural or unconscious biases in the workplace may complicate this estimation even further.[148]

But courts do have *some* information about (prob(X)) before the case begins. They have two items of information. First, they know the parties are litigating. The fact that the parties have not been able to reach agreement without resorting to litigation and have expended the resources to collect and present evidence at Phase I of the case sends at least some signal to the court about the case, before any other evidence is considered. There is some reason to believe that (prob(X)) and (prob($\sim X$)) are at least *reasonably* close to .5 each.[149] If the probabilities were at either extreme, then the case would be a likely candidate for settlement. Of course, this is not a satisfactory estimate. But it is at least one indicator available to the court of the relative prior likelihood of X.[150]

The second, and more specific, item of information courts have about (prob(X)) is a body of empirical research on the success rates of plaintiffs in federal employment discrimination litigation. Professors Kevin Clermont and Stewart Schwab have studied the rates of success for plaintiffs in federal employment litigation at the trial and appellate levels, and compared them with similar statistics

148. There is an ongoing debate in the literature about whether current antidiscrimination laws are effective against structural discrimination. *See generally* Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CAL. L. REV. 1 (2006); Tristin K. Green, *Discrimination in Workplace Dynamics: Toward a Structural Account of Disparate Treatment Theory*, 38 HARV. C.R.-C.L. L. REV. 91 (2003). If structural discrimination is unlawful under current antidiscrimination laws, then the estimate of the prior, unconditional probability that X (unlawful discrimination) occurred would arguably need to be adjusted upward accordingly.

149. *See* Posner, *supra* note 98, at 1507. A reasonable guess, in light of this information and *only* this information, might put (prob(X)) somewhere between, say, .3 and .7. Viewing the problem from a defense perspective, however, one might argue that because plaintiffs have an incentive to bring nonmeritorious claims in order to extract a settlement, a better guess would be that (prob(X)) falls between, say, .05 and .4. A pro-plaintiff view might argue the contrary, asserting that it is extremely difficult for plaintiffs with weak claims to find an attorney willing to bring the case before a court. Under that view, one might estimate (prob(X)) as falling somewhere between, say, .6 and .9. This range of possible opinions about prior probabilities highlights the difficulty of stating (prob(X)) with any mathematical precision, but it does not detract from the analytical power of the model. *See generally* Ramona L. Paetzold, *Problems with Statistical Significance in Employment Discrimination Litigation*, 26 NEW ENG. L. REV. 395, 413–14 ("Although it has often been argued in many contexts that frequentist procedures should be preferred because, not requiring the injection of prior beliefs, they are more 'objective,' the fact remains that virtually any decision rule that a frequentist employs is a Bayesian decision rule with respect to some assignment of prior probabilities. Simply pretending that prior beliefs are not being used does not make them or their effects go away . . . .") (footnote omitted).

150. Professor Browne does not consider the possibility that the prior unconditional probability may be affected by the signal to the court from the parties' litigation decisions. *See* Browne, *supra* note 144.

for other types of federal litigation.[151] In their most recent study, they found that for the years 1998–2006, the win rate for Title VII plaintiffs in federal district court was 10.88%, as compared with 9.12% for ADA plaintiffs, 11.67% for ADEA plaintiffs, and 10.96% for Section 1981 race discrimination plaintiffs.[152] Over the period from 1979–2006, plaintiffs in all types of federal employment cases won only about 15% of the time, while plaintiffs in other types of federal litigation won approximately 51% of the time.[153] For cases actually reaching trial, the win rates were approximately 28% for employment plaintiffs, compared to 45% for other federal plaintiffs.[154] These numbers can help give courts a basis for estimating (prob(X)).[155]

But using these win rate numbers as a rough gauge of (prob(X)) is not a perfect solution, either. Perhaps, as Clermont and Schwab and others have suggested, the win rates are much lower for plaintiffs in employment cases not because the claims are less likely to be meritorious, but because of some unique difficulty in overcoming certain types of hurdles or roadblocks in employment discrimination litigation, such as pretrial motions to dismiss and motions for summary judgment.[156] If that were the case, then the win rates might significantly understate the true (prob(X)). On the other hand, perhaps the win rates are lower in employment cases simply because those suffering adverse employment actions are more willing than other types of plaintiffs to bring non-meritorious

151. *See* Kevin M. Clermont & Stewart J. Schwab, *Employment Discrimination Plaintiffs in Federal Court: From Bad to Worse?*, 3 HARV. L. & POL'Y REV. 103 (2009) [hereinafter Clermont & Schwab, *From Bad to Worse*]; Kevin M. Clermont & Stewart J. Schwab, *How Employment Discrimination Plaintiffs Fare in Federal Court*, 1 J. EMPIRICAL L. STUD. 429 (2004) [hereinafter Clermont & Schwab, *Employment Discrimination Plaintiffs in Federal Court*]; *see also* Kevin M. Clermont et al., *How Employment-Discrimination Plaintiffs Fare in the Federal Courts of Appeals*, 7 EMP. RTS. & EMP. POL'Y J. 547 (2003).

152. Clermont & Schwab, *From Bad to Worse*, *supra* note 151, at 117.

153. *Id.* at 127.

154. *Id.* at 129. The movement of the win-rate toward 50% in cases reaching trial comports with Judge Posner's view that one-sided cases are more likely settled, leaving cases on the trial docket more likely to be toss-ups. Posner, *supra* note 98, at 1507.

155. Because these win rate numbers are based on individual discrimination cases, rather than on systemic cases, using them to determine when burdens should shift under *Teamsters* would not have an endogenous feedback effect on the win rates. *Cf.* Sanchirico, *supra* note 70, at 275 (expressing concern that using estimates of probabilities in the determination of burden allocation is susceptible to a "troubling circularity.").

156. *See* Clermont & Schwab, *Employment Discrimination Plaintiffs in Federal Court*, *supra* note 151, at 127 ("But perhaps [plaintiffs' lack of success] results from hurdles being placed before employment discrimination plaintiffs."); Selmi, *supra* note 147 (arguing that various biases can affect courts considering discrimination claims, thus skewing success rates downward); *see also* Calloway, *supra* note 30 (arguing that courts have, mistakenly, become skeptical about the prevalence of racial discrimination in the workplace).

claims.[157] If that were true, then the win rates might be an accurate reflection of the true (prob(X)), or might even overstate the true (prob(X)).

