

Denver Law Review

Volume 91
Issue 4 *Symposium - Revisiting Sex: Gender and
Sex Discrimination Fifty Years after the Civil
Rights Act Part I*

Article 4

December 2020

Hidden Priors: Toward a Unifying Theory of Systematic Disparate Treatment Law

Jason R. Bent

Follow this and additional works at: <https://digitalcommons.du.edu/dlr>

Recommended Citation

Jason R. Bent, Hidden Priors: Toward a Unifying Theory of Systematic Disparate Treatment Law, 91 *Denv. U. L. Rev.* 807 (2014).

This Article is brought to you for free and open access by Digital Commons @ DU. It has been accepted for inclusion in Denver Law Review by an authorized editor of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu, dig-commons@du.edu.

HIDDEN PRIORS: TOWARD A UNIFYING THEORY OF SYSTEMIC DISPARATE TREATMENT LAW

JASON R. BENT[†]

ABSTRACT

Did the Court's procedural decision in *Wal-Mart Stores, Inc. v. Dukes* undermine the substance of the systemic disparate treatment theory of employment discrimination? The answer to that question hinges on understanding the theoretical foundation for what one scholar calls the "most potent and least understood of the various Title VII causes of action." The current scholarly efforts to understand systemic disparate treatment law can be sorted into two distinct strands—methodological and contextualist. Scholars in the methodological strand question whether statistical techniques currently used by courts are sufficient to support an inference of discrimination. In the contextualist strand, scholars urge a conceptual expansion of the systemic disparate treatment theory that would impose liability on employers for wrongdoing located at the organizational level, rather than simply aggregating individual-level claims. These two strands have advanced independently, with scholars in each strand often overlooking the implications of progression in the other. This Article is the first attempt to unify these two scholarly strands. It does so by exposing the inescapable role of hidden Bayesian priors—preconceptions about background rates of discrimination—in the interpretation of statistical evidence. Taking a Bayesian view, the shortcomings of traditional statistical evidence identified by methodologists are not fatal. Yet, the Bayesian view also provides the conceptual space needed for further development of the organizational approach advanced by contextualists. The *Wal-Mart* decision presents an opportunity to radically rethink this misunderstood area of antidiscrimination law, and this Article takes the first step in developing of a coherent theory of systemic disparate treatment that embraces Bayesian priors.

[†] Associate Professor of Law, Stetson University College of Law; J.D., University of Michigan Law School (2000); B.A., Grinnell College (1997). Earlier versions of this paper were presented at the 2012 Annual Meeting of the Southeastern Association of Law Schools and at the 2012 Annual Colloquium on Current Scholarship in Labor and Employment Law. Thank you to the participants of both conferences for their helpful comments and suggestions. The paper was substantially improved by comments received from Theresa Beiner, David Kaye, Jason Palmer, J.J. Prescott, Gowri Ramachandran, Charles Sullivan, Louis J. Virelli III, Steven Willborn, and Michael J. Zimmer, as well as illuminating conversations on the subject with Shaina Bent, Deborah Weiss, Charlotte Alexander, Scott R. Bauries, Zev J. Eigen, Michael Z. Green, and Michael D. Smith. Lindsay Houser and Michael Labbee provided exemplary research assistance. The faculty and administration of the Stetson University College of Law generously supported this research project.

TABLE OF CONTENTS

INTRODUCTION.....	808
I. THE FOUNDATION OF SYSTEMIC DISPARATE TREATMENT LAW.....	811
II. SYSTEMIC DISPARATE TREATMENT AT A CROSSROADS	816
<i>A. Methodological Strand: The Limits of Statistical Evidence</i>	816
1. Binomial Distribution	818
2. Multiple Regression.....	819
3. The Transposition Fallacy Challenge	820
4. The Potential Outcomes Challenge	824
<i>B. Contextualist Strand: Organizational Causes of Discrimination</i>	828
<i>C. Wal-Mart: The End of Systemic Disparate Treatment?</i>	832
III. THE BAYESIAN PATH FORWARD.....	835
<i>A. Bayesian Inference</i>	835
<i>B. The Bayesian/Frequentist Divide</i>	837
<i>C. Hidden Priors in Judicial Use of Frequentist Statistics</i>	840
<i>D. Hidden Priors in Teamsters and Hazelwood: The Foundational Flaw</i>	842
<i>E. Hidden Priors in Wal-Mart: How Changing Priors Changed Substantive Doctrine</i>	843
<i>F. Reconciling the Methodological and Contextualist Strands by Exposing Hidden Priors</i>	847
IV. A SCHOLARLY AGENDA FOR THE MANAGEMENT OF PRIORS.....	849
CONCLUSION	851

INTRODUCTION

What, if anything, remains of the systemic disparate treatment theory of employment discrimination? This question is raised by the Supreme Court's decision in *Wal-Mart Stores, Inc. v. Dukes*.¹ Although *Wal-Mart* was ostensibly a procedural case about the commonality requirement for class certification, the opinion will have important, if deeply uncertain, implications for the substantive law of systemic disparate treatment.² Several scholars have questioned whether, as a practical matter, the theory survives *Wal-Mart* at all.³ One thing is certain: systemic disparate

1. 131 S. Ct. 2541 (2011).

2. See Tristin K. Green, *The Future of Systemic Disparate Treatment Law*, 32 BERKELEY J. EMP. & LAB. L. 395, 397 (2011) ("But the Court did more than pull the procedural rug out from under the decade-long lawsuit; it called into question the future of systemic disparate treatment law."); see also Noah D. Zatz, *Introduction: Working Group on the Future of Systemic Disparate Treatment Law*, 32 BERKELEY J. EMP. & LAB. L. 387, 387 (2011) ("Although *Wal-Mart* formally is a case about class certification, the procedural analysis takes shape in the shadow of the substantive theory of liability.'").

3. See Green, *supra* note 2, at 397-98; L. Camille Hébert, *The Supreme Court's 2010-2011 Labor and Employment Law Decisions: A Large and "Mixed Bag" for Employers and Employees*, 15 EMP. RTS. & EMP. POL'Y J. 279, 309 (2011); Natalie Bucciarelli Pedersen, *The Hazards of Dukes: The Substantive Consequences of a Procedural Decision*, 44 U. TOL. L. REV. 123, 137 (2012). Professor Pedersen states:

The majority has rewritten the systemic disparate treatment standard to require some type of corporate policy calling for the use of discrimination by its supervisors or else discrim-

treatment law now sits at a historic crossroads.⁴ Whether it survives, and in what form, depends on the articulation of a coherent theory of systemic disparate treatment that advances the remedial goals of antidiscrimination law without overstating the probative reach of statistical evidence of observed disparities in employment outcomes.⁵

The scholarship on systemic disparate treatment discrimination has, thus far, been unable to articulate such a coherent theory. Rather, the discourse has diverged into two separate analytical strands each focusing on a distinct aspect of systemic disparate treatment law. In the first strand, which can be described as the methodological strand, scholars question the ability of current statistical methodologies to support an inference of unlawful discrimination based only on observed quantitative disparities in employment outcomes.⁶ In the second strand, which may be called the contextualist strand, scholars argue that the systemic disparate treatment theory should be conceptually expanded to impose liability on employers for wrongdoing located at the organizational level, rather than simply functioning as a narrow doctrinal tool for aggregating individual-

inatory acts of one supervisor that reach an entire class of plaintiffs and can be held to constitute a pattern or practice. Neither seems very viable.

Pedersen, *supra*, at 137 (footnote omitted); *see also* Zatz, *supra* note 2, at 387 (“Now *Wal-Mart* threatens to turn that avenue into a dead end, in part by extending to Title VII this Court’s general hostility to the class action device.” (footnote omitted)); Michael J. Zimmer, *Wal-Mart v. Dukes: Taking the Protection out of Protected Classes*, 16 LEWIS & CLARK L. REV. 409, 457 (2012) (“If taken seriously, that interpretation of *Wal-Mart* would eliminate most systemic disparate treatment pattern or practice cases.”); Charles A. Sullivan, *Maybe Systemic Disparate Treatment Isn’t Dead Yet?*, WORKPLACE PROF BLOG (Dec. 10, 2012), http://lawprofessors.typepad.com/laborprof_blog/2012/12/maybe-systemic-disparate-treatment-isnt-dead-yet.html. *But see* Elizabeth Tippett, *Robbing a Barren Vault: The Implications of Dukes v. Wal-Mart for Cases Challenging Subjective Employment Practices*, 29 HOFSTRA LAB. & EMP. L.J. 433, 434–35 (2012) (analyzing lower “court opinions from 2005 to mid-2011” and questioning whether the *Wal-Mart* opinion will have a significant practical effect on lawsuits “challenging subjective employment practices”).

4. In recognition of this, a group of scholars led by Professors Tristin Green and Noah Zatz recently organized a *Working Group on the Future of Systemic Disparate Treatment Law* (Working Group). The stated purpose of the Working Group was to consider what exactly the substantive law of systemic disparate treatment means and to “get a head start on thinking about the post-*Wal-Mart* landscape.” *See* Zatz, *supra* note 2, at 387–88. The Working Group’s work was published in Volume 32 of the *Berkeley Journal of Employment and Labor Law*. *See id.* at 388; Richard Thompson Ford, *Beyond Good and Evil in Civil Rights Law: The Case of Wal-Mart v. Dukes*, 32 BERKELEY J. EMP. & LAB. L. 513, 513 (2011); Green, *supra* note 2, at 398–99 (advancing “a ‘context’ model of organizational wrongdoing” and direct liability as superior to models focused on identifying individual instances of discrimination); Melissa Hart, *Civil Rights and Systemic Wrongs*, 32 BERKELEY J. EMP. & LAB. L. 455, 457–58 (2011); Michael Selmi, *Theorizing Systemic Disparate Treatment Law: After Wal-Mart v. Dukes*, 32 BERKELEY J. EMP. & LAB. L. 477, 481 (2011) (arguing that a statistical showing alone is no longer sufficient to establish a systemic disparate treatment claim and that “it is incumbent upon plaintiffs to explain the story the statistical presentation is telling”).

5. *See* Green, *supra* note 2, at 454 (calling for an “open and frank debate about the theoretical grounding for and shape of systemic disparate treatment law”); Zatz, *supra* note 2, at 391 (“What systemic disparate treatment theory needs is an account that grounds employer liability in firm-level conduct—the connective tissue—[not simply in dispersed individual disparate treatment, and does so without relying exclusively on the extreme case of connective disparate treatment.” (emphasis omitted)).

6. *See infra* Part II.A.

level claims of wrongdoing.⁷ Although there is significant tension between these two strands of scholarship, authors writing in each strand rarely engage the key arguments or insights developed in the other. As a result, courts are left without a coherent theory to make sense of a perplexing subject lying at the intersection of antidiscrimination law and statistical theory. This Article is the first attempt to reconcile the key differences between the methodologists and the contextualists and represents the first step in developing a unified theory of systemic disparate treatment.

The path forward for systemic disparate treatment theory post-*Wal-Mart* requires acknowledging the role of Bayesian priors—preconceptions about background rates of employment discrimination.⁸ At present, the Bayesian priors held by judges and juries in systemic discrimination cases are obscured beneath a veneer of formalistic legal reasoning.⁹ But these hidden priors operate nonetheless, invisibly affecting the outcome of all systemic disparate treatment cases.¹⁰ As this Article will demonstrate, hidden priors can be located in the formative cases that defined systemic disparate treatment theory in the late 1970s,¹¹ as well as in the Court's recent *Wal-Mart* decision, which threatens to severely limit the theory.¹²

The objective of this Article is to develop a unifying path forward for systemic disparate treatment law by exposing the inescapable operation of hidden Bayesian priors. The Article urges courts to openly acknowledge the undeniable role of Bayesian priors in systemic disparate treatment litigation and further urges scholars and courts to take up the

7. See *infra* Part II.B.

8. As the concept is used in this Article, a Bayesian prior is generally a pre-formed or pre-conceived estimate of the likelihood that any given employer engaged in unlawful employment discrimination, made *before* considering any given piece of additional information. Bayesian priors may be used in an iterative fashion, such that the “prior probability” is informed by general preconceptions about background rates, as updated by a cumulative assessment of all the nonstatistical evidence in the case, but *before* considering the purely statistical evidence. See *infra* Part III.A.

9. See Deborah M. Weiss, *The Impossibility of Agnostic Discrimination Law*, 2011 UTAH L. REV. 1677, 1741 (2011) (observing that both the majority opinion and the dissent in *Wal-Mart* “clearly turn on background assumptions that are not well-grounded in evidence”).

10. See *id.* at 1701 (“[J]uries enter the courtroom with pre-existing views about the societal pattern of discrimination and cannot be forced by the fiat of evidentiary exclusion to turn off their views about that societal pattern.”). Judge Richard Posner explains that Bayesian priors can operate at an unconscious level, using discrimination cases as an example. See RICHARD A. POSNER, *HOW JUDGES THINK* 67–68 (2008) (“I used the example of a sex discrimination suit because it is the kind of suit in which judges’ priors are likely to differ along political lines or along racial, religious, or gender lines that are correlated with (and often influence) political leanings, or because of different personal or professional experiences or differences in personality.”).

11. See *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 300 (1977); *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 335 (1977). *Teamsters* is generally considered the seminal case on systemic disparate treatment theory. See, e.g., 1 CHARLES A. SULLIVAN ET AL., *EMPLOYMENT DISCRIMINATION: LAW AND PRACTICE* § 3.01, at 145 (3d ed. 2002); David G. Karro, *Common Sense About Common Claims*, 25 HOFSTRA LAB. & EMPL. L.J. 33, 45 (2007) (calling *Teamsters* “the seminal pattern-or-practice case”).

12. *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541 (2011); see also Weiss, *supra* note 9, at 1740–41.

difficult challenges posed by recognizing the operation of priors. This Article will be the first step in the development of a unified theory of systemic disparate treatment employment discrimination. Embracing priors will open up several challenging questions and new lines of scholarly inquiry about the proper management of priors in antidiscrimination legislation and litigation. As just one example of these challenging questions: whose priors should be determinative?¹³ This Article will highlight some of these second-order questions and frame the future debate over such questions. A related article includes my own suggestions for the proper management of priors in employment discrimination law.¹⁴ The point of this Article, however, is to convince courts and scholars that priors operate even when they are hidden or overlooked—meaning that difficult questions about the management of priors should be openly debated rather than simply ignored.

This Article proceeds in four parts. Part I provides a brief introduction to the systemic disparate treatment theory, including an examination of its foundational cases. Part II illustrates the critical juncture at which systemic disparate treatment law now sits, exploring the two divergent strands of legal scholarship on systemic disparate treatment theory and the potential substantive implications of the *Wal-Mart* decision. Part III lays out the Bayesian path toward a unified theory of systemic disparate treatment law by bringing hidden priors into clear focus. Part IV briefly identifies some of the difficult second-order questions regarding the management of priors and lays out a scholarly agenda for addressing those questions.

I. THE FOUNDATION OF SYSTEMIC DISPARATE TREATMENT LAW

To discern what remains of systemic disparate treatment law, we must first examine the theory's origins. Professor Michael Selmi, in his insightful contribution to the *Working Group on the Future of Systemic Disparate Treatment Law* (Working Group), aptly described the systemic disparate treatment theory as the “most potent and least understood of the various Title VII causes of action.”¹⁵ Professor Selmi, along with Professor Tristin Green, was no doubt correct in characterizing systemic disparate treatment law as “under-theorized.”¹⁶ Indeed, systemic disparate treatment law has been under-theorized from its inception, a point to be

13. Possible answers to this question include the trial judge, the fact-finder, the appellate court, or the legislature.

14. Jason R. Bent, *P-Values, Priors, and Procedure in Antidiscrimination Law*, 63 *BUFF. L. REV.* (forthcoming Dec. 2014).