Further empirical study could lead to a better understanding of what these win rates reflect, and how courts should approach the problem of estimating the prior probability that X occurred. In particular, empirical studies that could effectively gauge the relative prevalence of the various types of discrimination (age, sex, race, disability, etc.) in the workplace might provide a more useful starting point for estimating (prob(X)), and a better basis for determining the proper extent of the systemic disparate treatment theory. If courts or legislators have some reason to believe that (prob(X)) is significantly lower for certain types of discrimination (for example, age discrimination or disability discrimination) than it is for other types (for example, the type of race discrimination at issue in *Teamsters*), then that difference should be taken into account in determining whether the systemic disparate treatment method of proof is appropriate for those types of cases.

Despite the difficulties in estimating prior probabilities, courts considering systemic disparate treatment claims cannot simply ignore the prior probability that X (unlawful discrimination) occurred when making probability estimates and burden allocation decisions.[158] They should use what information is available to them and attempt to gauge a prior probability before updating that probability with the evidentiary signals they observe. In particular, the statistical evidence typically offered in Phase I of a systemic disparate treatment case cannot be viewed in a vacuum, without reference to an estimate of the prior probability of discrimination.

B. The Reference Class Problem

A particular problem in estimating prior probabilities is highlighted by the varying win rates presented in subpart A and is

157. This might be because plaintiffs suffering an adverse employment action, such as job termination, are more likely to be desperate for money, and more likely to have already severed their business association with the defendant (often the plaintiff's former employer).

158. *See* Paetzold, *supra* note 149, at 412–14 (urging a complete statistical revolution in antidiscrimination law, from the traditional frequentist statistical analysis to a Bayesian statistical approach). My proposal can be thought of as a hybrid approach, whereby the results of a traditional frequentist statistical analysis, such as a binomial distribution or a multiple regression analysis, can represent a piece of evidence or an input (like a fingerprint or a DNA match analysis) that can, in turn, be thought about in a Bayesian way, as advocated by scholars like Richard Friedman. *See* Friedman, *supra* note 138.

worthy of special attention: the reference class problem. Assume that the court is considering a pattern or practice claim of sexual discrimination against females, where the defendant employer is a trucking company. To what class of cases should the court refer when estimating the prior probability that unlawful discrimination occurred?

Should the court, when estimating the prior probability, look to: (a) the probability of unlawful discrimination of any kind in the employment environment generally; (b) the probability of unlawful discrimination on the basis of sex in the employment environment generally; (c) the probability of unlawful discrimination on the basis of sex in the trucking industry; (d) the probability of unlawful discrimination on the basis of sex in the trucking industry in cases that reach litigation; or (e) the probability of unlawful discrimination on the basis of sex in the trucking industry in cases reaching trial; or any number of other plausible reference classes? Any one of these reference classes might be appropriate, and the choice of reference class could have a significant effect on the court's estimate of the prior probability.[159] For example, the overall discrimination rate against females in the workplace in the U.S. might be .001, while the discrimination rate against females in industries traditionally dominated by males may be much higher, say .05. Moreover, the discrimination rate against females in cases that get to litigation might be .2, while the rate in cases that make it to trial may be .5. The choice of reference class will have a significant effect on (prob(X)), and therefore on the court's ultimate estimate of probabilities.[160]

159. For further discussion of the reference class problem, see Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 J. LEGAL STUD. 107, 114–16 (2007). For a broader discussion about the propriety of using formal probability theory to explain civil trial decision-making, see Ronald J. Allen, *A Reconceptualization of Civil Trials*, 66 B.U. L. REV. 401 (1986) (arguing that civil trials would better comport with formal probability theories if reconceptualized as contests between two competing, but fully specified, versions of the relevant facts); Ronald J. Allen & Brian Lieter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491, 1527–28 (2001) (arguing that relative plausibility theory better explains the legal fact-finding process than probability theory), and Friedman, *supra* note 138 (responding to Allen and Lieter).

160. This is the dilemma identified by Cleary, who posited that probabilities in litigated cases make a more appropriate baseline than probabilities of a situation in general. Cleary, *supra* note 100, at 13. Note that the reference class problem affects traditional frequentist statistical analysis as well. Thus, in *Hazelwood*, a significant question was determining the proper comparison population for conducting a binomial distribution analysis. Meier et al., *supra* note 30, at 146.

No one has yet devised a convincing solution to the reference class problem that works across all contexts.[161] Some contend that the reference class problem renders the value of any Bayesian mathematical models of evidence law "limited."[162] In some cases, the reference class problem appears intractable; there is no obviously "best" choice among several possible reference classes. However, in other cases the problem may be manageable. For example, in deciding whether to use the probability of discrimination in society in general or the probability of discrimination in employment cases reaching litigation, it appears that choosing the former reference class is really just electing to ignore probative and available information about the case for no good reason.[163] Generally speaking, the more specific the information used to define the reference class, the better its use as a basis for estimating probabilities[164]—at least for purposes of considering the application of the systemic disparate treatment doctrine.[165]

While it may not always be possible to locate the "best" reference class from the universe of all possible reference classes, in the context of our adversarial legal system choosing the "better" reference

161. For a proposed practical solution to the reference class problem for purposes of selecting a reference class for use in our adversarial legal system, see Edward K. Cheng, *A Practical Solution to the Reference Class Problem*, 109 COLUM. L. REV. 2081 (2009).

162. Allen & Pardo, *supra* note 159, at 107; Cheng, *supra* note 161, at 2083–84.

163. This may explain why Cleary believed that reference to probabilities in litigated cases was a superior baseline to probabilities of a situation in general. *See* Cleary, *supra* note 100. Using cases reaching trial (rather than just reaching litigation) may be inappropriate, however, because pattern or practice cases arguably do not fall within that reference class. At the time the burden-shifting decision is made in a pattern or practice case, the presentation of Phase I evidence has been completed, but the presentation of evidence on the individual claims (Phase II) has not yet commenced. There remains an opportunity for settlement of claims before the individual trials. A more conservative estimate of (prob(X)) based on the reference class of cases reaching litigation therefore seems more appropriate.

164. There is, however, a risk of overfitting, which refers to the use of highly detailed and specific information to arrive at a very small reference class that is defined by a number of very unique characteristics, and which is therefore unlikely to be a good representation of the true prior probability. *See* Cheng, *supra* note 161, at 2093 (likening the problem of overfitting in reference class selection to the problem of overfitting in statistical model selection). At least for the present time, that risk appears to be minimal in the context of employment discrimination, given the lack of available data. There is little data available about the frequency of unlawful discrimination in the workplace in general, let alone highly stratified data analyzing discrimination frequency in specific industries or specific geographic locations. *See id.* at 2097–98 (noting that the availability of datasets provides a practical limitation on the universe of potential reference classes that legal adversaries can "dream up").

165. Unlike some of the mathematical models critiqued by Allen and Pardo, the analysis herein is not aimed at assigning a specific, objective, and "correct" mathematical value to any individual piece of evidence. *See* Allen & Pardo, *supra* note 159, at 130 (acknowledging that Bayesian models can be used for "less problematic purposes" when arguments do not depend on assertions of the "true or correct probative value of evidence.").

class may be possible by viewing the choice of reference class as an exercise in model selection. Each side can present to the court its chosen reference class and advocate for it by explaining how it better "fits" the question at issue. The court can then choose which reference class is "better," and come up with its rough estimate of the prior probability (prob(X)) accordingly.[166]

C. Statistical Signaling in Systemic Disparate Treatment Cases

> *"Statistics per se cannot ascertain whether employment discrimination exists . . . . But statistical testing is, we would argue, a useful bench mark."*[167]

The second key factor to examine is the evidentiary signal: (prob($Y|X$)) as compared to (prob($Y|{\sim}X$)). For our purposes, the evidentiary signal (Y) is the evidence observed by the court at Phase I of a pattern or practice case. In the typical pattern or practice case, the Phase I evidence is primarily statistical evidence.[168] Without question, the majority of the controversy surrounding systemic disparate treatment cases center on how to properly consider, interpret, and weigh statistical evidence suggesting discrimination. Although the use of statistics in systemic discrimination cases is now commonplace, the propriety of shifting burdens of proof onto defendants based primarily upon statistical analyses is still disputed.

Statistical evidence can provide a "signal" (Y) for updating prior probabilities in order to permit courts to make estimates of posterior probabilities, and in turn, determine whether a shift of the burden of persuasion is appropriate. Statistical evidence comes in a variety of forms, but this subpart focuses on two types of statistical

166. *See* Cheng, *supra* note 161, at 2096–97.

167. Meier et al., *supra* note 30, at 162–63 (emphasis omitted).

168. Non-statistical evidence in pattern or practice cases may also be evaluated using this framework; for example, if there is undisputed evidence that an employer has an express, announced policy of discrimination (e.g., "Females may not be hired"), then that evidence can be plugged in as the evidentiary signal (Y). In the example given, (prob($Y|X$)) = 1, or at least some figure approaching 1, for any female applicant, and (prob($Y|{\sim}X$) = 0. *But see* Bates v. United Parcel Serv., Inc., 511 F.3d 974, 988 (9th Cir. 2007) (*Teamsters* burden-shifting framework is unnecessary where the case involves an employer's facially discriminatory policy, because the fact to be uncovered by the burden-shifting scheme has already been admitted). Disputed evidence of an unspoken or unwritten employer policy of discrimination might be evaluated in a similar manner, although that would require factoring in credibility determinations by the factfinder. *See generally* EEOC v. Joe's Stone Crab, Inc., 220 F.3d 1263, 1285 (testimony of multiple statements made by management could be evidence of a facially discriminatory policy, even though there was no formally announced facially discriminatory policy).

evidence most commonly offered in systemic employment discrimination cases: binomial distributions[169] and regression analyses.[170]

At the outset, it is important to recognize that these two frequentist statistical techniques do not provide exactly the information that courts need to update their prior probabilities. The two values of most significance are (prob($Y|X$)), the probability of observing the evidentiary signal when assuming unlawful discrimination, and (prob($Y|{\sim}X$)), the probability of observing the evidentiary signal when assuming no discrimination (i.e., a false positive in the statistical evidence).[171] These represent the information the court needs to update its priors and make probability estimates.

By contrast, binomial distributions and multiple regressions provide only the likelihood that the observed distribution or regression coefficient would be observed, assuming random selection. In other words, binomial distribution and regression analyses can provide information only about (prob($Y|$random selection)). This is not precisely the same as (prob($Y|{\sim}X$)), because we would not expect that nondiscriminatory employment decisions would necessarily be the same as *random* employment decisions.[172] Nonetheless, even those scholars with reservations about overreliance on statistical evidence acknowledge that models of random selection can still provide a useful *benchmark* for comparison purposes.[173] For practical reasons, it is reasonable to assume that the two are at least closely related. If random decisions could have easily led to the observed distribution or regression coefficient (i.e., the

169. *See* Meier et al., *supra* note 30; Paul Meier, *Damned Liars and Expert Witnesses*, 81 J. AM. STAT. ASS'N 269, 271 (1986); *see also* Browne, *supra* note 144, 479–82 (criticizing the use of binomial distributions as evidence of discrimination).

170. *See* Delores A. Conway & Harry V. Roberts, *Regression Analyses in Employment Discrimination Cases*, *in* STATISTICS AND THE LAW 107 (Morris H. DeGroot et al. ed., 1986). *Compare* D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533 (2008) (criticizing the continued use of regression analyses in employment discrimination cases and urging utilization of a more recently-developed statistical technique, known as potential outcomes, which manipulates observational data in a way that is designed to approximate a controlled experiment, from which inferences of causation might be drawn); *with* Steven L. Willborn & Ramona L. Paetzold, *Statistics is a Plural Word*, 122 HARV. L. REV. F. 48 (2009) (responding to Griener by arguing that regression and potential outcomes analyses are both useful techniques, yet both flawed in different ways).

171. *See supra* Part II.B.1.b (discussing the derivation of the signal comparison (Y/X to $Y/{\sim}X$) in the Hay & Spier model).

172. *See* Meier et al., *supra* note 30, at 153.

173. *Id.* at 25 ("Although hiring is not a process of random selection, it is not unreasonable to compare it to such a benchmark process. If the disparity between black and white hire rates is so small that it could have arisen by random sampling from an appropriate population, then it seems reasonable to conclude that the statistics, by themselves, provide no evidence of discrimination.").

(prob(Y|random selection)) is very high), then it can be assumed an employer's nondiscriminatory decisions could likewise have easily led to the observed distribution or coefficient (i.e., (prob(Y|~X)) is also high).

The most commonly accepted test for statistical significance for both binomial distribution and regression analysis is .05 (or, significance at the 5% level).[174] If the (prob(Y|random selection)) < .05, the results are usually said to be "statistically significant." Using random selection as a benchmark for nondiscrimination, it may be reasonable to assume that, in such a case, (prob(Y|~X)) is also reasonably likely to be less than .05.