15. Selmi, *supra* note 4, at 478 (referring to systemic disparate treatment claims as “pattern or practice claims”).

16. See Green, *supra* note 2, at 418, 421; Selmi, *supra* note 4, at 500. Prior to *Wal-Mart*, the Court and scholars had focused their attention primarily on the disparate impact theory of systemic discrimination, rather than systemic disparate treatment. See Jason R. Bent, *The Telltale Sign of Discrimination: Probabilities, Information Asymmetries, and the Systemic Disparate Treatment Theory*, 44 *U. MICH. J.L. REFORM* 797, 798–99, 799 n.5 (2011); Selmi, *supra* note 4, at 478.

demonstrated herein. The theoretical and statistical foundation for systemic disparate treatment law was never properly laid by the Court, leading to the current doctrinal and scholarly confusion.

As originally enacted, Title VII of the Civil Rights Act of 1964 prohibited discrimination “because of race, color, religion, sex, or national origin,” but the text of the statute said nothing about proving unlawful discrimination with statistical evidence of disparities in employment outcomes.¹⁷ In 1977, the Supreme Court recognized the “pattern or practice,” or systemic disparate treatment, theory as one type of proof framework that plaintiffs could employ to establish a prima facie case of discrimination.¹⁸ The theory originated as the Supreme Court’s gloss on the statutory prohibitions found in the text of Title VII.¹⁹ The birth of the systemic disparate treatment theory through judicial construction is similar to the judicial creation of the first individual disparate treatment proof framework in *McDonnell Douglas Corporation v. Green*²⁰ and the disparate impact theory in *Griggs v. Duke Power Company*.²¹ All of these proof frameworks developed originally as judicial scaffolding built on top of the vague and general statutory prohibitions found in Title VII.

The two formative cases for the systemic disparate treatment theory are *International Brotherhood of Teamsters v. United States*²² and *Hazelwood School District v. United States*,²³ both decided in 1977. After 1977, the Supreme Court would not directly address the systemic disparate treatment theory again until *Wal-Mart*, decided in 2011.²⁴ The basic

17. See Civil Rights Act of 1964, Pub. L. No. 88-352, § 701(b), 78 Stat. 241, 254 (1964) (codified as amended in scattered sections of 42 U.S.C., primarily 42 U.S.C. §§ 2000e to 2000e-17). In amendments enacted as part of the Civil Rights Act of 1991, Congress codified a proof structure for establishing unlawful discrimination under the disparate impact theory. See Civil Rights Act of 1991, Pub. L. No. 102-166, § 105(a), 105 Stat. 1071, 1074 (1991) (codified at 42 U.S.C. § 2000e-2(k)). Congress, however, has not explicitly codified a proof structure for the systemic disparate treatment theory. See 42 U.S.C. § 2000e (2012).

18. See *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 336, 358 (1977) (describing the burden-shifting framework of *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973), as only one method of establishing a prima facie case of discrimination in violation of Title VII).

19. In laying out a proof structure for systemic disparate treatment, or “pattern or practice” cases, the *Teamsters* Court drew upon language found in Section 707 of the Civil Rights Act of 1964, codified at 42 U.S.C. § 2000e-6(c). *Teamsters*, 431 U.S. at 328 n.1. This subsection originally granted the Attorney General the authority to investigate and act upon a charge of a “pattern or practice of discrimination.” Civil Rights Act of 1964, 42 U.S.C. § 2000e-6(c), (e) (1964). That authority was subsequently transferred to the Equal Employment Opportunity Commission (EEOC) in 1972. See *id.* § 2000e-6(e).

20. 411 U.S. 792 (1973).

21. 401 U.S. 424 (1971).

22. 431 U.S. 324 (1977).

23. 433 U.S. 299 (1977).

24. See Selmi, *supra* note 4, at 478 (“In no other area of substantive antidiscrimination case law—indeed, perhaps no other area of law—are the leading cases three decades old.”). Even in *Wal-Mart* the substance of the systemic disparate treatment theory was at issue only through the prism of a procedural class certification question. See *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2561–62 (2011). As discussed below, the Court did touch upon the systemic disparate treatment theory in its 1986 decision in *Bazemore v. Friday*, 478 U.S. 385 (1986), which approved the use of multiple

idea underpinning the systemic disparate treatment theory is that unexplained statistical evidence of a quantitatively skewed distribution of hiring, firing, promotion, or other employment outcomes can be evidence that the employer unlawfully and intentionally discriminated against a protected group.²⁵ As the *Teamsters* Court famously put it:

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.²⁶

In *Teamsters*, the Court considered the government's claim that an employer "had engaged in a pattern or practice of discriminating against minorities in hiring so-called line drivers."²⁷ The government alleged that minorities who had been hired by the employer "were given lower paying, less desirable jobs as servicemen or local city drivers, and were thereafter discriminated against with respect to promotions and transfers."²⁸

The *Teamsters* Court promulgated a two-phase framework for evaluating this type of systemic claim of employment discrimination. In Phase I, plaintiffs may establish a prima facie case of systemic discrimination by showing "that unlawful discrimination has been a regular procedure or policy followed by an employer or group of employers."²⁹ The Court unequivocally approved the use of statistical evidence in making this Phase I determination.³⁰ If plaintiffs meet this initial burden, "[t]he burden then shifts to the employer to defeat the prima facie showing of a pattern or practice by demonstrating that the [plaintiff's] proof is either inaccurate or insignificant."³¹ If the employer fails to defeat the plain-

regression analysis. See *infra* note 55 and accompanying text; see also *Bazemore*, 478 U.S. at 399-401.

25. See Deborah A. Calloway, *St. Mary's Honor Center v. Hicks: Questioning the Basic Assumption*, 26 CONN. L. REV. 997, 997 (1994).

26. *Teamsters*, 431 U.S. at 339 n.20. This Article focuses on the importance of the words "absent explanation" in that statement and approaches those words from a Bayesian statistical perspective. See also Paul Meier et al., *What Happened in Hazelwood: Statistics, Employment Discrimination, and the 80% Rule*, in STATISTICS AND THE LAW 1, 20-21 (Morris H. DeGroot et al. eds., 1986) (arguing that the assumption described in *Teamsters* is untenable and can only be taken as an aspirational statement of a court's impartiality); Kingsley R. Browne, *Statistical Proof of Discrimination: Beyond "Damned Lies"*, 68 WASH. L. REV. 477, 482-83, 503-05 (1993) (critiquing this "[c]entral [a]ssumption").

27. *Teamsters*, 431 U.S. at 329.

28. *Id.*

29. *Id.* at 360.

30. *Id.* at 339 ("[S]tatistical analyses have served and will continue to serve an important role' in cases in which the existence of discrimination is a disputed issue." (quoting Mayor of Phila. v. Educ. Equal. League, 415 U.S. 605, 620 (1974))).

31. *Teamsters*, 431 U.S. at 360.

tiffs' prima facie showing of a pattern or practice of discrimination in Phase I, then plaintiffs are entitled to an award of prospective relief, such as an injunctive order.³² In Phase II, plaintiffs may seek individual relief for each person in the alleged group of aggrieved individuals.³³ The determination that the employer engaged in a pattern or practice of discrimination in Phase I leads to a rebuttable presumption that the employer's pattern or practice of discrimination affected each individual claimant for purposes of Phase II.³⁴

In *Teamsters*, the Court held that the government had carried its initial burden in Phase I by offering evidence of a statistical disparity in the hiring of line drivers, along with the testimony of affected individuals to bring the "cold numbers convincingly to life."³⁵ The *Teamsters* case involved an extreme case of statistical disparity. The government showed that of 1,828 line drivers only eight were African-American, and all of those eight were hired after the litigation commenced.³⁶ Responding to the employer's challenge to the specific statistical comparisons used by the government, the Court noted that no amount of fine tuning of the statistical comparisons would change the "inexorable zero"—there had been no African-American line drivers prior to the lawsuit.³⁷

In *Hazelwood*, systemic disparate treatment theory's second formative case, the Court reiterated its approval of the use of statistical evidence to shift the burden of disproving discrimination onto the defendant.³⁸ While the statistical showing in *Teamsters* consisted of a very simple descriptive analysis highlighting the inexorable zero among a relatively large number of line drivers, the statistical analysis at issue in *Hazelwood* was more sophisticated.³⁹ There, the government alleged that the Hazelwood School District had engaged in a "pattern or practice" of discrimination in hiring teachers.⁴⁰ The government offered statistical evidence in the form of a binomial distribution analysis, a concept explored in depth in Part II of this Article.⁴¹ The government argued that a statistical comparison of the percentage of African-American teachers employed by the school district to the percentage of African-American teachers in the relevant comparison labor market could serve as proof of

32. *Id.* at 361.

33. *Id.*

34. Allan G. King, "Gross Statistical Disparities" as Evidence of a Pattern and Practice of Discrimination: Statistical Versus Legal Significance, 22 LAB. LAW. 271, 282 (2007) ("[T]he presumption created primarily by this statistical proof applies to each and every class member and requires the employer to rebut that presumption in each specific instance.").

35. *Teamsters*, 431 U.S. at 339–40, 342–43.

36. *Id.* at 337.

37. *Id.* at 342 n.23 (citation omitted) (internal quotation mark omitted).

38. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307–08 (1977).

39. See, e.g., Zimmer, *supra* note 3, at 422.

40. *Hazelwood*, 433 U.S. at 301 (internal quotation marks omitted).

41. *Id.* at 303, 308; see *infra* Part II.A.I.

a discriminatory pattern or practice in the district's hiring of teachers.⁴² The Court drew upon the reasoning in *Teamsters* and stated that “[w]here gross statistical disparities can be shown, they alone may in a proper case constitute prima facie proof of a pattern or practice of discrimination.”⁴³ The *Hazelwood* Court endorsed the use of binomial distribution analysis to show the required “gross statistical disparities.”⁴⁴ If the difference between the expected number of African-American teachers (calculated by reference to the relevant labor pool, or “reference class”) and the observed number of African-American teachers actually hired by the district “is greater than two or three standard deviations,” the Court stated, the “hypothesis that teachers were hired without regard to race would be suspect.”⁴⁵

Teamsters and *Hazelwood* remained the leading cases on the systemic disparate treatment theory until *Wal-Mart*.⁴⁶ As statistician Paul Meier explained, after *Teamsters* and *Hazelwood*, “a flood of statistical tests of significance, confidence intervals, and multiple regressions thundered forth from the lower courts.”⁴⁷ More recently, systemic theories of discrimination have become a centerpiece of the EEOC’s enforcement strategy. The EEOC’s newly-released Strategic Plan for Fiscal Years 2012–2016 emphasizes the importance of systemic cases, and it includes performance measures unequivocally requiring that an increasing percentage of the EEOC’s cases be systemic cases.⁴⁸ How the *Wal-Mart* decision will impact private and EEOC systemic disparate treatment liti-

42. *Hazelwood*, 433 U.S. at 303, 308–09.

43. *Id.* at 307–08 (citing *Teamsters*, 431 U.S. at 339); see also Selmi, *supra* note 4, at 480 (noting that the early systemic disparate treatment cases were “almost entirely statistical in nature”). After *Hazelwood*, lower courts have frequently acknowledged that a prima facie case of systemic disparate treatment discrimination may be established by statistical evidence alone. See, e.g., EEOC v. Olson’s Dairy Queens, Inc., 989 F.2d 165, 168 (5th Cir. 1993) (relying on statistical comparisons alone to reverse district court’s decision that no prima facie case had been established under the *Teamsters* framework); Satchell v. FedEx Express, No. C 03-2659 SI, 2006 WL 3507913, at *1 (N.D. Cal. Dec. 5, 2006) (plaintiffs can prove their prima facie case “with or without anecdotal testimony”); McReynolds v. Sodexo Marriott Servs., Inc., 349 F. Supp. 2d 1, 17 (D.D.C. 2004) (“[P]laintiffs may sustain their burden at the prima facie stage exclusively on statistical evidence, for ‘no sound policy reason exists for subjecting the plaintiff to the additional requirement of either providing anecdotal evidence or showing gross disparities.’” (quoting Segar v. Smith, 738 F.2d 1249, 1278 (D.C. Cir. 1984))).

44. *Hazelwood*, 433 U.S. at 307–08.

45. *Id.* at 308 n.14 (quoting *Castaneda v. Partida*, 430 U.S. 482, 497 n.17 (1977)) (internal quotation mark omitted).

46. See Selmi, *supra* note 4, at 492–93.

47. See Meier et al., *supra* note 26, at 3.

48. *Strategic Plan for Fiscal Years 2012–2016*, U.S. EQUAL EMP. OPPORTUNITY COMMISSION, http://www.eeoc.gov/eeoc/plan/strategic_plan_12to16.cfm (last visited Aug. 1, 2014). The Strategic Plan anticipates that the EEOC will adopt a performance measure based on the percentage of the cases in the litigation docket that are systemic cases and anticipates that the target for this performance measure will increase every year from 2012 to 2016. See *id.* (Performance Measure 4 for Strategy I.A.3 of Strategy Objective I). As envisioned by the Strategic Plan, the EEOC adopted an accompanying Strategic Enforcement Plan that also emphasizes systemic discrimination litigation. *Id.* (app. B – Strategic Enforcement Plan & Quality Control Plan Timelines).

gation remains to be seen and is the subject of Part II.C below.⁴⁹ But it is clear that systemic disparate treatment law is at a critical juncture. The next Part illustrates the current divide in the legal scholarship about the proper course of systemic disparate treatment law and considers how *Wal-Mart* may represent a turning point.

II. SYSTEMIC DISPARATE TREATMENT AT A CROSSROADS

A. Methodological Strand: The Limits of Statistical Evidence

Statistical evidence forms the cornerstone of most systemic disparate treatment cases.⁵⁰ But the mechanism through which statistical evidence is thought to justify an inference of unlawful discrimination has been frequently misunderstood, or at least glossed over, by courts and scholars.⁵¹ The first major strand of systemic disparate treatment scholarship to develop in the wake of *Teamsters* and *Hazelwood* carefully detailed the inferential limitations of the statistics offered in systemic discrimination cases. Authors writing in this methodological strand took courts to task for misinterpreting and misapplying statistical evidence to arrive at unwarranted conclusions that discrimination, rather than chance or luck, was the most likely explanation for a statistically significant disparity in employment outcomes.⁵²

Understanding the nature of the methodological critiques requires a closer look at the types of statistical evidence commonly offered in systemic discrimination cases. Two types of statistical evidence have been explicitly approved by the Court for use in employment discrimination cases: binomial distributions, approved in *Hazelwood*,⁵³ and multiple regressions, approved by the Supreme Court in *Bazemore v. Friday*,⁵⁴ decided in 1986.⁵⁵ In both binomial distribution and multiple regression, statistical experts look for statistically significant results. Courts often

49. Notably, *Wal-Mart's* holding on the commonality issue is inapplicable to systemic claims brought by the EEOC. The EEOC's systemic discrimination cases need not comply with the class certification requirements of Federal Rule 23. *See Gen. Tel. Co. of the Nw., Inc. v. EEOC*, 446 U.S. 318, 323 (1980).

50. MICHAEL J. ZIMMER ET AL., *CASES AND MATERIALS ON EMPLOYMENT DISCRIMINATION* 135 (7th ed. 2008); Meier et al., *supra* note 26, at 3; Selmi, *supra* note 4, at 485.

51. *See Bent*, *supra* note 16, at 833 n.145 (explaining how courts misinterpret statistical evidence of discrimination); Browne, *supra* note 26, at 486-88 (illustrating where "courts have gone wrong" in the use of statistical evidence of discrimination); Kinsley R. Browne, *The Strangely Persistent "Transposition Fallacy": Why "Statistically Significant" Evidence of Discrimination May Not Be Significant*, 14 LAB. LAW. 437, 437, 441-42 (1998); Richard Lempert, *The Significance of Statistical Significance: Two Authors Restate an Incontrovertible Caution. Why a Book?*, 34 LAW & SOC. INQUIRY 225, 237-41 (2009).

52. *See infra* Part II.A.3-4.

53. *See Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307-08 (1977).

54. 478 U.S. 385 (1986).