The other variable of interest from the Hay and Spier model is (prob(Y|X)). That is, the probability of observing (Y) given an assumption that the employer did unlawfully discriminate (X). In what percentage of cases would we expect to see the observed statistical evidence of disparity if the employer were actually discriminating? In what percentage of cases would we expect false negatives? The relationship between (prob(Y|~X)) and (prob(Y|X)) is complicated, but one statistician has estimated that where statistical significance is set to the 5% level, the probability of observing a false negative is approximately .5.[175] Thus, (prob(Y|X)) is approximately .5. In the case of observed statistical evidence showing statistical significance at the 5% level, it is therefore reasonable to assume that (prob(Y|X)) is substantially higher than (prob(Y|~X)), which is likely to be less than .05.[176]

---

174. *See* Posner, *supra* note 98, at 1510–11. As Judge Posner notes, "The five percent test is a convention employed in academic research, though not one to which the research community adheres rigidly. Social scientists often report results that are significant only at the ten percent level. And if the results are significant at the two percent or one percent level, the social scientist will point out that these results are more robust than those that are significant only at the five percent level; they are 'highly significant' rather than just 'significant.' There is no magic to the five percent criterion . . . ." *Id.* at 1511.

175. John M. Dawson, *Scientific Investigation of Fact—The Role of the Statistician*, 11 FORUM 896, 907–08 (1976); *see also* SULLIVAN ET AL., *supra* note 1, at 178 ("Although the risks of Type I and Type II errors are inversely related in that increasing one decreases the other, they are not simple complements . . . . One statistician has created a model for an employment discrimination case by setting the risk of Type I error at 5 percent. With the confidence level at 95 percent, he estimated the risk of Type II error to be approximately 50 percent." (citing Dawson, *supra*)).

176. Hay & Spier suggest just such a result in the context of an ordinary *McDonnell Douglas* framework for individual discrimination cases. They note that, "[r]oughly speaking, if the plaintiffs can show they were treated differently from a group of similarly situated male employees [by establishing a prima facie case under *McDonnell Douglas*], the defendant must come forward with evidence that the plaintiff was not fired for discriminatory reasons. In such a setting, prob(Y|X) is relatively high compared to prob (Y|~X) . . . ." Hay & Spier, *supra* note 90, at 427 (footnote omitted).

D. Information Asymmetries in Disparate Treatment Cases

The third and final key factor in burden allocation is an assessment of relative access to information on the fact in question. This is represented as the cost to plaintiff of showing that *X* occurred, as compared to the cost to defendant of showing that *X* did not occur. As noted above, procedural discovery rules can complicate this question, and their effect should be taken into account to the extent possible.

Employment discrimination cases vary, but in the typical disparate treatment discrimination case plaintiff alleges that the employer intentionally took some adverse employment action against plaintiff *because of* plaintiff's membership in a protected class. In other words, the employer made a decision to not hire, not promote, terminate, or pay less compensation to, the plaintiff because of her sex, race, religion, age, or other protected characteristic. The critical factual question in most disparate treatment cases is not whether plaintiff suffered an adverse employment action, but rather whether the employer acted with a discriminatory intent.[177] It is on this factual question that the shifting of the burden of production and/or persuasion can be critical.

It will often be the case, in the disparate treatment context, that the defendant will have better access to the evidence on the relevant point. Usually the information about why employment decisions are made is exclusively in the possession of the employer and its agents.[178] The relevant decision-maker will likely know the true reason for the adverse action, and that information may be reflected in the employer's documents, such as personnel files or possibly notes kept by the decision-maker. Generally speaking, employers will have better access to evidence than plaintiffs will in the disparate treatment context.[179] Although liberal discovery may lessen this information asymmetry somewhat, it seems unlikely that it

177. This has been described as a "causation" requirement for disparate treatment cases. *See* Katz, *supra* note 45, at 860 ("Almost all disparate treatment statutes include an element of causation. They do not prohibit adverse employment actions, such as firing, in all instances. Rather, they prohibit adverse employment actions only where those actions occur 'because of' a protected characteristic, such as race, sex, or age. In other words, these statutes all require causation.") (footnote omitted); Martin J. Katz, *The Fundamental Incoherence of Title VII: Making Sense of Causation in Disparate Treatment Law*, 94 GEO. L.J. 489 (2006).

178. *See* Katz, *supra* note 45, at 881 ("Causation occurs in the mind of the decision-maker/defendant. And most of the relevant evidence tends to be under the control of the defendant. This lack of access to evidence makes proving any type of causation difficult, and therefore makes burden-shifting normatively desirable.").

179. *See id.* at 881; Hay & Spier, *supra* note 90, at 427 ("[I]n addition, the employer has lower costs of producing evidence on the motive for the discharge.").

would completely equalize the parties' relative access to information.[180]

Interestingly, courts considering the application of the *Teamsters* burden-shifting doctrine often fail to consider the parties' relative access to information and instead focus only on the quality of, or flaws in, the statistical or other evidence offered by plaintiffs to show a pattern or practice of discrimination.[181] These courts are overlooking a key component of the inquiry. Because the question is whether to shift the burdens of production and persuasion onto the defendant, courts should consider not only how the statistical evidence affects the probabilities, but also the asymmetries in access to information inherent in the element of discriminatory intent.

E. Illustration: A Hypothetical Application to a Single Systemic Disparate Treatment Case

The foregoing framework can be used by courts to determine whether burden-shifting is appropriate in any given systemic disparate treatment case before the court. Consider the following hypothetical case as an illustration. It is crucial to note that this hypothetical application is an illustration only. The chief purpose is to urge courts and policymakers to think about systemic disparate treatment cases in these terms, and to substitute the proper figures according to their beliefs and the available empirical information, as well as the evidence submitted to them in a particular case.

In a systemic discrimination case, a class of female employees alleges that the defendant employer intentionally and systematically discriminated against women in making promotions to management positions. The court estimates that the prior probability of unlawful discrimination (prob(X)) in this case is roughly .15, given the historical data on success rates for Title VII discrimination plaintiffs in cases reaching litigation.[182]

180. *See supra* notes 134–136 and accompanying text.

181. *See, e.g.*, Morgan v. United Parcel Serv. of Am., Inc., 380 F.3d 459 (8th Cir. 2004) (affirming summary judgment in favor of employer because of weaknesses in statistical evidence and lack of testimonial evidence regarding intentional discrimination, without considering the possibility of asymmetries in access to information); EEOC v. McDonnell Douglas Corp., 191 F.3d 948 (8th Cir. 1999) (affirming summary judgment in favor of employer on pattern or practice disparate treatment age discrimination claim in the context of a RIF due to weaknesses in statistical and testimonial evidence, without mentioning potential disparities in access to evidence on motivation for terminations).