55. *Id.* at 387 (involving a claim of salary discrimination); *see also* Roberto Corrada, Ricci's *Dicta: Signaling a New Standard for Affirmative Action Under Title VII?*, 46 WAKE FOREST L. REV. 241, 249 n.59 (2011) (identifying binomial distribution and multiple regression as the two types of statistical analysis approved by the Supreme Court).

consider statistical significance sufficient to establish a prima facie case of systemic discrimination under Phase I of *Teamsters*, thus establishing liability and shifting onto the defendant the burden of disproving discrimination as to the individual claimants or class members.⁵⁶

Many, though not all, antidiscrimination scholars consider these two statistical methodologies adequate to support an inference of unlawful discrimination.⁵⁷ But it is important to understand exactly how such an inference has been justified. The inference of discrimination arises not because binomial distribution or regression analysis directly proves the likelihood that the employer in question discriminated. Instead, the inference is indirect.⁵⁸ The statistical techniques of binomial distribution and multiple regression allow a statistician to conclude that chance alone would be relatively unlikely to lead to the observed disparities given the important assumption that employment decisions were made at random.⁵⁹ This assumption—that employment decisions were made at random—is called the null hypothesis.⁶⁰ If, *assuming the null hypothesis to be true*, pure chance would lead to the observed disparity in less than some predetermined statistically significant level of repetitions (often set at 5% or .05), then a (frequentist)⁶¹ statistician will believe the evidence supports rejection of the null hypothesis.⁶²

If an observed disparity is statistically significant and the null hypothesis of random decision-making can be rejected, proponents of statistical evidence argue, reason and logic should allow the court or fact

56. See King, *supra* note 34, at 272 (noting that in pattern or practice cases applying the *Teamsters* framework “lower courts frequently have turned to ‘statistical significance’ as the measuring rod”); Meier et al., *supra* note 26, at 20–21; Ramona L. Paetzold, *Problems with Statistical Significance in Employment Discrimination Litigation*, 26 NEW ENG. L. REV. 395, 395–96 (1991) (“Although [hypothesis testing] has only been used for a few decades in legal proceedings, it has become the predominant method of statistical analysis used in employment discrimination cases.”); Selmi, *supra* note 4, at 481–82; see also Green, *supra* note 2, at 403 (reading *Teamsters* and *Hazelwood* as “instruct[ing] that if the disparity after accounting for legitimate factors is statistically significant—meaning that it is unlikely due to chance—then an inference of internal systemic disparate treatment can be drawn and entity liability imposed”).

57. See, e.g., RAMONA L. PAETZOLD & STEVEN L. WILLBORN, *THE STATISTICS OF DISCRIMINATION: USING STATISTICAL EVIDENCE IN DISCRIMINATION CASES* § 4.13 (2005); ZIMMER ET AL., *supra* note 50, at 144–45 (“When any one of these techniques is used to conclude that the null hypothesis . . . should be rejected, the next step, based on reason and logic, should be to draw the inference that systemic disparate treatment discrimination has occurred.”). *But see* Browne, *supra* note 26, at 488, 489 n.41 (arguing that neither binomial distribution analysis nor regression analysis should give rise to an inference of unlawful discrimination because the court cannot know the prior unconditional probability of discrimination).

58. See Steven L. Willborn & Ramona L. Paetzold, *Statistics Is a Plural Word*, 122 HARV. L. REV. F. 48, 48 (2009) (“Statistics is imperfect as proof of causation in the same way that every other type of proof is imperfect—it is messy, indirect, uncertain, and subject to varying interpretations.”).

59. See PAETZOLD & WILLBORN, *supra* note 57, § 4.13, at 38–41, § 12.03, at 7–8; ZIMMER ET AL., *supra* note 50, at 142; Corrada, *supra* note 55, at 249 n.59.

60. PAETZOLD & WILLBORN, *supra* note 57, § 4.11, at 36; ZIMMER ET AL., *supra* note 50, at 142.

61. See *infra* Part III.B for a brief discussion of the ongoing philosophical debate between frequentists and Bayesians in the statistics literature.

62. See PAETZOLD & WILLBORN, *supra* note 57, § 4.12, at 37.

finder to take the next step and indirectly infer that the employment decisions at issue were likely the result of discrimination and were not likely caused by chance or some other non-discriminatory factor.⁶³ Not all scholars are comfortable making this indirect inference from evidence of a statistically significant disparity in employment outcomes. Some authors writing in the methodological strand have repeatedly highlighted the limitations of the binomial distribution and multiple regression methodologies currently accepted by courts, and some argue that these statistical techniques are simply inadequate to perform the role prescribed for them by the *Teamsters* and *Hazelwood* opinions.⁶⁴ The statistical techniques at issue, and the methodological critiques, are discussed in further detail below.

1. Binomial Distribution

In a binomial distribution analysis, the statistician typically compares the *observed* percentage of members of a protected group, say African-Americans, that were actually hired by the employer (or promoted, terminated, or subjected to other employment action) to the *expected* percentage of members of that group that would have been hired if the employment decisions were made at random.⁶⁵ Thus, in *Hazelwood*, the government offered evidence comparing the percentage of African-American teachers hired by the Hazelwood School District to the percentage of African-American teachers in St. Louis County and the City of St. Louis, which the government considered to be the proper comparison labor market.⁶⁶ Defendants disputed the relevance of that comparison market by arguing that the City of St. Louis had made special attempts to seek a 50% African-American teaching staff, thus distorting that comparison pool.⁶⁷ The Supreme Court ultimately approved the use of binomial distribution analysis for comparing the racial composition of Hazelwood's teaching staff to "the racial composition of the qualified public school teacher population in the relevant labor market."⁶⁸ The Court, however, left it to the district court on remand to decide the proper comparison labor market.⁶⁹

In performing a binomial distribution analysis, a statistician might find that the disparity between the observed and expected percentage of

63. See ZIMMER ET AL., *supra* note 50, at 144–45.

64. See *infra* Part II.A.3–4.

65. See Meier et al., *supra* note 26, at 6–7; ZIMMER ET AL., *supra* note 50, at 145–51. There are several variants on this simple binomial distribution analysis, including one-sample models, two-sample binomial model, and one- and two-sided significance tests. See Meier et al., *supra* note 26, at 6–15. The methodological details of the variants are beyond the scope of this Article, and the differences between the variants are not important for purposes of the analysis that follows in this Article.

66. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 310 (1977).

67. *Id.* at 310–11.

68. *Id.* at 308.

69. *Id.* at 312–13 ("It is thus clear that a determination of the appropriate comparative figures in this case will depend upon further evaluation by the trial court.")

African-American hires is statistically significant. A finding of statistical significance at the .05 level in a binomial distribution analysis means that random chance would lead to the observed disparity (or an even more skewed disparity) only 1 time in 20 repetitions of the hiring process, assuming that employment decisions in each repetition of the hiring process were made at random.⁷⁰ Based on this result, the statistician has some basis for rejecting the null hypothesis that employment decisions were made at random. Importantly, however, a finding of statistical significance at the .05 level *does not* mean that there is a 95% likelihood that the employer unlawfully discriminated, nor does it mean that there is a 5% likelihood that the disparity was caused by chance.⁷¹

2. Multiple Regression

Multiple regression analysis involves a more sophisticated technique that can model the relationship of multiple independent, or explanatory, variables—such as education level, experience, and gender—with the employment outcome or other dependent variable in question—such as salary—given certain modeling assumptions. In a regression analysis the statistician constructs a model that includes a number of selected independent variables that are thought to have some predictive relationship with the dependent variable in question and then attempts to isolate the predictive effects of each particular independent variable by conditioning on the other included independent variables.⁷² The multiple regression model attempts to determine the characteristics of a mathematical function that best fits or explains all the observed data points for all the included variables in the model.⁷³

One reason that multiple regression models are appealing is that they are able to provide results in the form of a coefficient for each inde-

70. See Paetzold, *supra* note 56, at 400 (“In other words, if the p-value is 0.05 or smaller, the discrepancy between actual hiring practices and assumed hiring practices is considered to be statistically significant. If the p-value is greater than 0.05, then the discrepancy is not statistically significant.”). Setting the cutoff for statistical significance “is somewhat arbitrary.” *Id.* There is no mathematical rule requiring that it be set at .05, but that level is commonly used in the social sciences to identify significant results. *Id.* For a discussion of the historical origins of the .05 level of statistical significance, the work of R.A. Fisher, and Fisher’s acknowledgment that the choice of .05 was subjective, see Michael I. Meyerson & William Meyerson, *Significant Statistics: The Unwitting Policy Making of Mathematically Ignorant Judges*, 37 PEPP. L. REV. 771, 823–24 (2010).

71. E.g., Browne, *supra* note 51, at 449. These common misstatements of statistically significant results are examples of the transposition fallacy. See *infra* Part II.A.3.

72. See ZIMMER ET AL., *supra* note 50, at 151–55; Delores A. Conway & Harry V. Roberts, *Regression Analyses in Employment Discrimination Cases*, in STATISTICS AND THE LAW, *supra* note 26, at 107, 126 (“Multiple regression, as opposed to simple regression, permits simultaneous investigation of the relationship between salaries and multiple job qualifications.”). See generally D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533, 535–36 (2008) (“[A] perusal of the hornbooks and looseleafs discussing the use of statistics as evidence in civil rights litigation suggests that the field is fixated on methods introduced decades ago, particularly regression, despite judicial dissatisfaction.” (footnote omitted)); Willborn & Paetzold, *supra* note 58, at 49 (referring to regression as the “predominant statistical technique used in civil rights litigation”).

73. See Greiner, *supra* note 72, at 541–42.

pendent variable that tell the statistician the direction and estimated magnitude of the effect of a unit change in each particular independent variable on the dependent variable of interest—salary, in this example.⁷⁴ A regression result may indicate that a change in the gender variable from male to female has a negative effect on annual salary (direction) and that a change in the gender variable from male to female results in a decrease in annual salary of \$1,000 (magnitude). A statistician would have some degree of confidence in these coefficient estimates only if certain diagnostic tests for model fit⁷⁵ are satisfied and only if the results found for the independent variable of interest—in this example, gender—are statistically significant.⁷⁶ As with a binomial distribution model, determining whether the coefficient is statistically significant means that the statistician tests the regression results against a null hypothesis. Here, the null hypothesis is that the independent variable we are interested in (gender) has no correlative relationship with the dependent variable (salary). In other words, the null hypothesis is that the coefficient for the gender variable equals 0. A multiple regression result for a given independent variable will be statistically significant at the .05 level—or have a *p*-value of less than .05—if, *assuming the null hypothesis to be true*, pure chance or randomness would have led to the observed correlative results—or even stronger evidence of a correlative relationship—in 5% or less of many repetitions of the analysis.⁷⁷

3. The Transposition Fallacy Challenge

A key critique levied by authors in the methodological strand, and persuasively articulated by Professor Kingsley Browne, is that courts regularly misinterpret statistically significant results obtained from binomial distribution or regression analysis by committing what is often referred to as the “transposition fallacy.”⁷⁸ Courts commit the transposition fallacy, sometimes with the inadvertent help of testifying statistical experts, by forgetting that binomial distribution and regression analyses

74. *Id.* at 541.

75. *See generally id.* at 542 (“When a [regression] diagnostic shows a lack of fit . . . a conscientious analyst alters the model.”).

76. *See* Willborn & Paetzold, *supra* note 58, at 59.

77. For additional information on interpreting the results of multiple regression analyses, see *Interpreting Regression Output*, PRINCETON UNIV., http://dss.princeton.edu/online_help/analysis/interpreting_regression.htm (last visited Apr. 8, 2014).

78. *See* DAVID H. KAYE ET AL., *THE NEW WIGMORE: A TREATISE ON EVIDENCE: EXPERT EVIDENCE* § 12.8.3.a nn. 22–23 (2d ed. 2010) (collecting cases where the error is committed); Browne, *supra* note 51, at 447 (“The error committed by courts engaging in the transposition fallacy is remarkable, however, for its ubiquity; the vast majority of courts that describe the meaning of a *p*-value explain it in terms embodying the fallacy.”); David H. Kaye, *Statistical Significance and the Burden of Persuasion*, 46 *LAW & CONTEMP. PROBS.* 13, 21–23 (1983).

For an interesting discussion of the transposition fallacy in the context of DNA evidence in a criminal case, see David H. Kaye, Commentary, “False but Highly Persuasive”: *How Wrong Were the Probability Estimates in* *McDaniel v. Brown?*, 108 *MICH. L. REV. FIRST IMPRESSIONS* 1 (2009).

are limited to testing a null hypothesis.⁷⁹ Professor Browne documents “cases from virtually all the circuits” that misstate the import of statistical evidence by committing the transposition fallacy.⁸⁰ Thus, in the case of a binomial distribution analysis, a typical misstatement by a court about statistical significance is: “[T]he .05 level of significance . . . [is] certainly sufficient to support an inference of discrimination. [T]he .05 level . . . indicates that the odds are one in 20 that the result could have occurred by chance.”⁸¹

This statement is incorrect. As Professor Browne points out, it confuses two different conditional probabilities: First is the conditional probability that the employer selected employees randomly given the observed disparity—what the court incorrectly claims is shown by the statistics.⁸² Second is the conditional probability that we would see the observed disparity given the assumption that the employer selected employees randomly—what the statistical results actually show.⁸³ A finding of statistical significance at the .05 level means that an employer known or assumed to hire employees at random would nonetheless arrive at results as skewed with respect to the protected characteristic as the actual observed disparity in no more than 1 in 20 repetitions of the hiring procedure. Importantly, statistical significance at the .05 level *does not mean* that it is 95% likely that this employer’s actual hiring in this particular case was non-random, discriminatory, or otherwise suspect. Statistical significance at the .05 level does not mean—contrary to the D.C. Circuit’s assertion—that “the odds are one in 20 that the result could have occurred by chance.”⁸⁴

As Professor Browne emphasizes, the probability that the defendant selected employees discriminatorily is actually a function of another

79. Browne, *supra* note 51, at 447, 447 n.25 (documenting expert testimony that appears to include the transposition fallacy).

80. *Id.* at 447; see also KAYE ET AL., *supra* note 78, § 12.8.3.a. nn.22–23 (collecting state and federal opinions committing the transposition fallacy); Browne, *supra* note 26, at 491–92 (collecting examples from appellate courts, district courts, statistical expert witnesses, and commentators).

81. *Palmer v. Shultz*, 815 F.2d 84, 92 (D.C. Cir. 1987) (citations omitted) (quoting *Segar v. Smith*, 738 F.2d 1249, 1282–83 (D.C. Cir. 1984)) (internal quotation marks omitted), *quoted in* Browne, *supra* note 25, at 491 n.46.