182. If the court believes that these success rates are artificially skewed toward defendants because of unwarranted procedural or other obstacles facing employment discrimination plaintiffs, it may place the estimate of (prob(X)) closer to .5, making a

In Phase I, both parties offer statistical evidence presented by competing statistical experts. The court finds the statistical evidence presented by plaintiffs to be more convincing because the statistical model is better-designed.[183] Plaintiffs' statistical evidence consists of a multiple regression analysis, using what the court determines to be an appropriate set of independent variables. The results show that the gender of an applicant for promotion was a very strong predictor of whether the applicant received a promotion, and the regression coefficient for gender is highly statistically significant at the 5% level, meaning that it was not likely produced by chance. Based on this statistical result, the court might reasonably assume that (prob($Y|{\sim}X$)) is probably .05 or less.[184] The court likewise reasons that (prob($Y|X$)) must be substantially higher than (prob($Y|{\sim}X$), and therefore estimates (prob($Y|X$)) as .5.[185] Plugging these estimated values into the Hay and Spier formula, the court should shift the burden to defendant if, and only if:

.5 x .15 x plaintiff's information costs > .05 x .85 x defendant's information costs

which reduces to:

.075 x plaintiff's information costs > .0425 x defendant's information costs

and further reduces to:

1.76 x plaintiff's information costs > defendant's information costs

---

stronger case for burden-shifting. On the other hand, if the court believes that the .15 win rate overstates the percentage of Title VII claims brought to litigation that are actually meritorious, then it may arrive at a lower estimate of (prob(X)), weakening the case for burden-shifting.

183. For the sake of simplicity, this illustration assumes a bench trial, where the court is the factfinder. Whether a jury should be permitted to evaluate complex statistical evidence of discrimination in Phase I of a pattern or practice case poses another question beyond the scope of this Article. Studies suggest that judges and juries are equally poor at evaluating complicated statistical evidence. *See* Andrew Jurs, *Judicial Analysis of Complex & Cutting-Edge Science in the* Daubert *Era: Epidemiologic Risk Assessment as a Test Case for Reform Strategies*, 42 CONN. L. REV. 49, 73–75 (2009); Jennifer K. Robbennolt, *Evaluating Juries by Comparison to Judges: A Benchmark for Judging?*, 32 FLA. ST. U. L. REV. 469, 491–92 (2005).

184. On this point, the court should be careful to recognize that it is actually using (prob($Y|$*random selection*)) as a benchmark for estimating (prob($Y|{\sim}X$)).

185. The court is assuming that even where the employer actually does unlawfully discriminate against females in handing out promotions a statistical analysis would reveal a highly statistically significant result at the 5% level only about half of the time, and would return a false negative about half of the time. This assumption is a reasonable one, according to Dawson's calculations. *See supra* note 175 and accompanying text.

In other words, given the court's estimates of probabilities, updated in light of the statistical evidence presented at Phase I, the court should leave the burdens of production and persuasion on plaintiff only if it believes that the plaintiff has 1.76 times *better* access to information (i.e., lower information costs) on the question of discriminatory intent than the defendant. If the court believes that the defendant actually has better access to information on the element of discriminatory intent—the reasons for not promoting any particular female applicant—then the court should shift the burdens of production and persuasion onto defendant on that point.

IV. Application to Groups of Cases: Resolving Uncertainties in Systemic Disparate Treatment Law

The usefulness of this theoretical framework is not necessarily limited to resolving burden-shifting decisions in individual cases. Policy makers and appellate courts can use the principles derived herein to generalize about *types* of discrimination cases, in order to make principled decisions about the proper reach of the *Teamsters* systemic disparate treatment theory. To demonstrate, reconsider the questions in pattern or practice doctrine outlined in Part II.A., using the insights offered by the framework developed in this Article.[186]

A. Age Discrimination

Should the *Teamsters* framework be available to plaintiffs in age discrimination cases under the ADEA? Does it make sense as a policy matter to treat systemic age discrimination claims differently than Title VII systemic race, sex, national origin, or religion claims?

The theoretical framework developed above can guide this policy analysis. Is there any reason to think that the prior probability of discrimination (prob(X)) is lower for age claims than it is for other kinds of discrimination claims? Congress may have thought so.

186. In this Part, I consider the differences between cases like *Teamsters* and cases involving age discrimination, disability discrimination, systemic harassment, or individual nonclass plaintiffs, and analyze whether those differences are relevant under the framework developed in this Article. The discussion to follow assumes that legislators and courts (and the reader) believe that *Teamsters* was correctly decided, and that shifting the burdens of persuasion and production can be justified in appropriate systemic disparate treatment cases.

Before enacting the Civil Rights Act of 1964, Congress considered and rejected proposals to include older workers as a protected group under Title VII.[187] Instead, Congress requested that the Secretary of Labor study the issue of age discrimination in employment and provide a report.[188] The Secretary's report indicated that there was little intentional discrimination against older workers arising from dislike or intolerance of older people.[189] Rather, the report found that some "arbitrary" age limitations could result in age discrimination.[190] The Supreme Court has noted that, as a historical matter, "intentional discrimination on the basis of age has not occurred at the same levels as discrimination against those protected by Title VII."[191]

However, this may not necessarily equate to a significant difference in the prior probability (prob(X)) of unlawful discrimination *in cases reaching litigation*. As discussed above, cases falling at either extreme are likely to settle.[192] A comparison between win rates for ADEA plaintiffs in litigated cases versus win rates for litigated Title VII cases may provide additional insight. In their most recent study Clermont and Schwab found that ADEA plaintiffs actually had *higher* win rates (11.67%) than Title VII plaintiffs (10.88%) for the years 1998–2006.[193] This suggests that, at least for cases reaching litigation, the prior probability (prob(X)) for age discrimination is likely not significantly different from (prob(X)) for the types of discrimination prohibited by Title VII. Further empirical research could shed more light on the relative prevalence of unlawful intentional age discrimination in the workplace, as compared to other types of discrimination.

Next, consider the type and quality of the evidentiary signal (the statistical evidence) available in systemic age discrimination cases. The types of statistical evidence offered in systemic age discrimination cases are generally the same as the types of statistical evidence offered in race, sex, national origin, and religion cases—usually consisting of either a binomial distribution analysis or a multiple regression analysis, with the same inherent strengths and weaknesses that attend those statistical techniques. These statistical

187. 110 CONG. REC. 2596–99, 9911–13 (1964); Smith v. City of Jackson, 544 U.S. 228, 232 n.2 (2005).

188. *Smith*, 544 U.S. at 232 (citing Civil Rights Act of 1964, § 715, 78 Stat. 265).

189. *Id.* (citing Report of the Secretary of Labor, The Older American Worker: Age Discrimination in Employment 5 (June 1965), *reprinted in* U.S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION, LEGISLATIVE HISTORY OF THE ADEA (1981), Doc. No. 5).

190. *Id.*

191. *Id.* at 241; *see also* Hazen Paper Co. v. Biggins, 507 U.S. 604, 610–11 (1993).

192. *See supra* note 149 and accompanying text.

193. *See* Clermont & Schwab, *From Bad to Worse*, *supra* note 151, at 117.

analyses can provide the same type of signal suggesting the existence of the discriminatory intent element. There is no reason to think that (prob($Y|X$)) or (prob($Y|{\sim}X$)) will be systematically different in age cases as compared with race or sex cases.[194]

Finally, there is no apparent reason to believe that age discrimination cases are significantly different from other types of discrimination cases on the parties' relative access to information about discriminatory intent. It may be no easier for the victim of age discrimination to learn the true reason for his termination than it is for the victim of sex or race discrimination.[195] As in Title VII cases, the employer in an ADEA will have greater access to evidence on the question of discriminatory intent than the plaintiff.

Applying the theoretical framework developed above, there is no compelling policy reason to treat systemic disparate treatment ADEA cases any differently than systemic disparate treatment Title VII cases. From a policy perspective, courts should allow ADEA plaintiffs to utilize the *Teamsters* burden-shifting method of proving systemic disparate treatment. If necessary, the legislature should amend the ADEA to resolve any textual ambiguity and make it clear that the *Teamsters* framework is available in age cases.

B. Disability Discrimination

Should plaintiffs under the ADA be able to utilize the *Teamsters* framework? Resolving this question is difficult, even with the

194. Age sometimes correlates with factors that may be legitimately considered by employers in making employment decisions. *See generally Hazen Paper Co.*, 507 U.S. at 611. Thus, it could be argued that statistical evidence of age disparities in hiring or terminations might seem to have less probative force than in the context of race or sex. But this argument ignores two points. First, a well-modeled regression analysis can control for any quantifiable legitimate variables that may correlate with age. Second, even if a regression analysis control is impracticable in Phase I, the purpose of Phase II in a systemic disparate treatment case is to permit the employer to demonstrate any legitimate factors that led to the adverse employment action suffered by any individual claimant—a question on which the employer has superior access to information. Thus, burden-shifting on the basis of statistical evidence in age discrimination cases furthers the policy purposes of systemic disparate treatment theory.

195. *See generally* Michael J. Zimmer, *Chaos or Coherence: Individual Disparate Treatment Discrimination and the ADEA*, 51 MERCER L. REV. 693, 713 (2000). There might, however, be some reason to think that employers or their agents are somewhat less likely to hide their true beliefs about older employees, and therefore more likely to create direct evidence of discriminatory intent in age cases than in race or sex cases. *See supra* notes 187–191. If true, this would mean that the parties' relative access to information on discriminatory intent is marginally more symmetrical in age cases.

benefit of the theoretical framework developed herein.[196] Nonetheless, the framework can help sharpen analysis of the question.

First, consider prior probabilities. Like age discrimination, there may be some reason to believe that animus-based intentional discrimination against disabled individuals in the workplace is generally less prevalent than animus-based intentional discrimination based on race, sex, or national origin.[197] Narrowing the focus to cases that reach litigation, however, there is less reason to believe that (prob(X)) is significantly different for disability cases than for Title VII cases. Clermont and Schwab found win rates of 9.12% for ADA plaintiffs versus 10.88% for Title VII cases for the years 1998–2006.

The second factor—the strength of the evidentiary signal—poses the major hurdle to applying systemic disparate treatment theory in ADA cases. Obtaining reliable estimates of (prob($Y|{\sim}X$)) and (prob($Y|X$)) is substantially complicated by the definition of unlawful discrimination under the ADA. Recall that (X) represents not just discrimination, but *unlawful* discrimination—as defined by the legal rule in question. The ADA prohibits "discriminat[ing] against a qualified individual with a disability."[198] As the *Hohider* court correctly explained, the ADA differs from Title VII in that it does not categorically prohibit discrimination against any disabled individuals, but rather only protects "qualified individuals with a disability."[199]

Presenting statistical evidence (Y) that gives some reliable indication of whether discrimination in violation of the ADA (X) has occurred therefore poses a challenge. A binomial distribution or regression analysis reflecting a disparity in an employer's hiring or promotion of all disabled individuals, without any analysis of whether those disabled individuals were "qualified individuals with a disability" as defined by the ADA, does not give any reliable signal (Y) about whether the employer is acting in violation of the statute (X). This is because the law recognizes that, unlike race, sex, or

196. This question will likely need to be resolved in the near future, because the EEOC is currently exploring ways in which it can pursue more ADA cases as part of its Systemic Initiative. *See* EEOC Transcript, *supra* note 12 (Testimony of Commissioner Christine M. Griffin)("And as I sit here thinking about that wonderful report and the remarks about individual ADA cases versus Systemic, I was actually trying to think would that, would the passage of that [ADA Restoration Act] create more of a Systemic opportunity within the ADA versus what's traditionally been mostly individual cases.").

197. *See* Alexander v. Choate, 469 U.S. 287, 295 (1985) (noting that discrimination against the disabled is "most often the product, not of invidious animus, but rather thoughtlessness and indifference . . . .").

198. 42 U.S.C. § 12112(a) (2006).

199. *See* Hohider v. United Parcel Serv., 574 F.3d 169, 191 (3d Cir. 2009).

national origin, in many cases a person's disability actually *is* relevant to job decisions; the disability itself might be a legitimate reason for taking an adverse employment action.

Few cases have considered the possible application of the *Teamsters* doctrine in the ADA context.[200] Indeed, in *Hohider* the court noted that no federal court of appeals had directly addressed whether or how the *Teamsters* framework might apply in ADA cases.[201] However, it might be possible to envision an observational statistical study of an employer's workforce that could provide some reliable indication of (prob($Y|{\sim}X$)) in a disability discrimination case. Perhaps a statistical study could be undertaken after individual rulings are made (or agreements are reached) on which employees or applicants included in the study meet the statutory requirement as "qualified individuals with a disability." In this hypothetical study, the independent variable in question would not be simply "Disabled," but would be a more appropriate independent variable that comports with the statutory requirement: "Qualified Individual with a Disability." However, no such studies have yet been offered as statistical proof of systemic disability discrimination in the reported cases to date.[202]

Finally, consider the parties' relative access to information on the element of discriminatory intent. Again, evidence of the employer's reasons for taking an adverse employment action is generally in the hands of the employer. There is no apparent reason to believe that information asymmetries on the element of discriminatory intent are systematically different for disability cases than for other types of discrimination.