82. See Browne, *supra* note 26, at 507.

83. See *id.*

84. *Palmer*, 815 F.2d at 92 (quoting *Segar*, 738 F.2d at 1282) (internal quotation mark omitted). Notably, the district court in *Wal-Mart* committed the transposition fallacy when describing statistical significance. See *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137, 156 n.23 (N.D. Cal. 2004), *aff’d* 474 F.3d 1214 (9th Cir. 2007), *opinion withdrawn and superseded on denial of reh’g*, 509 F.3d 1168 (9th Cir. 2007), *reh’g en banc*, 603 F.3d 571 (9th Cir. 2010), *rev’d* 131 S. Ct. 2541 (2011) (“Statistical significance is measured by standard deviations. The standard deviation is a number that quantifies the probability that chance is responsible for any difference between an expected outcome and the observed outcome in a sample containing two groups.”). For examples of courts committing the transposition fallacy in the context of DNA match evidence in criminal cases, see KAYE ET AL., *supra* note 78, § 12.8.5 n.96, § 14.1.2.a, and Andrea Roth, *Safety in Numbers? Deciding when DNA Alone Is Enough to Convict*, 85 N.Y.U. L. REV. 1130, 1150–56 (2010) (urging courts to apply Bayes’ Theorem in DNA match cases).

number: “the likelihood of discrimination prior to making the employment decision.”⁸⁵ In other words, the base rate or background rate of discrimination matters when interpreting statistical evidence. If it were known as fact that only 1% of all employers are discriminators, for example, then a binomial distribution analysis finding a disparity statistically significant at a p -value = .05 would mean that for every 100 employers tested we would expect to find 6 “positive” test results—5 false positives due to chance plus 1 true positive due to the employer being guilty of discrimination.⁸⁶ If we assume the base rate of discrimination is only 1%, a positive binomial distribution test (one finding a statistically significant disparity with p -value = .05) does not make it *more likely than not* that the defendant before the court is guilty of discrimination. Rather, in the absence of any evidence other than the base rate assumption and the positive statistical test the probability would be 1 in 6—or about 0.17—that the employer is a discriminator.⁸⁷ Given the base rate assumption of 1%, the statistical evidence from the binomial distribution should therefore be insufficient by itself to establish liability, contrary to *Teamsters* and *Hazelwood*.⁸⁸

A leading employment discrimination text counters Professor Browne’s critique by noting that there is no reason necessarily to assume “that base rate discrimination is especially rare.”⁸⁹ If the base rate of dis-

85. Browne, *supra* note 26, at 488 (“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.”).

86. It should be noted here that this is a simplified example designed to illustrate the importance of the base rate, or background rate, of discrimination. This simplified example ignores the possibility of false negatives—cases in which the employer actually discriminates but the statistical test results are not statistically significant. The relationship between false negatives (or Type II error) and false positives (or Type I error) is complex. “By adopting a standard that minimizes the number of Type I errors, the chance of identifying discrimination when it actually occurs is reduced.” PAETZOLD & WILLBORN, *supra* note 57, § 2.04. One statistician’s model estimates that if the level of statistical significance is set at .05, meaning that there is a 5% risk of Type I error, the corresponding risk of Type II error will be approximately 50%. John M. Dawson, *Scientific Investigation of Fact—The Role of the Statistician*, 11 FORUM 896, 906–08 (1976). Despite the simplifying assumption used in these examples, the central point remains that prior probabilities will necessarily influence the interpretation of the statistical data, and in some cases different prior probabilities will lead to drastically different conclusions about the likelihood of discrimination in any given case, even though the statistical evidence remains exactly the same. This simplified example is used in ZIMMER ET AL., *supra* note 50, at 144.

87. See ZIMMER ET AL., *supra* note 50, at 144.

88. It is not clear precisely where the threshold should be drawn for shifting the burden of proof onto the defendant under the *Teamsters* framework. Absolute certainty is, of course, not required. But what estimate of the probability of discrimination should be considered sufficient to shift the burden onto defendant? In earlier work, I articulated one possible test for determining when to shift the burden of proof under *Teamsters* that considered three factors: the unconditional prior probability of discrimination, or estimates of background rates of discrimination; the strength of the statistical evidence of disparity; and the parties’ relative access to information. Bent, *supra* note 16, at 802.

89. See ZIMMER ET AL., *supra* note 50, at 144 (acknowledging that Professor Browne is “theoretically correct” but pointing out that he “fails to demonstrate why the legal system should conclude that base rate discrimination is especially rare”). Professor Zimmer et al., also point out the limited legal effect of a statistical showing establishing the plaintiffs’ prima facie case. The defendant has

crimination was assumed to be 10% instead of 1%, then for every 100 employers tested we would find 15 “positive” results—5 false positives due to chance plus 10 true positives due to the employer being guilty of discrimination.⁹⁰ In the absence of any other evidence, the probability would then be 10 in 15—or about 0.67—that the defendant-employer is guilty of discrimination.⁹¹ Given this higher base rate assumption of 10%, it might make sense to shift the burden of proof onto the defendant upon a showing of statistical evidence with a p -value = .05. But rather than refuting Browne’s argument, this point only underscores the importance of the base rate, i.e., the prior probability, of discrimination. A change in base rate from 1% to 10% made all the difference!

The base rate problem is a flaw that undermines our current understanding of the *Teamsters* approach to systemic disparate treatment law. Reasonable people can disagree about the true base rate of employment discrimination and therefore can disagree about the prior probability of discrimination in any given case. Hence, reasonable people can disagree about whether an inference of discrimination or the imposition of liability for systemic disparate treatment should attach upon a showing of a statistically significant disparity with a p -value = .05.⁹²

Professor Browne’s attack leads him to conclude that current statistical methods for proving systemic disparate treatment should probably be “abandoned altogether.”⁹³ Browne also suggested, however, that traditional statistical proof might be acceptable if a more rigorous standard of statistical significance than .05 were applied and if courts required “strong[] anecdotal evidence” to buttress the statistical showing.⁹⁴

“the opportunity to rebut by offering proof that it does not discriminate. Sufficiently strong testimony might convince the jury that it was chance that explained the disparity.” *Id.*

90. *See id.*

91. *See id.*

92. *See id.* (“The problem is a ‘base rate’ one . . .”).

93. Browne, *supra* note 26, at 553 (“Hypothesis testing, with its reliance on the assumption that the resultant p -value represents the probability that the observed distribution was a consequence of chance and its declaration of results as ‘statistically significant,’ should be abandoned altogether. Such evidence is simply irrelevant to the ultimate question.” (footnote omitted)).

94. *Id.* at 541–42, 554. Browne appears to acknowledge the possibility of a continuing role for statistical evidence but only where “substantially more rigorous criteria” are applied. *Id.* at 554. He contends that if statistical analyses are to be used, courts should require that they show “gross statistical disparities,” rather than just ordinary statistical significance (often set at the .05 level), that they be accompanied by “strong anecdotal evidence” of discrimination, and that courts adhere to proper allocations of burdens of proof. *See id.* at 541–42, 549, 554. Professor Browne concludes:

If statistical proof of discrimination is still to be acceptable at all in court—which perhaps is doubtful—courts must pay more than lip service to the principle that throughout the litigation it is the plaintiff’s burden to demonstrate that impermissible discrimination is “the company’s standard operating procedure—the regular rather than the unusual practice.”

Id. at 555 (quoting *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 336 (1977)).

4. The Potential Outcomes Challenge

Professor Browne is not alone in criticizing traditional statistical evidence offered to show systemic disparate treatment.⁹⁵ Professor D. James Greiner recently levied an attack on the use of multiple regression analysis in discrimination cases.⁹⁶ He argues that multiple regression analysis often “give[s] the wrong answer, or contradictory answers” to the questions that are actually important in systemic discrimination litigation.⁹⁷ The key problem, as Professor Greiner sees it, is that regressions “lack . . . a framework for causal inference.”⁹⁸ According to Greiner, the legal community seems to believe, erroneously, “that one can measure the causal effect of any variable by including it on the right-hand side of the equals sign in a regression equation.”⁹⁹

Professor Greiner focuses on several specific shortcomings of multiple regression models, ranging from the judgment calls required in selecting which independent variables to include in the model to the potential for analyst bias in specifying the regression model and testing for fit.¹⁰⁰ Fundamentally, he argues that regression models cannot support a causal inference and therefore cannot tell judges and juries what they actually would like to know—how likely is it that the observed disparities in employment outcomes were caused by the defendant’s unlawful discrimination?¹⁰¹

Professor Greiner recommends an alternative quantitative tool for attempting to measure causal inference in discrimination cases—potential outcomes analysis.¹⁰² In contrast to multiple regression, the potential outcomes technique attempts to approximate, to the extent possible, a randomized experiment using only observational data drawn from observed employment outcomes.¹⁰³ The potential outcomes model matches up pairs—or sometimes small groups—of observational data points that are closely correlated in all potentially relevant explanatory covariates (e.g., position, performance evaluations, years of experience,

95. See also Ben Ikuta, Note, *Why Binomial Distributions Do Not Work as Proof of Employment Discrimination*, 59 HASTINGS L.J. 1235, 1235 (2008).

96. See Greiner, *supra* note 72, at 534.

97. *Id.* at 538.

98. *Id.* at 543.

99. *Id.* at 544.

100. See *id.* at 544–56.

101. *Id.* at 556–57.

102. *Id.* at 557. Potential outcomes analysis is also sometimes referred to in the literature as “matching,” the “counterfactual” model, or the Rubin Causal Model. See, e.g., ROBERT B. SMITH, MULTILEVEL MODELING OF SOCIAL PROBLEMS: A CAUSAL PERSPECTIVE 13 (2011). The modern revival of the potential outcomes method of causal inference is often attributed to Donald Rubin. See Guido W. Imbens & Donald B. Rubin, *Causal Inference in Statistics and Social Sciences* § 1.1–1.12 (Jan. 30, 2012) (unpublished manuscript) (on file with author), for a detailed explanation of the potential outcomes model.

103. See Greiner, *supra* note 72, at 537–38, 557–58.

education, training) but that diametrically differ on the explanatory variable of interest (e.g., gender).¹⁰⁴

In a potential outcomes model, this matching is typically performed by creating a sort of index statistic that measures the combination of each individual's covariate characteristics, called a "propensity score."¹⁰⁵ Once the data is sorted into matched pairs using propensity scores, it can then be considered a rough approximation of a randomized, controlled experiment in which one group is the control group (male) while the other is the treatment group (female).¹⁰⁶ The benefit of matching is that it allows the researcher to "borrow" the unobserved, counterfactual potential outcome for any given individual from the other individual in the matched pair.¹⁰⁷ In other words, a potential outcomes model provides some basis for inferring what the counterfactual state of the world would look like—a female's salary, if that female had been male, but in all other respects identical.¹⁰⁸ After matching, employment outcomes are compared across the matched pairs to estimate the average treatment effect i.e., the average effect on salary of any given individual being female, rather than male.¹⁰⁹ The researcher can further narrow the results by looking at the average treatment effect on salary for only those in the treatment group—females—or for only those in certain portions of the salary distribution.¹¹⁰

By isolating matched pairs, the potential outcomes technique has the benefit of controlling for covariates using research model design, rather than relying on a multiple regression analysis with built-in assumptions about linearity and constant effects across the salary distribution.¹¹¹ Further, a statistician conducting a multiple regression analysis will usually need to peek at the regression coefficients and *p*-values before evaluating whether the regression model was a satisfactory fit to the data set.¹¹² Professor Greiner emphasizes the importance of this difference, noting that an expert using a potential outcomes analysis is more credible because the model can be fully specified *before* peeking at the results.¹¹³

The potential outcomes approach may represent an improvement over regression in many cases, and it has unquestionably increased in

104. *See id.* at 570–73.

105. *See id.* at 574–75.

106. *See id.* at 575–76.

107. *See id.* at 562 (referring to donated values for counterfactual states).

108. *See id.*

109. *See id.* at 558–60.

110. *See id.* at 567–68.

111. *See id.* at 548, 580–81.

112. *See id.* at 544.

113. *See id.*

popularity in the social sciences.¹¹⁴ Nonetheless, it still falls short of the “gold standard” of a completely randomized experiment, as does any technique that relies on observational data.¹¹⁵ Whenever a research design includes an “assignment mechanism”—that is, the force that determines whether an individual ends up in the treatment group or the control group—that is outside the researcher’s control and that is non-random, the results will necessarily be open to challenge as a basis for causal inference.¹¹⁶ Of course, in the context of employment discrimination litigation the assignment mechanism is always non-random and is always out of the researcher’s control. A researcher obviously cannot randomly assign individual employees to a gender treatment group—male or female—and then observe the results. Without a controlled, randomized experiment, the potential exists that unobserved covariates are confounding the analysis making the results misleading. Unobserved covariates cannot be controlled for in a regression model and cannot be included in a propensity score calculation to match individuals in a potential outcomes model.

The potential outcomes model suffers from other drawbacks as well, as highlighted by Professors Steven Willborn and Ramona Paetzold.¹¹⁷ Importantly, a potential outcomes analyst must make “complex decisions about which variables to include” in performing the task of matching pairs.¹¹⁸ These decisions are similar to the difficult choices that regression analysts make when determining which covariates to include in their regression models. If the calculation of propensity scores misses certain important unobserved variables, then it is open to bias in much the same way as a regression.¹¹⁹ Second, the potential outcomes approach usually requires a process called “trimming,” which throws out some potentially relevant data—those data points for which there is no match close enough on the covariates in the other group.¹²⁰ The outliers in the tails of the distribution of covariates are simply disregarded on the justification that they do not provide much useful information for drawing inferences anyway.¹²¹ Throwing out data may amount to ignoring

114. See STEPHEN L. MORGAN & CHRISTOPHER WINSHIP, COUNTERFACTUALS AND CAUSAL INFERENCE: METHODS AND PRINCIPLES FOR SOCIAL RESEARCH 87 (2007) (“[A]mong social scientists who adopt a counterfactual perspective, matching methods are fast becoming an indispensable technique for prosecuting causal questions, even though they usually prove to be the beginning rather than the end of causal analysis on any particular topic.”).

115. See Willborn & Paetzold, *supra* note 58, at 50; Imbens & Rubin, *supra* note 102, § 15.1.

116. See Imbens & Rubin, *supra* note 102, § 8.1, § 15.1.

117. See Willborn & Paetzold, *supra* note 58, at 48–56.

118. See *id.* at 52 (“Inferences about outcomes are only as good as the covariates that have been included. A similar point is also true for regression—variables that are not included in the regression model are not controlled and their effects on the outcome cannot be ascertained.”); see also KAYE ET AL., *supra* note 78, § 12.5.3 n.43 (noting that a potential outcomes “mode of analysis also demands an appreciation of potential confounders and adequate data if it is to ‘balance covariates’”).

119. See MORGAN & WINSHIP, *supra* note 114, at 122 (noting that matching is vulnerable to a “selection on the unobservables” bias).

120. See Willborn & Paetzold, *supra* note 58, at 55.

121. See *id.*; Greiner, *supra* note 72, at 566.

potentially relevant evidence of discrimination.¹²² Further, by trimming the data to only those areas of the distribution of covariates where the treatment group and the control group have sufficient overlap, a conscientious potential outcomes analyst will properly report the results as the estimate of a defined local average treatment effect—that is, the average treatment effect only for those individuals falling in the overlap region.¹²³ This means that the analyst cannot make causal claims about the effect of gender on wages for those who fall outside the region of covariate overlap, which may very well include some plaintiffs.

Finally, and most fundamentally, as Professors Willborn and Paetzold explain, the potential outcomes model ultimately requires the same indirect causal inference to discrimination that regression models require.¹²⁴ After matching, trimming, and generating an estimated local average treatment effect, the final step in a potential outcomes analysis is to calculate the standard error for the estimate so that the researcher can determine whether the estimated effect is statistically significant.¹²⁵ That is, the analyst will attempt to determine how often random chance would lead to the observation of an average treatment effect as large or larger than the one observed and test the null hypothesis that the treatment—gender = female—has zero effect. If the researcher rejects the null hypothesis, then it may permit an indirect inference that unlawful discrimination caused the observed outcome disparities. This indirect mode of inferring a causal effect means that potential outcomes results are vulnerable to the same type of misinterpretation via the transposition fallacy as the examples discussed in the preceding subpart.

So where does this discourse in the methodological strand leave us? Professors Browne and Greiner both appear to believe that the currently-used statistical methodologies of binomial distribution and regression modeling are inadequate to justify an inference of discrimination. Professor Greiner advocates the use of potential outcomes models to avoid implausible assumptions and potential analyst bias involved in regression analysis.¹²⁶ Professors Willborn and Paetzold freely admit the messy and indirect nature of statistical inference using the traditional techniques, but nonetheless contend that flawed statistical evidence—whether binomial distributions, regressions, or potential outcomes—can still convey *some*

122. Willborn & Paetzold, *supra* note 58, at 55.

123. See MORGAN & WINSHIP, *supra* note 114, at 117 (“Resulting estimates are then interpreted as estimates of a narrower treatment effect: the common-support treatment effect for the treated.”).

124. See Willborn & Paetzold, *supra* note 58, at 60 (stating that nonstatistical circumstantial evidence, regression, and potential outcomes all use the same causal framework by “answering the question of how likely it is that we would see this outcome in the absence of discrimination”).

125. See MORGAN & WINSHIP, *supra* note 114, at 118 (“After computing a matching estimate of some form, most researchers naturally desire a measure of its expected variability across samples of the same size from the same population, either to conduct hypothesis tests or to offer an informed posterior distribution for the causal effect that can guide subsequent research.”).

126. See *supra* notes 98–117 and accompanying text.

relevant information and therefore has *some* value in litigation, just like any piece of nonstatistical, circumstantial evidence of discriminatory intent.¹²⁷ Meanwhile, in the absence of a coherent methodological critique from the legal academy, courts continue to rely on binomial distributions and multiple regressions and regularly misinterpret the meaning of statistically significant results obtained therefrom.¹²⁸

B. Contextualist Strand: Organizational Causes of Discrimination

The second strand of scholarship on systemic disparate treatment theory has an entirely different focus. Led by Professor Tristin Green's contribution to the Working Group, the contextualist strand represents an attempt to expand the conception of systemic disparate treatment law to impose liability on employers at the entity level.¹²⁹ Under this view, systemic disparate treatment was not meant to be a simple aggregation of multiple instances of discrimination occurring at the individual level. Rather, systemic disparate treatment law serves a broader purpose: "[T]he employer's responsibility under this model turns not on identification of a single instance or even multiple instances of disparate treatment; rather, its responsibility turns on its own role in producing disparate treatment within its walls."¹³⁰ Professor Green, drawing upon the literature on corporate criminality and organizational studies, urges that employers be held *directly* responsible for what the entity itself has done wrong—encouraging or allowing discrimination to thrive by the entity's "set of attitudes and positions, which influence, constrain, and at times even define the modes of thinking and behavior of the people who populate it."¹³¹

Instead of asking whether an entity can be held vicariously liable for negligence in supervising its individual policymakers or employees,¹³² Professor Green's context model would seek to locate wrongdoing at the

127. See *supra* notes 118–29 and accompanying text. The difficulties identified by scholars writing in the methodological strand are nicely encapsulated by Professor Selmi in a story he relates in his contribution to the Working Group:

The role statistics play in systemic discrimination cases has always been a bit of a mystery, a fact that was brought home to me recently at a conference I attended with mostly philosophers present. During a discussion of the *Wal-Mart* case, one of the philosophers asked, rather incredulously, how can statistics prove intent? . . . [W]e had a difficult time providing an answer, other than to point to some of the cases.

Selmi, *supra* note 4, at 480.

128. See *supra* notes 82–86 and accompanying text.

129. See Green, *supra* note 2. Professor Green's proposed entity-level conceptualization of systemic disparate treatment appears to be an idea with which Professors Hart and Ford generally concur. Hart, *supra* note 4, at 456 n.2; see Ford, *supra* note 4, at 513–14 (advancing a "collective justice" approach to systemic disparate treatment).

130. Green, *supra* note 2, at 439.

131. *Id.* at 439 (quoting Eli Lederman, *Models for Imposing Corporate Criminal Liability: From Adaptation and Imitation Toward Aggregation and the Search for Self-Identity*, 4 BUFF. CRIM. L. REV. 641, 686 (2000)) (internal quotation mark omitted).

132. As Professor Green observes, "[E]ntity liability for systemic disparate treatment has never turned on these concepts." Green, *supra* note 2, at 429.

entity level. In her words, the context model “asks whether the entity is producing wrongdoing on an aggregate basis within its walls rather than asking exclusively whether an identifiable high-level policy maker or low-level decision maker acted with purpose or intent to harm or whether the entity has done enough to police wrongdoing of individual actors.”¹³³

Professor Green’s proposal of a contextual model is an important step in systemic disparate treatment scholarship. By drawing upon studies of organizational dynamics, the contextual model appropriately recognizes that organizations can produce disparate treatment within the organization even though it may be “difficult to isolate identifiable wrongdoers.”¹³⁴ The *Wal-Mart* case itself may present an example of such a situation. There, Walmart left promotion decisions to the subjective discretion of store managers.¹³⁵ The plaintiffs alleged that this decentralized, discretionary promotion policy resulted in discrimination against female employees in the aggregate at Walmart stores across the country.¹³⁶ When each individual promotion decision in a case like *Wal-Mart* is examined in isolation it may not be possible to affirmatively identify a definitive instance of intentional disparate treatment discrimination. Yet, social science evidence suggests that an observed company-wide disparity in promotion outcomes may nonetheless be caused by certain organizational influences, attitudes, and cultures that “creat[e] the environment in which interactions and decisions take place.”¹³⁷

Professor Green’s approach places a focus on understanding the causal relationship between organizational dynamics and employment outcomes, but it does not tackle the fundamental problems of inference highlighted by the ongoing debate in the methodological strand. In her context model of systemic disparate treatment, Professor Green would determine liability by “ask[ing] whether the entity is producing wrongdoing on an aggregate basis within its walls,” but Professor Green would appear to accept statistically significant results from binomial distributions and multiple regression analyses to answer that question.¹³⁸ She argues that “[s]tatistics serve as evidence of regular, widespread internal disparate treatment.”¹³⁹ Professor Green explains her view of the power

133. *Id.* at 398–99. Professor Selmi would not ascribe to this view if it would go so far as to impose direct liability on employers for “passively facilitating discrimination” that is not at least causally connected to the employer’s “broader cultural norms within the firm.” Selmi, *supra* note 4, at 503. Professor Selmi argues that such a conceptualization of systemic disparate treatment “would come close to requiring employers to implement some form of affirmative action, something no court is likely to require.” *Id.*

134. Green, *supra* note 2, at 436.

135. See *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2547 (2011).

136. See *infra* Part II.C (discussing the facts and holding from *Wal-Mart*).

137. Green, *supra* note 2, at 440; see also Tristin K. Green, *Targeting Workplace Context: Title VII as a Tool for Institutional Reform*, 72 *FORDHAM L. REV.* 659, 662–63 (2003).

138. Green, *supra* note 2, at 398.

139. *Id.* at 444. Professor Green refers to “sophisticated statistical analyses” that can control for external factors, an apparent reference to multiple regression. *Id.*

of statistical evidence “[s]tatistics cannot tell us whether one would see a particular observed disparity in the absence of discrimination, but *they can indicate the likelihood that an observed disparity (after accounting for legitimate variables) is due to chance.*”¹⁴⁰

Other scholars have made similarly inflated claims about the power of statistics in systemic discrimination cases to prove the likelihood that observed disparities are due to chance. Professor Selmi may not fully endorse Professor Green’s contextual model,¹⁴¹ but he nonetheless echoes her claim about statistical evidence: “In the context of discrimination claims, that [statistically significant] disparity will generally be attributed to discrimination since the function of the standard deviation analysis is to *rule out chance fluctuations.*”¹⁴² Professor Michael Zimmer, discussing binomial distribution analysis in a recent article about *Wal-Mart*, states: “If there is a statistically significant relationship between sex and promotions, that relationship makes it *extremely unlikely* to be the result of chance.”¹⁴³

The foregoing statements all exaggerate, to greater or lesser degrees, the power of statistical evidence obtained from binomial distribution and multiple regression analyses to cast doubt on chance as the likely explanation for observed disparities. Just like the D.C. Circuit in the passage quoted above in Part II.A.,¹⁴⁴ scholars frequently convert statistical significance into something that it is not. A statistically significant result *does not* justify the conclusion that chance can be ruled out as an explanation for an observed disparity in a particular case, at least not absent some hidden assumption about background rates of discrimination.¹⁴⁵ Nor can a statistically significant result, in the absence of base rate information, provide us with a quantified likelihood that observed disparities are due to chance rather than discrimination.¹⁴⁶ To rule out chance as an unlikely explanation for observed disparities, one must consider the statistical results in light of information or estimates about the

140. *Id.* at 403 (emphasis added).

141. Selmi, *supra* note 4, at 503 n.107. Professor Selmi would require that the disparate employment outcomes are somehow traced to particular employees or at least to “broader cultural norms within the firm.” *Id.* at 503.

142. *Id.* at 482 (emphasis added). Likewise, Professor Ford in his contribution to the Working Group writes: “Of course that [statistical evidence] can be and often is manipulated too, but good statistical analysis can distinguish real evidence from spin.” Ford, *supra* note 4, at 521. Ford explains that as sample sizes increase, the “potential for chance to affect the analysis” decreases. *Id.* This is a recognition that large sample sizes tend to more easily produce statistically significant binomial distribution or regression results. See PAETZOLD & WILLBORN, *supra* note 57, § 4.15.

143. Zimmer, *supra* note 3, at 442 (emphasis added). Professor Zimmer makes a more modest claim about the results of multiple regression analysis: “Holding all the variables other than sex constant, the technique shows whether there is a statistically significant relationship between pay and sex. If there is, the null hypothesis that sex and pay are unrelated should be rejected.” *Id.* at 442–43.

144. See *supra* text accompanying note 82.

145. See *supra* Part II.A.3; see also Weiss, *supra* note 9, at 1746.

146. See *supra* text accompanying note 87.

base rate of discrimination.¹⁴⁷ Unfortunately, statistical analysis of observational data alone, without some hidden assumption about background rates, cannot “indicate the likelihood that an observed disparity (after accounting for legitimate variables) is due to chance.”¹⁴⁸

Recall that if the background rate of discrimination is assumed to be 1% in the simplified example, then a statistically significant finding at exactly $p\text{-value} = .05$ would mean that the likelihood is 1 in 6—or about 17%—that the observed disparity is due to discrimination (true positive) and 5 in 6—or about 83%—that the observed disparity is due to chance (false positive).¹⁴⁹ This result patently falls short of truly “rul[ing] out chance fluctuations.”¹⁵⁰ When operating with a 1% base rate assumption, chance is actually a much more likely explanation for the observed disparity than discrimination. What the statistically significant results do show, given a 1% base rate assumption, is that it now appears relatively more likely that the employer in question discriminated (17%) than we would have estimated without the benefit of the observed statistical evidence of disparity (1%). The statistical showing of a disparity does offer valuable and probative information that should be considered, but it simply does not and cannot rule out chance as an explanation for observed disparities.¹⁵¹

Neither Professor Green nor Professor Selmi, in their contributions to the Working Group, nor Professor Zimmer, in his separate work on the substantive implications of *Wal-Mart*, directly address the methodological arguments about the limitations of statistical evidence advanced by Professors Browne and Greiner. Although authors in both scholarly strands are trying to make sense of systemic disparate treatment law, the two scholarly strands have, unfortunately, rarely intersected. The methodological critics focus intently on the limitations of statistical evidence without addressing legitimate, larger questions about the importance of organizational influences on observed employment outcomes and the difficulty of identifying and proving discrimination.¹⁵² Yet, the contextualists advocating a more expansive conceptualization of systemic disparate treatment tend to gloss over the very real limitations of statistical evidence as a tool for reliably identifying employers that are actually guilty of “producing disparate treatment within [their] walls.”¹⁵³ For sys-

147. See *supra* text accompanying note 87.

148. Green, *supra* note 2, at 403; see also *supra* text accompanying note 143.

149. See *supra* text accompanying notes 88–94.

150. Selmi, *supra* note 4, at 482.

151. See KAYE ET AL., *supra* note 78, § 12.8.2.b (“Chance affects the data, not the hypothesis. With the frequency interpretation of chance, there is no meaningful way to assign a numerical probability to the null hypothesis.”).

152. See *supra* Part II.A.3–4.

153. See Green, *supra* note 2, at 439. Interestingly, Professor Ford chalks up general skepticism toward statistical evidence as a product of cognitive bias. See Ford, *supra* note 4, at 519–20. Under Professor Green’s conceptualization of systemic disparate treatment we would, by the definition of statistical significance at the .05 level, necessarily expect at least 5% of all employers covered

temic disparate treatment law to survive as an important tool in the anti-discrimination toolbox, the important theoretical advancements made in both strands must be harmonized. With apologies to Professor Zatz, what systemic disparate treatment *really* needs is an inclusive rethinking of its theoretical foundations that incorporates the contributions of both the methodological critics and the contextualists.¹⁵⁴

C. Wal-Mart: *The End of Systemic Disparate Treatment?*

Unexplored tensions between the methodologists and the contextualists were manifested in the *Wal-Mart* decision. *Wal-Mart* was nominally a case about the requirements of commonality and typicality under Federal Rule of Civil Procedure 23 governing class certification.¹⁵⁵ But Justice Scalia's majority opinion in *Wal-Mart* has potentially profound implications for the substantive law of systemic disparate treatment given the skepticism with which the majority viewed statistical evidence of observed outcome disparities.¹⁵⁶

Wal-Mart involved claims of discrimination on the basis of sex in the promotion and pay practices at the giant retail chain.¹⁵⁷ Walmart is the largest private employer in the United States, and the purported class seeking Rule 23 certification included about 1.5 million female current and former Walmart employees.¹⁵⁸ Plaintiffs claimed that a decentralized decision-making structure at Walmart led to discrimination against women.¹⁵⁹ Plaintiffs brought claims under both the systemic disparate treatment theory and the disparate impact theory—a point that is somewhat obscured by the Court's decision.¹⁶⁰ The plaintiffs' theory, as described by Justice Scalia, was "that a strong and uniform 'corporate culture' permits bias against women to infect, perhaps subconsciously, the discre-

by antidiscrimination laws to be held liable for "producing disparate treatment within [their] walls" based on a finding of statistically significant outcome disparities. See Green, *supra* note 2, at 439. This is true because, assuming that decisions were made at random, *by chance alone* we would expect to observe such statistically significant disparities at a rate of 5%, regardless of the number of covered employers that, in truth, are guilty of unlawful discrimination.

154. Cf. Zatz, *supra* note 2, at 391 ("What systemic disparate treatment theory needs is an account that grounds employer liability in firm-level conduct—the connective tissue—[n]ot simply in dispersed individual disparate treatment, and does so without relying exclusively on the extreme case of connective *disparate treatment*.").

155. To the extent that *Wal-Mart* is truly limited to application of Rule 23, it should have no bearing on systemic disparate treatment actions brought by the EEOC on behalf of groups of aggrieved persons. EEOC pattern or practice claims are not governed by the class certification requirements of Rule 23. See Gen. Tel. Co. of the Nw., Inc. v. EEOC, 446 U.S. 318, 327 (1980).

156. Green, *supra* note 2, at 405. Professor Green states:

If the Supreme Court continues down the path set by the majority opinion in *Wal-Mart* and adopts an individualistic theoretical foundation for systemic disparate treatment law . . . the Court's 'procedural' decision will result in drastic change in the substantive law of systemic disparate treatment as it has been practiced for more than three decades.

Id.

157. *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2547 (2011).

158. *Id.*

159. *Id.* at 2548.

160. Green, *supra* note 2, at 407 n.45.

tionary decisionmaking of each one of Walmart's thousands of managers."¹⁶¹

The district court certified the proposed class under Rule 23(b)(2), and a divided en banc Ninth Circuit Court of Appeals affirmed the certification.¹⁶² A majority of the Supreme Court reversed, finding that the Rule 23 requirement of commonality had not been met.¹⁶³ The Court, relying heavily on *General Telephone Co. of the Southwest v. Falcon*,¹⁶⁴ determined that commonality in a case like this could be shown in one of two ways: (1) where the purported class members were subjected to the same biased testing procedure for purposes of evaluation by the employer; or (2) where the plaintiffs could show "[s]ignificant proof that an employer operated under a general policy of discrimination" it might be conceivable to have a broad class including both applicants for hire and incumbent employees who are denied promotion.¹⁶⁵

From this, the Court reasoned that commonality was lacking because plaintiffs had adduced no "significant proof" that Walmart "operated under a general policy of discrimination," as the majority thought the *Falcon* Court's second prong demanded.¹⁶⁶ Instead, the Court thought there was evidence of the opposite, finding it significant that "Wal-Mart's announced policy forbids sex discrimination."¹⁶⁷ In a revealing passage, the Court explained why commonality must be lacking in Walmart's decentralized decision-making system:

[L]eft to their own devices *most managers in any corporation—and surely most managers in a corporation that forbids sex discrimination—would select sex-neutral, performance-based criteria for hiring and promotion that produce no actionable disparity at all.* Others may choose to reward various attributes that produce disparate impact—such as scores on general aptitude tests or educational achievements. And still other managers may be guilty of intentional

161. *Wal-Mart*, 131 S. Ct. at 2548.