The theoretical framework developed in this Article can assist courts and policymakers in considering whether ADA plaintiffs should be allowed to use the *Teamsters* systemic disparate treatment method of proof. But the reported cases on systemic

200. In *Hohider*, the plaintiffs did not offer statistical proof of a pattern or practice of discrimination against the disabled. Instead, they argued that the defendant maintained certain unwritten policies that resulted in systemic discrimination against the disabled. *See id.* at 172 (plaintiffs alleged that UPS enforced an "unwritten policy, which prohibits employees from returning to UPS in any vacant position unless the employee can return to his or her last position without any medical restrictions."). See *supra* note 168 for a discussion of how testimonial evidence regarding unwritten employer policies of discrimination might be evaluated as an evidentiary signal under the framework developed herein.

201. *Hohider*, 574 F.3d at 179 n.11.

202. Alternatively, a study might examine statistical disparities only with respect to the termination of individuals who had been employed by defendant while disabled. The independent variable in such a study would therefore naturally be confined to only those disabled individuals whom the employer had already determined were qualified, with or without reasonable accommodation, to perform the essential functions of their position. *See* 42 U.S.C. § 12111(8).

disability discrimination highlight the difficult challenge of obtaining reliable evidentiary signals (prob($Y|{\sim}X$)) for making accurate judicial estimates of the probability of unlawful behavior sufficient to justify burden-shifting. However, a more finely-tuned statistical model that could closely track the "qualified individual with a disability" requirement of the ADA might offer a more reliable evidenevidentiary signal at Phase I, justifying burden-shifting in systemic disability cases. Therefore, appellate courts and legislators should not foreclose the potential application of the *Teamsters* method of proof in appropriate ADA cases.

C. Hostile Work Environment

Should *Teamsters* be used to shift the burdens of production and persuasion in systemic harassment cases? Applying the framework developed herein, the answer should be no. In systemic harassment cases, there are two major differences from ordinary Title VII systemic discrimination cases: (1) a general inability to obtain reliable estimates of (prob($Y|{\sim}X$)) and (prob($Y|X$)) with statistical evidence and (2) the lack of information asymmetries on the relevant factual points at issue.

First, in systemic harassment cases the statistical evidence available (if any) gives no reliable basis for estimating (prob($Y|{\sim}X$)) or (prob($Y|X$)). The court has no reliable information with which to update its estimate of the prior probability of unlawful harassment (prob(X)). The only statistical evidence available in a systemic harassment case consists of the number of harassment complaints and the distribution of those complaints, as well as the size and distribution of the workforce.[203] But the number and distribution of harassment complaints alone does not give a reliable indication that unlawful harassment in violation of Title VII has occurred. A complaint is just that—a charging party's expression that it believes a violation has occurred. It does not provide the same "telltale sign" of employer discrimination that a statistically significant disparity in the distribution of hirings, firings, or promotions can provide.[204]

---

203. *See, e.g.*, EEOC v. Carrols Corp., No. 5:98 CV 1772, 2005 WL 928634, at *2 n.4 (N.D.N.Y. April 20, 2005); *see also* Bent, *supra* note 10 at 176–77.

204. Some might contend that plaintiffs in a systemic harassment case can present nonstatistical evidence of either an express or unwritten policy of tolerating harassment in the workplace, thereby providing the court with at least some information from which a rough estimate of (prob($Y|{\sim}X$)) might be made. *See supra* note 168. Even if this is taken to be true, other characteristics unique to hostile work environment law, discussed in the text immediately below, make estimation of (prob($Y|{\sim}X$)) especially challenging.

Further, like the "qualified individual with a disability" requirement of the ADA, the definition of an unlawful hostile work environment in violation of Title VII contains intricacies not captured by the available statistical evidence. A plaintiff may indeed have been harassed by a co-worker or supervisor, and might even have complained about it, without an employer violating Title VII.[205] Usually plaintiffs in systemic harassment cases offer only a series of anecdotal accounts of harassment and testimony about an employer's inadequate responses to individual instances of harassment. But these anecdotal accounts, even if taken as true, tell the court very little about whether any other class member suffered from harassment rising to the level of an objectively and subjectively hostile work environment. As one district court put it: "[A] findfinding that an employer had a pattern or practice of tolerating sexual harassment in violation of Title VII does not necessarily establish than an individual claimant was exposed to harassment or that the harassment an individual claimant suffered violates Title VII."[206] In terms of the Hay and Spier model, where (Y) consists only of anecdotal testimony offered by other class members in Phase I, then as to any other non-testifying class member there is little reason to be confident that ($\text{prob}(Y|X)$) is greater than ($\text{prob}(Y|{\sim}X)$). Or, restated, the observation that several class members testified in Phase I that they were exposed to an objectively and subjectively hostile work environment does not tell the court much about whether a different class member, who did not testify in Phase I, was also exposed to a work environment that was both objectively and subjectively hostile and abusive (i.e., whether the employer acted unlawfully as to this particular individual). The evidentiary signal from Phase I is weak.

Second, and perhaps more importantly, systemic harassment cases do not present the same sort of asymmetrical access to evidence on

205. First, a plaintiff in a hostile work environment case must establish that the work environment to which he was subjected was both objectively and subjective hostile. *See* Harris v. Forklift Sys., Inc., 510 U.S. 17, 21 (1993). Second, even if the plaintiff can show that he was subjected to an objectively and subjectively hostile work environment due to harassment by a supervisor, the employer has not violated Title VII if (1) the employer took reasonable preventative and corrective measures and (2) the plaintiff unreasonably failed to take advantage of the employer's preventative and corrective measures. *See* Burlington Indus., Inc. v. Ellerth, 524 U.S. 742, 765 (1998); Faragher v. City of Boca Raton, 524 U.S. 775, 807 (1998). For alleged harassment by a co-worker rather than a supervisor, the standard triggering employer liability is even higher. The employer is liable for such conduct only if it was negligent in failing to prevent the harassment. *See, e.g.*, Noviello v. City of Boston, 398 F.3d 76, 95 (1st Cir. 2005); Ocheltree v. Scollon Prods., Inc., 335 F.3d 325, 333–34 (4th Cir. 2003); Hall v. Bodine Elec. Co., 276 F.3d 345, 356 (7th Cir. 2002). *See generally* Bent, *supra* note 10.