162. *Id.* at 2549.

163. *Id.* at 2556–57.

164. 457 U.S. 147 (1982).

165. *Id.* at 159 n.15. The *Falcon* case was focused on the question whether, given Rule 23's commonality and typicality requirements, a single class under Rule 23 could include two separate groups of individuals: (1) those who were denied hire by the defendant, and (2) those who were denied promotion by the defendant. *Id.*; see also Green, *supra* note 2, at 410. In the early development of systemic disparate treatment law, cases would often proceed as "across the board" discrimination cases, where the class of aggrieved individuals would include combinations of different protected groups and combinations of different adverse employment actions (such as failure to hire, failure to promote, and termination). Melissa Hart, *Will Employment Discrimination Class Actions Survive?*, 37 AKRON L. REV. 813, 818 (2004).

166. *Wal-Mart*, 131 S. Ct. at 2553 (quoting *Falcon*, 457 U.S. at 159 n.15). The first prong of *Falcon* was inapplicable, because the plaintiffs in *Wal-Mart* did not allege that they were subjected to any biased uniform testing procedure or evaluation method. *Id.*

167. *Id.* Importantly, the defendant in *Hazelwood* had a similar official policy of non-discrimination, providing that the School District would "hire all teachers on the basis of training, preparation and recommendations, regardless of race, color or creed." *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 303–04 (1977).

discrimination that produces a sex-based disparity. In such a company, demonstrating the invalidity of one manager's use of discretion will do nothing to demonstrate the invalidity of another's.¹⁶⁸

Professor Green rightly notes that Justice Scalia's majority opinion appeared to embrace a policy-required view of the systemic disparate treatment theory, representing a break with the Court's precedent.¹⁶⁹ Nothing in *Teamsters* or *Hazelwood* required the plaintiffs to identify a specific policy or a specific individual decision-maker that caused the observed disparity. Rather, the Court in *Hazelwood* plainly allowed statistical evidence of a gross disparity alone to establish a prima facie case of systemic disparate treatment discrimination without requiring the plaintiff to single out any particular policy or mechanism responsible for generating the observed disparity.¹⁷⁰ Identifying a particular employment practice or policy leading to a disparity is required in *disparate impact* cases,¹⁷¹ but until *Wal-Mart*, it had never been required in systemic disparate treatment cases.¹⁷²

The *Wal-Mart* opinion nicely highlights a tension between the two scholarly strands. On a contextual view, the *Wal-Mart* plaintiffs established all that was necessary to show that Walmart was producing disparate treatment within its walls—statistically significant employment outcome disparities. Under Professor Green's model, this would be sufficient to hold *Wal-Mart* directly liable for its own entity-level wrongdoing. The various differences between class claimants, including their store location, their supervisors, their managers, their job duties, and so on, would be entirely irrelevant to this determination, making class certification appropriate. Methodological scholars criticizing overreliance on tests of statistical significance, however, would likely reach exactly the opposite conclusion. For scholars like Professor Browne, the statistically significant results of traditional statistical techniques, without something more, would not be sufficient to justify an inference of disparate treatment. The *Wal-Mart* majority appears to have sided with the methodological critics. Justice Scalia demanded more than just statistical evi-

168. *Wal-Mart*, 131 S. Ct. at 2554 (emphasis added) (citation omitted).

169. Green, *supra* note 2, at 408, 410; see also Selmi, *supra* note 4, at 503 (discussing Judge Ikuta's dissent in the Ninth Circuit's en banc decision).

170. *Hazelwood*, 433 U.S. at 307–08 (“Where gross statistical disparities can be shown, they alone may in a proper case constitute prima facie proof of a pattern or practice of discrimination.”).

171. Civil Rights Act of 1964, Pub. L. No. 88-352, 42 U.S.C. § 2000e-2(1)(A)(i) (2012); see also Michael Selmi, *Theorizing Systemic Disparate Treatment Law 5* (2011) (unpublished draft of Selmi, *supra* note 4) (on file with author) (“In a disparate impact claim, it is generally possible to pinpoint a decision—to locate an agent or an explicit policy that is at the center of the allegations, and it is mostly a specific act.”).

172. Green, *supra* note 2, at 397 (calling the majority's policy-required view a “drastic reshaping of systemic disparate treatment law”); Selmi, *supra* note 4, at 503 (“This is why many, like Judge Ikuta of the Ninth Circuit and the late scholar Richard Nagareda, seek a policy or practice as proof of discrimination, although a formal policy is not required and if there was one it should be adjudicated under other causes of action.” (footnote omitted)).

dence of a disparity. He required some “glue” to bind the class members’ claims together—evidence of a policy of discrimination, a common rogue decision-maker, or perhaps anecdotal evidence on a much larger scale.¹⁷³ As demonstrated below, the majority’s demand for some “glue” binding together the class members’ claims reflects a change in the Court’s hidden priors from *Teamsters* to *Wal-Mart*.

III. THE BAYESIAN PATH FORWARD

*“Bayesian theory is a way of systematizing the elementary point that preconceptions play a role in rational thought.”*¹⁷⁴

Can systemic disparate treatment law survive *Wal-Mart*? If, as Professor Green argues, *Wal-Mart* represents a substantive shift to a new, policy-required view of systemic disparate treatment law, then the proof framework elucidated in *Teamsters* and *Hazelwood* is no more. This Part argues, however, that there is a viable path forward for systemic disparate treatment law that explains the outcome in *Wal-Mart* and that can also successfully reconcile some of the key concepts advanced by methodologists and contextualists. That path will require courts to embrace a Bayesian view of statistical evidence that acknowledges the importance of preconceptions.

A. Bayesian Inference

Recall the simplified base rate examples set forth in Part II.A. In the base rate examples, we see that when identical statistical evidence is viewed in the context of two different base rate assumptions, we arrive at very different likelihoods that the defendant discriminated.¹⁷⁵ This outcome is explained by the operation of Bayes’ Theorem. Bayes’ Theorem describes in mathematical terms how a decision-maker can rationally process new information—here, an observed disparity in employment outcomes—by combining it with the decision-maker’s prior knowledge or belief—here, the assumed base rate of discrimination—and use the end result (usually called a “posterior probability”) to make a decision under uncertainty.¹⁷⁶ In a Bayesian model of discrimination analysis the

173. *Wal-Mart*, 131 S. Ct. at 2552.

174. POSNER, *supra* note 10, at 67. Judge Posner later substitutes the term “Bayesian priors” for the term “preconceptions,” given the pejorative connotation associated with the word “preconception.” *Id.*

175. See *supra* notes 88–92 and accompanying text.

176. How a judge or jury processes information received in the form of pleadings, testimonial evidence, documentary evidence, or statistical evidence is the subject of a continuing debate in evidence literature. See, e.g., Ronald J. Allen, *A Reconceptualization of Civil Trials*, in *PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE: THE USES AND LIMITS OF BAYESIANISM* 21, 43 (Peter Tillers & Eric D. Green eds., 1988) (arguing that juries engage in a “relative plausibility” analysis, in which they choose from competing, but fully specified, alternative stories by deciding which story best explains the evidence). Some scholars suggest that a relative plausibility model is not necessarily inconsistent with Bayesian processing. Peter Tillers, *Trial By Mathematics—Reconsidered*, 10 *LAW, PROBABILITY & RISK* 167, 170 (2011) (“[I]t was a mistake [for critics of trial by mathematics]

decision-maker begins by assigning some prior probability distribution to the question at issue and then updates that probability distribution with observed information about employment outcome disparities.

The calculations are generally more complex in Bayesian statistical analysis than in the traditional analyses described above, but several authors have provided examples in the discrimination context.¹⁷⁷ Professor Ramona Paetzold succinctly illustrates Bayesian inference with the following example. Imagine an employer makes twenty new hires, six of whom were women and fourteen of whom were men, even though the employer drew from a pool of qualified applicants that contained 52% women and 48% men.¹⁷⁸ The question is: should an inference of discrimination arise from this evidence outcome disparity? The answer likely depends on the decision-maker's priors, or preconceptions about the likelihood that the employer discriminated, before viewing the observed statistical data.

A Bayesian analyst begins by assigning a prior probability distribution.¹⁷⁹ In Professor Paetzold's example, she first assumes that the decision-maker assigns a 0.5 probability to the proposition that the hypothetical employer did not discriminate at all in making hiring decisions (i.e., the true chance that any particular hire would be female from the qualified applicant pool was exactly equal to 0.52) and that the remaining 0.5 probability is spread uniformly over all the remaining possible values (other than 0.52) for the chances that any particular hire would be female.¹⁸⁰ Applying Bayes' Theorem to update the prior in light of the observed disparity, the decision-maker in this case would arrive at a posterior probability of 0.357 that the employer did not discriminate.¹⁸¹

The posterior probability in the foregoing example might still be sufficient to support an inference of discrimination, but it is sensitive to changes in the prior. If the assigned prior probability distribution is changed—either by building in a higher or lower probability that the employer did not discriminate or by changing how the remaining probability is distributed among the possible values for the true chance that any particular hire would be female—then the posterior outcome changes. Where the observed disparity in employment outcomes is not over-

to suppose that storytelling is inconsistent with Bayesian or mathematical analysis of evidence with cardinal numbers . . ." (footnote omitted).

177. See, e.g., Paetzold, *supra* note 56, at 406-07; David Kaye, *Statistical Evidence of Discrimination*, 77 J. AM. STAT. ASS'N 773, 779 (1982).

178. Paetzold, *supra* note 56, at 406-07.

179. *Id.*

180. *Id.* at 407.

181. *Id.* at 407 n.59 (providing the Bayesian calculation). Based on the assumed prior distribution, combined with the observed data reflecting an employment outcome disparity, the decision-maker would estimate that there is approximately a 64% chance that the employer discriminated in the hiring process, and approximately a 36% chance that the employer did not discriminate. *Id.* at 407.

whelming, the effects of varying the assigned prior can be dramatic. As Professor Paetzold notes, it is possible for traditional statistical analysis and Bayesian analysis of the same data to reach diametrically opposite conclusions about whether an employer likely discriminated.¹⁸² A traditional statistical analysis of a given data set might yield a statistically significant result with a *p*-value of .05, while a Bayesian analysis of that very same data might (depending on priors) yield a posterior probability that the employer did *not* discriminate of 0.95.¹⁸³

As the foregoing illustrates, estimates of priors can make the difference between deciding that a defendant more likely than not engaged in systemic disparate treatment and deciding *exactly the opposite*, even where the parties offer the exact same statistical evidence. Recognition of the limitations on indirect statistical inference, the base rate problem, and the role of prior probability estimates is critical to understanding the evolution of systemic disparate treatment theory from *Teamsters* and *Hazelwood* through *Wal-Mart*. The future of systemic disparate treatment law depends on courts and scholars recognizing this fundamental problem and developing a theoretically sound way of managing it. To begin this task first requires a brief introduction to a long-running philosophical debate in the field of statistics.

B. The Bayesian/Frequentist Divide

The problem of assigning a prior probability distribution, which will necessarily be at least somewhat subjective and possibly deeply uncertain, is the primary criticism of the Bayesian—or subjectivist—philosophy of statistical analysis.¹⁸⁴ Another school of statistical thought, frequentism—or objectivism—represents the more traditional approach to statistical inference.¹⁸⁵ A full discussion of the contours of the frequentist–Bayesian divide is beyond the scope of this Article; however, a short introduction to this philosophical debate is in order before examin-

182. *Id.* at 396–97 (citing Dennis V. Lindley, *A Statistical Paradox*, 44 *BIOMETRIKA* 187 (1957)).

183. *Id.* at 408.

184. See PAETZOLD & WILLBORN, *supra* note 57, § 12.05 (noting that traditional frequentist methods are often thought to be more objective than Bayesian methods because “they appear to rely on the sample evidence alone, without the inclusion of prior probabilities”); see also MICHAEL J. ZIMMER ET AL., *EMPLOYMENT DISCRIMINATION: SELECTED CASES AND STATUTES*—2011 47 (Supp. 2011); Mikel Aickin, *Issues and Methods in Discrimination Statistics*, in *STATISTICAL METHODS IN DISCRIMINATION LITIGATION* 159, 163 (D. H. Kaye & Mikel Aickin eds., 1986) (“Satisfactory rules for formulating such ‘prior’ probabilities are not well-developed, and it is far from clear that people do, or should, formulate and use them in the ‘Bayesian’ fashion stipulated by subjectivists.”).

185. See PAETZOLD & WILLBORN, *supra* note 57, § 12.01 (describing the “traditional school of thought in statistics” as “based on frequentist notions of probability”); Aickin, *supra* note 183, at 161 (“Speaking very roughly, most statisticians can be categorized with regard to their attitudes towards probability as being either *frequentists* or *subjectivists*.”); Rory Bahadur, *The Scientific Impossibility of Plausibility*, 90 *NEB. L. REV.* 435, 457 (2011) (“The two main philosophies of probability theory are the Frequentist and Bayesian models of probability.”).

ing how courts can manage the role of Bayesian priors in systemic disparate treatment cases.

The frequentist view eschews the notion of specifying prior probabilities altogether.¹⁸⁶ Instead, the frequentist looks at the statistical data merely as evidence in favor of rejecting the null hypothesis without attaching any particular probability to the likelihood that the employer *in this particular case* actually discriminated when it made its hiring decisions. A frequentist finding a gender disparity statistically significant at the .05 level might believe the statistical evidence is sufficient reason to reject the null hypothesis that employers hire from the labor pool at random with respect to gender.¹⁸⁷ But the frequentist does not form any specific opinion about the likelihood that *this* employer, during *this* particular set of hiring decisions, acted discriminatorily. Instead, the frequentist rejects the null hypothesis using the following *predictive* logic:

[T]o say that an event [here, random hiring leading to the observed disparity by chance] has probability 0.05 is not so much a statement about any *particular* occurrence or nonoccurrence of the event, but is rather a *prediction* about future repetitions of the setting in which the event might occur.¹⁸⁸

In other words, the frequentist imagines an infinite repetition of the set of hiring decisions and determines the frequency with which random hiring would lead to the observed disparate result. This frequentist view of statistics appears, at least initially, to be more objective than the Bayesian view because it does not require the decision-maker to specify and employ a subjective, uncertain, or imprecise prior probability.

The frequentist critique of Bayesianism holds force for settings where experiments can be repeated many times, especially when using experimental conditions rather than observational data, but the critique is less potent in the context of judicial decision-making. A judge or jury must make a ruling based on the evidence presented in the single case before it, and does not have the luxury of repeating a hypothetical hiring process or a controlled hiring experiment many times over to observe a multitude of outcomes.¹⁸⁹ Thus, some scholars, including Professor Rory

186. See *infra* text accompanying notes 195–96.

187. See Paetzold, *supra* note 56, at 399–400.

188. Aickin, *supra* note 184, at 162 (second emphasis added); see also KAYE ET AL., *supra* note 78, § 12.8.2.b. (“With the frequency interpretation of chance, there is no meaningful way to assign a numerical probability to the null hypothesis.”); Paetzold, *supra* note 56, at 401 (“It should be noted that this traditional method of testing is referred to as the ‘frequentist’ method of testing because the p-value represents a ‘long-run frequency’ interpretation of probability. . . . In order for the probability to be accurately interpreted, it must be possible to conceive of an infinite number of relevant, nearly identical hiring decisions facing the employer.”).