206. EEOC v. Int'l Profit Assocs., No. 01 C 4427, 2007 WL 3120069, at *3 (N.D. Ill. Oct. 23, 2007).

the disputed questions as do ordinary Title VII discrimination cases. In the ordinary Title VII disparate treatment discrimination case the key disputed factual question is discriminatory intent—that is, whether the employer took adverse action against the plaintiff *because of* the plaintiff's sex, race, national origin, or religion. The question of discriminatory intent is the question that presents information asymmetries, and it is the fact upon which the *Teamsters* burden-shift operates. The question of whether an individual plaintiff suffered an adverse action is usually not disputed; it is generally easy to determine whether the plaintiff was fired, not hired, not promoted, or transferred, and the burden-shift is not intended to assist in the resolution of that question. The burden-shift operates on the question of the employer's discriminatory intent in taking the adverse action based on (1) statistical evidence suggesting a sufficiently high probability of discriminatory intent and (2) because it is generally difficult for plaintiffs to find and present evidence of that discriminatory intent.[207]

Whether the harassment directed at the plaintiff was because of sex or race or some other protected characteristic is not usually the critical issue in typical harassment cases. Rather, empirical evidence suggests that the decisive issues in harassment cases are more often: (1) whether the work environment to which the plaintiff was subjected rose to the level of subjectively and objectively hostile and (2) whether the plaintiff took advantage of any reasonable preventive and corrective measures the employer had in place.[208] These are both questions on which the plaintiff and defendant have roughly equivalent access to the evidence. Inherently, the plaintiff has personal knowledge of the work environment to which she was subjected. And the plaintiff has personal knowledge of any steps she took, such as making a complaint to her supervisor or to the personnel department, to take advantage of the employer's preventative and corrective measures. The key disputed facts in harassment cases are about what happened to plaintiff (whether she actually suffered an unlawful adverse action), not about what *motivated* the adverse action. The access to information factor does not favor burden-shifting in the context of systemic harassment cases.

207. *See* Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 339 n.20 (1977).

208. See David J. Walsh, *Small Change: An Empirical Analysis of the Effect of Supreme Court Precedents on Federal Appeals Court Decisions in Sexual Harassment Cases*, 30 BERK. J. EMP. & LAB. L. 461, 488, 492–503 (2009). Same sex harassment and "equal-opportunity harasser" cases are the primary exceptions, and more frequently turn on questions of whether the harasser had a discriminatory motive. *Id.* at 488.

The framework developed in this Article shows that the *Teamsters* pattern or practice method of proof is inappropriate for systemic harassment cases. Two factors counsel against burden-shifting in such cases: the weakness of the evidentiary signal offered in Phase I, and the lack of information asymmetries on the key disputed questions of fact.

D. Individual (Non-Class Action) Private Plaintiffs

Should individual employment discrimination plaintiffs have the opportunity to use the *Teamsters* method of proof if they can produce convincing statistical evidence of discrimination, or should they be limited to the weaker *McDonnell Douglas* framework? Applying the foregoing analysis, individual plaintiffs should be permitted to use the *Teamsters* framework in appropriate cases, even though the federal appellate courts have generally held to the contrary.

The only difference between an individual private discrimination claim involving statistical proof of a disparity and a class action discrimination claim involving statistical evidence is that in a class action the plaintiffs have satisfied the procedural requirements for class certification pursuant to Fed. R. Civ. P. 23.[209] In the individual private case, the individual has evidence of a systemic violation but is litigating alone or with too few co-plaintiffs to satisfy the numerosity requirement.

Analyzing this situation using the theoretical framework developed here, there is no reason to treat the individual non-class private plaintiffs any differently than class plaintiffs. There is no principled reason to believe that (prob(X)) is any higher for any given class claimant in a race discrimination class action than it is for an individual race discrimination plaintiff, all else being equal. Before considering the statistical evidence, the (prob(X)) for any given type of discrimination is no different for an individual plaintiff than it is for a plaintiff who also happens to be a member of a certified class. Nor is there any reason to believe that the information asymmetries are any different. It is just as difficult, and perhaps even more difficult, for an individual discrimination plaintiff to find and present evidence of an employer's true reasons for

209. These requirements include: (1) that the class is so numerous as to make joinder of all members impracticable; (2) the existence of questions of law or fact common to the class; (3) that the claims or defenses of the class representatives are typical of the claims or defenses of the class; and (4) that the class representatives will fairly and adequately protect the interests of the class. FED. R. CIV. P. 23(a).

taking an adverse employment action as it is for plaintiffs who are members of a certified class.

Finally, the evidentiary signal, (prob($Y|X$)) as compared to (prob($Y|{\sim}X$)), offered by a single plaintiff at Phase I can be just as strong as the evidentiary signal offered by class plaintiffs. In fact, there is no reason that an individual's statistical evidence need be any different *at all* from what could be offered in a class action. It is possible, for example, that an individual plaintiff in a sex discrimination case could obtain statistical evidence sufficient to support a class action for failure to promote on the basis of sex, but that all of her fellow female employees decline to sue their employer, for reasons personal to them. The Phase I statistical evidence in the individual case would be *identical* to the Phase I evidence that could have been offered if it were a class action.

If an individual plaintiff can make a strong showing that (prob($Y|X$)) is likely much higher than (prob($Y|{\sim}X$)) using statistical evidence, just as in a class action systemic disparate treatment case, then there is no principled reason why that individual should not obtain the same benefit of a shift in the burden of persuasion, rather than being relegated to the less powerful *McDonnell Douglas* scheme. Requiring individual plaintiffs to proceed under the *McDonnell Douglas* framework unjustifiably forces plaintiff to bear the burden of persuasion at all times, despite the strength of the statistical evidentiary signal. As shown by the analysis developed above, there is no policy reason to preclude individual plaintiffs from using the systemic disparate treatment theory.

CONCLUSION

Systemic disparate treatment theory is in disarray. The EEOC's current emphasis on systemic discrimination cases makes it imperative to devise a principled approach to systemic disparate treatment cases. Traditional burden allocation considerations, as well as the *Teamsters* opinion itself, suggest that estimates of probabilities and the parties' relative access to information should play an important role in determining how the systemic disparate treatment theory is applied. The use of Bayesian probability in the Hay and Spier model allows us to think even more rigorously about how burdens should be allocated in systemic disparate treatment cases. Drawing upon that model, this Article identifies the three key factors for determining whether to permit an inference of discriminatory intent under the systemic disparate treatment theory: (1) the prior probability (or estimated prevalence) of discrimina-

tion; (2) the relative strength of the evidentiary signal observed in Phase I of the case; and (3) the asymmetries in the parties' access to evidence on the point at issue.

Courts can make more informed judgments about whether an inference of discriminatory intent is appropriate in any particular systemic discrimination case by analyzing systemic disparate treatment cases through the lens of this theoretical foundation. This foundation will assist courts in considering the weight of statistical evidence and avoiding statistical fallacies by being mindful of the importance of estimates of prior probabilities. The broader effects and applications of this approach will not just serve to remind courts of the importance of information asymmetries and prior probabilities, but should also guide policymakers in defining the appropriate contours of the systemic disparate treatment theory.