189. Peter Donnelly & Richard D. Friedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 MICH. L. REV. 931, 969 (1999) (“[T]he classical statistical method seeks to ascertain recurrent patterns, using observations about occurrences in the past to predict

Bahadur and Professor Ramona Paetzold, have argued that Bayesian reasoning is the better way to view statistical proof in the courtroom. Professor Bahadur argues:

Frequentist models are ill-suited for use in the legal system because they involve computing probability in idealized, non-real-world contexts, and they are incapable of incorporating preexisting information into the decision process. . . . Bayesian probability analysis, by contrast, is an approach to statistics which “formally seeks to utilize prior information.” . . . Bayesian probability, like legal inference, essentially seeks to process information in a manner that yields the most optimal inferences based on the information.¹⁹⁰

Likewise, Professor Paetzold has argued for a switch from frequentism to Bayesianism in the specific context of employment discrimination cases.¹⁹¹ She notes that the information provided by frequentist statistical methods, including testing the null hypothesis of random hiring decisions, is so “at odds” with the purpose of a judicial proceeding that it leads to confusion and error, including the transposition fallacy.¹⁹² Professor Paetzold writes:

In order for the . . . [*p*-value from frequentist analysis] to be accurately interpreted, it must be possible to conceive of an infinite number of relevant, nearly identical hiring decisions facing the employer. In other words, the frequentist approach requires consideration of hypothetical evidence (i.e., hypothetical hiring decisions) that were not actually a part of the employer’s past hiring practices. The correct interpretation of *p*-values is sufficiently at odds with the factfinder’s needs that courts often have difficulty in interpreting the statistical evidence.¹⁹³

Although many statisticians can generally be categorized as frequentist or Bayesian in their philosophies, some take a more nuanced view by using “whichever approach seems to be appropriate to the nature of the problem at hand.”¹⁹⁴ The Bayesian approach seems especially apt in systemic disparate treatment cases in light of the base rate and causal inference problems highlighted by the methodological critics. Further, the frequentist disdain for using subjective prior probabilities is misleading in the context of judicial decision-making. Making liability decisions about an individual employer’s set of employment actions based on a frequentist rejection of the null hypothesis necessarily carries with it the

the future. . . . Adjudication, by contrast, usually seeks to determine the truth about a particular dispute, which usually concerns events in the past.”)

190. Bahadur, *supra* note 185, at 457–58 (footnotes omitted).

191. Paetzold, *supra* note 56, at 412 (“There are strong reasons for switching to Bayesian methods of statistical inference in employment discrimination cases.”).

192. *Id.* at 401.

193. *Id.* (footnote omitted).

194. Aickin, *supra* note 184, at 162.

use of an implied, hidden prior probability. The following subparts illustrate the hidden role of Bayesian priors in judicial decision-making in systemic disparate treatment cases, including *Teamsters*, *Hazelwood*, and *Wal-Mart*.

C. Hidden Priors in Judicial Use of Frequentist Statistics

Frequentists decry Bayesians' reliance on unknowable, subjective prior probabilities.¹⁹⁵ But when a decision-maker draws an inference of discrimination in a particular systemic disparate treatment case based on a finding of statistical significance at the .05 level, the frequentist encourages the use of a built-in, unstated, and unexamined prior probability. Courts and juries relying on frequentist statistical significance to shift the burden of proof or determine liability use built-in priors without realizing it. As Professors Paetzold and Willborn explain:

Although traditional methods are often viewed as more objective in the sense that they appear to rely on the sample evidence alone, without the inclusion of prior probabilities, virtually any inference that a frequentist makes implies a corresponding Bayesian inference with respect to some assignment of prior probabilities. In other words, the traditional approach can be viewed as operating with a built-in set of prior beliefs. The fact-finder in a discrimination case would arguably be better served by examining explicitly a wide range of prior beliefs, for it is only then that the fragility or robustness of the inference about discrimination can be assessed.¹⁹⁶

This point is illustrated by revisiting one of the simplified examples from Part II.A.3. Assume that a systemic disparate treatment case is brought against Employer A. In Phase I of the *Teamsters* framework, the plaintiffs present statistical evidence in the form of a binomial distribution analysis of Employer A's hiring patterns. The frequentist statistician expert testifies that the *p*-value for the observed disparity in Employer A's hiring data is exactly .05, making the results just statistically significant at the .05 level. Recall that the correct interpretation of this statistical finding is as follows: Assuming that Employer A hired randomly from the relevant labor pool, we would expect to see results showing this level of disparity only 5 times in every 100—or 1 in 20—repetitions of the hiring process. Assume that no other evidence is offered by either plaintiffs or Employer A. From this statistical evidence alone, the fact-finder determines in Phase I that Employer A *more likely than not* engaged in an unlawful pattern or practice of discrimination, and Employer A is therefore liable for a systemic violation of Title VII. Such a finding

195. See, e.g., Samuel R. Lucas, *The Road to Hell . . . : The Statistics Proposal as the Final Solution to the Sovereign's Human Rights Question*, 30 WIS. INT'L L.J. 259, 319–20 (2012).

196. PAETZOLD & WILLBORN, *supra* note 57, § 12.05 (footnotes omitted). As Paetzold and Willborn note, this is a point that commentators often miss when dismissing Bayesian methods. *Id.* § 12.05 n.2.

appears to be justified, and possibly even prescribed, by the Court's opinion in *Hazelwood*. Per *Teamsters*, this finding in Phase I has the effect of establishing Employer A's wrongdoing, justifying prospective relief, and shifting the burden of proof to Employer A for Phase II, in which individual relief for the class members will be considered.

By drawing this inference of systemic discrimination and imposing liability at Phase I from only the frequentist statistical proof presented, the fact-finder has actually applied a built-in Bayesian prior probability without realizing or acknowledging as much. The fact-finder in this simplified example has implicitly assumed that the prior probability of discrimination must be greater than 5%. At a base rate of exactly 5% discrimination, with an observed statistical disparity having a p -value of .05, one would expect exactly 5 true positives and 5 false positives when performing the statistical test on 100 employers. A fact-finder considering this statistical evidence under a 5% base rate assumption would therefore believe the evidence to be precisely in equipoise as to whether Employer A engaged in an unlawful pattern or practice of discrimination. It is exactly as likely that the defendant before the fact-finder is one of the true positives as it is one of the false positives. Given that plaintiffs bear the initial burden of establishing by a preponderance of the evidence that the defendant engaged in a pattern or practice of unlawful discrimination,¹⁹⁷ the fact-finder should find in favor of Employer A if it believed the base rate to be exactly 5% or less.¹⁹⁸ But if the fact-finder assumed any base rate of discrimination higher than 5%, it would find that plaintiffs succeeded in showing that it is more likely than not that Employer A engaged in unlawful discrimination.

By determining that Employer A engaged in systemic discrimination based on only the statistical evidence with a p -value of .05, the fact-finder in this simplified example has implicitly assumed a base rate of discrimination exceeding 5%. It may well be that assuming a base rate of systemic discrimination higher than 5% is justified, but this point is never considered by the litigants, the statistical experts, the court, or the fact-finder. Instead, the implied prior probability assumption is obscured—hidden within the steps of indirect logical inferences drawn from the statistics. The frequentist interpretation of statistical probability, at least as it has been applied by courts in systemic discrimination cases, is no

197. *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 358 (1977).

198. This analysis would be complicated if the inquiry involved not just consideration of whether the employer more probably than not engaged in unlawful discrimination. I have previously argued elsewhere that because the systemic disparate treatment theory is fundamentally a burden allocation scheme, at least as to recovery for individual class members, the parties' relative access to information should also be considered in Phase I. Bent, *supra* note 16, at 818. The addition of other factors, such as the parties' information costs, would complicate the formula and change the critical cutoff number from > 0.5 (representing the preponderance of the evidence) to some higher or lower figure, but it would not change the fact that hidden priors will play a role in the probability component of the analysis.

less subjective than the Bayesian approach; it just appears that way on the surface. Such hidden priors have influenced the development of systemic disparate treatment law beginning with *Teamsters* and *Hazelwood* and continuing through *Wal-Mart*.

D. Hidden Priors in Teamsters and Hazelwood: The Foundational Flaw

The systemic disparate treatment theory of unlawful discrimination was first clearly articulated by the Supreme Court in 1977 in *Teamsters* and *Hazelwood*.¹⁹⁹ Title VII of the Civil Rights Act of 1964 was still a relatively new development in the employment landscape, and intentional, open discrimination against minorities and women was still prevalent.²⁰⁰ Professor Selmi, in his contribution to the Working Group, notes the importance of the era in which these seminal cases were decided:

[I]n the 1970s, it was relatively easy to draw inferences of discrimination based on statistics, even relatively crude statistics as were offered in *Teamsters* and *Hazelwood*. In these early cases, there was not much of a need to explain the source of the statistical disparity given that employers routinely discriminated against African Americans and women prior to the passage of the 1964 Act, and those habits appeared to die quite slowly.²⁰¹

In other words, the possibility of intentional discrimination on the basis of race or gender was an ever-present and obvious potential explanation for observed disparities in hiring, promotion, or other employment practices. The role of statistics in the early cases, Professor Selmi argues, was to demonstrate that chance was not a very likely explanation for the observed disparity.²⁰²

Of course, as explained above, the statistics do not *eliminate* the possibility that observed disparities were the product of chance. The statistics do not even necessarily make it more likely than not that the observed disparities were caused by chance. That will depend on the level of statistical significance (or the *p*-value) obtained in the statistical analysis, combined with estimated background rates or priors.

Teamsters and *Hazelwood* were built upon hidden priors. Professor Selmi is surely correct when he notes: "It was not just the companies' own history of discrimination that allowed for an inference of discrimination but it was also *the prevalence of discrimination at the time*."²⁰³ As he puts it, the foundational cases "were of a particular era" and "social conditions have surely changed."²⁰⁴

199. See *supra* Part I.

200. Selmi, *supra* note 4, at 485–86.

201. *Id.*

202. *Id.*

203. *Id.* at 486 (emphasis added).

204. *Id.* at 487.

This observation reveals the importance of prior probabilities operating beneath the surface in the foundational cases. The prevalence of discrimination at the time *Teamsters* and *Hazelwood* were decided generally influenced estimates or preconceptions about how likely it was that any given employer was engaged in unlawful discrimination, before any consideration of other evidence in the case. Statistical showings in cases like *Hazelwood* had to be considered in the context, and against the backdrop, of those prior beliefs about the background rates, or prevalence, of employment discrimination. As Professor Selmi points out, once a statistically significant disparity was shown, and if an employer could not convince the court of some alternative explanation for the observed skew in hiring numbers, the Court was prepared to accept the next most obvious explanation at that time: intentional disparate treatment.²⁰⁵

Unsurprisingly, the Court did not openly discuss prior probabilities or base rates of discrimination in *Teamsters* or *Hazelwood*. In both cases, the Court appears to begin with an unstated assumption that the base rate of intentional discrimination was relatively high; hence, discrimination becomes the obvious explanation for observing statistically significant hiring disparities.²⁰⁶ The very structure that the *Teamsters* opinion prescribes for analyzing systemic disparate treatment cases is underpinned by an unstated prior belief that sufficient baseline levels of discrimination existed to make discrimination a reasonable explanation in the face of statistically significant disparities. In *Hazelwood*, the Court reinforced that notion, reflecting a similar, unstated prior baseline belief.²⁰⁷ The theoretical and statistical foundations of systemic disparate treatment law were flawed from the beginning because of the Court's failure to recognize the hidden role of priors. Unstated prior probabilities have been lying beneath the surface of systemic disparate treatment law since 1977.

E. Hidden Priors in Wal-Mart: How Changing Priors Changed Substantive Doctrine

Professor Suzanna Sherry argues that shifts in the Supreme Court's "intuitions about how the world works," or to use her term, "foundational facts," can result in shifting legal doctrine even where the Court denies that a doctrinal shift has occurred.²⁰⁸ This sort of shift in foundational facts is precisely what occurred in *Wal-Mart*. The Court's changing intuitions about the background likelihood of discrimination led it to apply systemic disparate treatment doctrine in ways that threaten to undercut *Teamsters* and *Hazelwood* entirely.

205. See *id.* at 485–87.

206. See *supra* text accompanying note 202.

207. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307–09 (1977).

208. Suzanna Sherry, *Foundational Facts and Doctrinal Change*, 2011 U. ILL. L. REV. 145, 146–47 (2011).

Why did the *Wal-Mart* Court demand more of plaintiffs than it did in *Teamsters* or *Hazelwood* for an inference of systemic discrimination?²⁰⁹ Professor Selmi contends that discrimination has become a less obvious explanation for observed statistical disparities than it was in the 1970s and much of the 1980s.²¹⁰ He notes:

[O]nce we moved farther away from the era of plain and open exclusionary policies, it has become less clear, at least to the courts, that discrimination provides the best explanation for the observed disparities. Importantly, while social conditions have surely changed, the theory underlying the pattern or practice cause of action has not, and indeed, the mid-1970s cases of *Teamsters* and *Hazelwood* remain surprisingly relevant today.²¹¹

Neither Professor Selmi nor the Court refers expressly to changes in Bayesian prior probabilities—no doubt because the Bayesian view of statistical inference has not been widely accepted in employment discrimination law²¹² and because the hidden priors are buried beneath frequentist inferential reasoning.²¹³ But Professor Selmi's basic point can be recast in Bayesian terms: the prior probability of discrimination, or the base rate against which the observed statistical disparity should be judged, has significantly declined since the 1970s. The estimated background likelihood that any particular employer is engaged in systemic discrimination has declined enough to make it "less clear" that "discrimination [always] provides the best explanation for the observed disparities."²¹⁴

Glimpses of changed priors can be observed by comparing language in *Wal-Mart* to language found in the early cases. As Professor Deborah Weiss notes, Justice Scalia's ruminations about what "most managers in any corporation" would do are particularly revealing.²¹⁵ One can certainly question the empirical accuracy of Justice Scalia's assertion that most managers, if left to their own devices, would choose "sex-neutral, performance-based criteria" to evaluate employees,²¹⁶ but there can be little doubt that the same statement would not likely have appeared in a Su-

209. On this point, it may be argued that *Wal-Mart* did not speak directly to the substance of systemic disparate treatment law, but only to the procedural question involving Rule 23. But given that most systemic disparate treatment claims (other than those filed by the EEOC) are brought as class actions under Rule 23, and that the merits inquiry often merges with the certification inquiry, it may make little difference. See Selmi, *supra* note 4, at 480, 493.

210. *Id.* at 487.

211. *Id.*

212. See PAETZOLD & WILLBORN, *supra* note 57, § 12.01 ("One increasingly popular school of statistical inference is the Bayesian school. Although the school is well-known among statisticians and some social scientists, it is only marginally recognized at law.")

213. See *supra* Part II.B.

214. See Selmi, *supra* note 4, at 487.

215. Weiss, *supra* note 9, at 1687, 1741; see also *supra* text accompanying note 172 (quoting *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2554 (2011)).

216. See *supra* text accompanying note 172 (quoting *Wal-Mart*, 131 S. Ct. at 2554).

preme Court opinion in 1977. Whether Scalia's unsupported observation is accurate or not, it is reasonable to presume that it is at least *closer* to a true statement about *intentional* discrimination now than it would have been if the same words had been written in 1977. Justice Scalia's unsupported assertion reflects his intuition about how the world works, and his intuition is that discrimination is not common.²¹⁷

Further evidence of changed priors can be found in the importance that the majority places on Walmart's official policy of non-discrimination, as compared to its treatment of a similar School District policy in *Hazelwood*. The *Wal-Mart* majority noted: "Wal-Mart's announced policy forbids sex discrimination, and as the District Court recognized the company imposes penalties for denials of equal employment opportunity."²¹⁸ As described above, the *Wal-Mart* majority believed this official company policy prohibiting discrimination undermined the commonality requirement because individual managers would not likely have discriminated in the face of the official policy.²¹⁹

Contrast this to *Hazelwood*. There, the Hazelwood School District had an "officially promulgated policy 'to hire all teachers on the basis of training, preparation and recommendations, regardless of race, color or creed.'"²²⁰ Yet, this evidence did not appear to affect the Court's assessment of the requirements for a prima facie showing of systemic disparate treatment in the slightest. After mentioning the existence of this policy, the Court focused exclusively on the relative merits of the statistical showing.²²¹ The Court devoted substantial attention to identifying the proper comparison labor pool—or reference class—for purposes of conducting a binomial distribution²²² and to considering whether hiring discrimination taking place before Title VII prohibited public employers from discriminating might have affected observed disparities.²²³ But nowhere did the *Hazelwood* Court suggest that the government would need to identify particular rogue decision-makers who allegedly acted discriminatorily in contravention of the official non-discrimination policy.²²⁴ Rather, the Court made clear that statistical evidence of gross disparity alone may "in a proper case" constitute a prima facie showing of a pattern or practice of discrimination—apparently in spite of an express non-discrimination policy.²²⁵

217. See Weiss, *supra* note 9, at 1687.

218. *Wal-Mart*, 131 S. Ct. at 2553 (citation omitted).

219. *Id.* at 2554.

220. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 303–04 (1977).

221. *Id.* at 306–12.

222. *Id.* at 308, 311–12.

223. *Id.* at 309, 309 n.15.

224. See *id.* at 306–13.

225. *Id.* at 307–08.

Finally, the Court in *Wal-Mart* played up the importance of anecdotal evidence of discrimination in systemic discrimination cases. While *Hazelwood* indicated that statistical evidence alone could in a proper case establish a prima facie case of systemic discrimination,²²⁶ the *Wal-Mart* ruling appears to call that statement into question. Justice Scalia points out the weaknesses of the anecdotal evidence offered by plaintiffs:

In *Teamsters v. United States*, in addition to substantial statistical evidence of company-wide discrimination, the Government (as plaintiff) produced about 40 specific accounts of racial discrimination from particular individuals. That number was significant because the company involved had only 6,472 employees, of whom 571 were minorities, and the class itself consisted of around 334 persons[.] The 40 anecdotes thus represented roughly one account for every eight members of the class. . . . Here, by contrast, respondents filed some 120 affidavits reporting experiences of discrimination—about 1 for every 12,500 class members—relating to only some 235 out of Wal-Mart's 3,400 stores. . . . Even if every single one of these accounts is true, that would not demonstrate that the entire company "operate[s] under a general policy of discrimination."²²⁷

Traditionally, a prima facie case of systemic disparate treatment has "relied almost entirely on statistics."²²⁸ According to Professor Selmi, the "anecdotal evidence is always of marginal significance in a pattern or practice claim."²²⁹ Although the *Teamsters* Court highlighted the anecdotal evidence that "brought the cold numbers convincingly to life,"²³⁰ that anecdotal evidence did not appear to be dispositive and the Court in *Hazelwood* clarified that statistical evidence alone could be sufficient.²³¹ One potential reading of *Wal-Mart* is that it has changed the substantive law of systemic disparate treatment such that some "significant" or sufficient level of anecdotal evidence is now required to explain what is going on in the observed statistical disparity.²³²

The Court now seems less willing to infer systemic discrimination on the basis of only statistical evidence of an observed company-wide, statistically significant disparity in hiring, promotions, or pay, without something more—the identification of a particular policy, mechanism, or rogue decision-maker that produces the observed disparity. The Court's

226. *Id.*

227. *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2556 (2011) (alteration in original) (citations omitted) (quoting *Gen. Tel. Co. of Sw. v. Falcon*, 457 U.S. 147, 159 n.15 (1982)). Justice Scalia also notes that in *Teamsters* the forty anecdotes came from employees "spread throughout" the company. *Id.* (internal quotation marks omitted).

228. Selmi, *supra* note 4, at 485.

229. *Id.* at 501.

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

231. *Hazelwood*, 433 U.S. at 307–08; *see also supra* note 43 (listing lower court decisions stating that statistical evidence alone may be sufficient to establish a prima facie case of systemic disparate treatment discrimination).

232. *See Wal-Mart*, 131 S. Ct. at 2556.

intuitions about the way the world works—or its hidden Bayesian priors—appear to have changed.²³³ Statistically significant evidence suggesting a rejection of the null hypothesis alone was not enough for the *Wal-Mart* Court to endorse the inferential leap to discrimination as the next most obvious explanation for observed disparities.²³⁴ Instead, the Court looked harder for other potential explanations for observed statistically significant disparities.²³⁵ After *Wal-Mart*, the Court may demand evidence of a policy of discrimination, or at least some story or narrative to explain the data, as Professor Selmi suggests. Alternatively, it might demand some additional pieces of evidence such as a higher volume of anecdotal testimony to nudge the probability of discrimination over the line.

The *Wal-Mart* Court's approach is generally consistent with, and can be explained by, a Bayesian view of systemic disparate treatment law with two major caveats: (1) the *Wal-Mart* Court did not openly acknowledge the role of prior probabilities, and (2) the *Wal-Mart* Court used its own unstated priors, but disregarded social framework evidence that would have helped a fact-finder form or adjust its own priors. The majority's unsupported assertion about what most managers would do was not recognized as a challengeable, and empirically testable, assumption about the non-case-specific background likelihood of discrimination.²³⁶ Yet, the majority cast aside expert testimony on social frameworks that would have provided valuable information about the background likelihood of discrimination in certain organizational settings precisely because the expert was not able to offer case-specific information.²³⁷

F. Reconciling the Methodological and Contextualist Strands by Exposing Hidden Priors

As the preceding discussion shows, Bayesian priors are at work in systemic disparate treatment cases whether courts choose to

233. The idea that the Court's prior estimate of the base rate of discrimination would change over the course of more than thirty years is generally consistent with statements the Court has made regarding changing societal attitudes toward racial classifications. See *Grutter v. Bollinger*, 539 U.S. 306, 343 (2003) ("We expect that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today."); see also Sherry, *supra* note 208, at 164–65. ("It is reasonable to suppose, however, that the passage of time might lessen the likelihood that employers are deliberately discriminating, especially in the context of disparate impact. By the late 1980s, was it still more likely than not that any employer who adopted an employment practice with a disparate impact had a covert discriminatory intent? The Court apparently thought not."). Professor Sherry highlights changes in disparate impact doctrine and individual disparate treatment burden-shifting doctrine that she contends are attributable to the Court "chang[ing] its mind about the overall prevalence of racially discriminatory motives among American employers." *Id.* at 166.

234. See *supra* text accompanying notes 169–73.

235. See *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2554–55 (2011).

236. Weiss, *supra* note 9, at 1687.

237. *Id.* at 1678.

acknowledge them or not.²³⁸ A Bayesian view of systemic disparate treatment would expose the role of hidden priors, rebuild the flawed theoretical and statistical foundations of the doctrine, and bring coherence to the most misunderstood category of antidiscrimination claims. Further, a Bayesian view also holds the potential to reconcile methodological and contextualist strands of legal scholarship described in Part II.

For critics writing in the methodological strand, use of Bayesian priors would prevent courts and experts from falling into the transposition fallacy trap and would focus fact-finders on the conditional probability in which they are truly interested—the probability that the employer unlawfully discriminated, given the observed disparities in employment outcomes. Professor Browne may remain skeptical of our judicial system's ability to generate usable estimates of prior probabilities and convey them in an understandable way, but these are questions about the source and management of priors. Part IV of this Article identifies some of these difficult practical questions and proposes a scholarly agenda for addressing them. For now, it is enough to note that the use of Bayesian statistics is not unheard-of in litigation, especially in paternity cases and DNA match criminal cases.²³⁹ Despite potential difficulties in implementation, the introduction of Bayesian logic in systemic disparate treatment cases is now overdue.

The Bayesian approach also provides conceptual space for the contextualist view advanced by Professor Green. The contextualist view posits that certain organizational or cultural dynamics operating at the entity-level can generate disparate employment outcomes, even though individual intentional wrongdoers cannot be identified.²⁴⁰ This is, essentially, an argument that society underperceives the extent and relative likelihood of employment discrimination because we cannot identify individual wrongdoers in all cases.²⁴¹ In formulating this argument, Professor Green draws on the work of social scientists studying organizational behavior.²⁴² Taking a Bayesian approach in the courtroom, those same social scientists could provide expert testimony about the ways in which organizational dynamics can and do lead to disparate treatment in employment outcomes, even where individual intentional wrongdoers are

238. *See id.* at 1678–79 (contending that the application of priors are unavoidable in discrimination law generally).

239. *See infra* Part IV.

240. *See supra* Part II.B.

241. *See Green, supra* note 2, at 433. Professor Green states:

As a practical matter, disparate treatment is often difficult to discern on an individual basis—it occurs subtly in day-to-day interactions, in decisions that do not lend easily to immediate comparison, and in unstated judgments and perceptions of value and skills—and therefore can frequently only be identified in the aggregate, where it can be shown that members of a particular group are being denied more promotions or provided less pay.

Id. (citation omitted).

242. *See supra* text accompanying note 138.

not observed. This expert testimony would help fact-finders estimate prior probabilities. For example, expert, non-case-specific social framework testimony about organizational causes of discrimination could appropriately adjust the fact-finder's estimates of prior probabilities.

One of the strengths of the Bayesian approach is its ability to incorporate newly-learned information in a rational and logically consistent way by updating probability estimates. This built-in flexibility in the Bayesian approach provides a distinct advantage over the status quo, with its more rigid decision rule that turns on a finding of frequentist statistical significance, and the accompanying doctrinal shifts that occur periodically as the Supreme Court's hidden priors change over time. If social scientists make additional breakthroughs in our understanding of the extent or causes of employment discrimination, that new information can be conveyed to the fact-finder in the courtroom (and, of course, subjected to challenge and rebuttal by the opposing party). Priors can be updated accordingly, and statistical evidence can be correctly interpreted in a way that avoids the transposition fallacy and is consistent with Bayes' Theorem. In this way, social framework evidence like that offered by Dr. William Bielby in the *Wal-Mart* case can and should be used to adjust the fact-finder's priors, even though it is not case-specific.²⁴³

The Bayesian view of systemic disparate treatment law provides a workable path forward that accommodates the contributions made by both the methodological critics and the contextualists. It also explains the doctrinal arc of systemic disparate treatment law from *Teamsters* to *Wal-Mart*. Changing priors have changed the substantive law. Acknowledging the influence of priors is the first step in developing a coherent theory of systemic disparate treatment law that accords statistical evidence of outcome disparities its appropriate weight, yet also incorporates our changing understandings of how discrimination operates in the workforce.

IV. A SCHOLARLY AGENDA FOR THE MANAGEMENT OF PRIORS

Acknowledgment of the importance of priors is only the first step in developing a unified theory of systemic disparate treatment law. A num-

243. Weiss, *supra* note 9, at 1679 ("Social framework evidence promises to go some way to replacing the personal views of judges and juries with a more objective perspective."). Social framework evidence can be relevant even though it is not case-specific, because it provides information about potential misperceptions of background rates of discrimination. The *Wal-Mart* majority dismissed the social framework testimony offered by plaintiffs' expert, Dr. Bielby, because he "could not . . . determine with any specificity how regularly stereotypes play a meaningful role in employment decisions at *Wal-Mart*." *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2553 (2011) (emphasis added) (internal quotation mark omitted). Once the role of Bayesian priors is exposed, however, it becomes clear that Dr. Bielby's testimony was relevant to a proper interpretation of the observed statistical disparities in employment outcomes, even though he could not offer case-specific testimony about the true causes of disparity at *Wal-Mart* specifically. See Weiss, *supra* note 9, at 1683–87 (noting the relevance, under a Bayesian view, of "pure social framework evidence").

ber of difficult second-order questions are immediately raised by the recognition that prior probabilities influence the interpretation of statistical evidence of outcome disparities. These questions include:

1. Whose priors matter?²⁴⁴ Possible answers include the trial judge, the trial fact-finder, appellate judges, and the legislature.
2. Relatedly, how should evaluation of priors fit into civil litigation pretrial procedure, including key dispositive procedures such as motions to dismiss and motions for summary judgment?
3. How can Bayesian statistical inference be presented to fact-finders at trial? Other scholars have discussed possible techniques for conveying Bayesian information to jurors, including the use of charts that provide different posterior probabilities for a number of selected possible prior probability distributions.²⁴⁵ Something similar has even been required by at least one court in the context of Bayesian statistical inference in the paternity testing context.²⁴⁶
4. What are legitimate or illegitimate sources of priors? Here, potential answers range from pure unsupported guesses to empirical evidence on background rates of discrimination to expert testimony about social framework causes of discrimination. Which sources are legitimate? Should some sources receive more deference than others?

In a related paper, I offer my views on these difficult second-order questions.²⁴⁷ Although these questions pose practical challenges to the use of Bayesian statistical inference in systemic disparate treatment litigation, they are not intractable. Courts already use Bayesian statistical inference regularly in some DNA matching and paternity testing cases.²⁴⁸ Some scholars and statisticians have long been optimistic that courts can adapt to the use of Bayesian statistics in discrimination cases.²⁴⁹

244. Professors Paetzold and Willborn identify this question, but do not attempt an answer. *See* PAETZOLD & WILLBORN, *supra* note 57, § 12.05 n.10 (“An important legal issue would involve whose prior probabilities should be represented. Because the Bayesian view of probability is subjective (represents an individual’s uncertainty), courts would need to decide whose uncertainty the prior distribution should represent.”).

245. *See, e.g.*, KAYE ET AL., *supra* note 78, § 14.3.1 (discussing the potential use of prior probability charts in DNA match case).

246. *Plemel v. Walter*, 735 P.2d 1209, 1219 (Or. 1987); *see also* D.H. Kaye, *Plemel as a Primer on Proving Paternity*, 24 WILLAMETTE L. REV. 867, 875–82 (1988).

247. Bent, *supra* note 14.

248. The following search in the ALLCASES Westlaw database returned 114 results: (BAYES! or “PRIOR PROBABILIT!” or “PRIOR ODDS”) & (“DNA MATCH” or “PATERNITY”). *See also* KAYE ET AL., *supra* note 78, §§ 14.3.1–14.4.3.

249. *See, e.g.*, Stephen E. Fienberg, *The Increasing Sophistication of Statistical Assessments as Evidence in Discrimination Litigation*, 77 J. AM. STAT. ASS’N. 784, 784 (1982) (“I believe that with

The existence of difficult second-order questions does not counsel in favor of ignoring prior probabilities. As demonstrated above, when decision-makers ignore prior probabilities and draw inferences to impose liability based on frequentist statistical significance they are, in effect, operating with a hidden, built-in, unexamined prior. The point of this paper is not to answer all of the difficult second-order questions that come with acknowledging priors but rather to get the discussion started. Once courts and scholars come to recognize the inescapable role of prior probabilities, they can turn their attention to the second-order questions and develop a more fully-formed system for managing priors in the courtroom.

CONCLUSION

Wal-Mart represents a substantive move away from the systemic disparate treatment theory as formulated in *Teamsters* and *Hazelwood*. While evidence of a statistically significant disparity was once sufficient to establish a prima facie case of systemic disparate treatment, the *Wal-Mart* Court demanded more. This Article shows that the Court's substantive doctrinal shift can be explained by recognizing the cracks in the statistical and theoretical foundations of systemic disparate treatment law tracing back to 1977. The foundational cases rested on an unstated and unexamined assumption about prior probabilities that a majority of the Court seemingly no longer holds. A change in hidden priors has led to a change in the law.

This Article challenges courts and scholars to openly acknowledge the importance of priors in evaluating systemic discrimination cases so that the discussion of the difficult challenges we face in managing priors can begin. The future shape of systemic disparate treatment law will depend on whether courts make the operation of priors more transparent. Acknowledging hidden priors will bring together the academic insights of the methodologists and the contextualists, clearing the way for consideration of the second-order issues of prior management and for the inclusion of social framework and other evidence of organizational causes of discrimination. *Wal-Mart* should be viewed not as the death knell for systemic disparate treatment law but rather as the instigator for a fundamental change in how statistical evidence of an outcome disparity is interpreted in antidiscrimination law.

