

# ***P*-Values, Priors, and Procedure in Antidiscrimination Law**

JASON R. BENT†

## INTRODUCTION

A recent, high-profile article in *Nature* questioned scientific journals' continued practice of publishing statistically significant findings of hypothesis testing studies.<sup>1</sup> The article was sharply critical of stubborn overreliance on *p*-values, and posited that “[m]ost scientists would look at” study results with a “*P* value of .01 and say that there was just a 1% chance of [the] result being a false alarm.”<sup>2</sup> Those scientists would be wrong.

Hopefully, this claim in the *Nature* article is overly pessimistic. One would expect that most research scientists have a sufficient understanding of statistics to avoid falling

---

† Associate Professor of Law, Stetson University College of Law; J.D., University of Michigan Law School (2000); B.A., Grinnell College (1997). This paper is the second in a series of two papers addressing the influence of priors in systemic disparate treatment discrimination law. The first paper, *Hidden Priors: Toward a Unifying Theory of Systemic Disparate Treatment Law*, 91 DENV. U. L. REV. 807 (forthcoming 2015), was selected for presentation and publication as part of a symposium hosted by the *Denver University Law Review* in January 2014, entitled *Revisiting Sex: Gender and Sex Discrimination Fifty Years After the Civil Rights Act*. The concepts advanced in this paper were refined as a result of comments received at the *Denver University Law Review* symposium, the 2012 Annual Meeting of the Southeastern Association of Law Schools, and the 2012 Annual Colloquium on Current Scholarship in Labor and Employment Law, and the suggestions of Charlotte Alexander, Kingsley Browne, Regina Nuzzo, Gowri Ramachandran, Charles A. Sullivan, Louis Virelli III, Steven Willborn, and Michael J. Zimmer. Both papers in this series benefited from the research assistance of Michael Labbee and Lindsay Houser. The faculty and administration of the Stetson University College of Law generously supported this research project.

1. See Regina Nuzzo, *Scientific Method: Statistical Errors*, 506 NATURE 151 (Feb. 12, 2014), available at [www.nature.com/news/scientific-method-statistical-errors-1.14700](http://www.nature.com/news/scientific-method-statistical-errors-1.14700).

2. *Id.* at 151.

into this trap, often called the transposition fallacy.<sup>3</sup> Then again, there is reason for pessimism. In a classic study of Harvard Medical School physicians and students, 45% of the respondents were fooled by the transposition fallacy when considering the following question:

One in a thousand people has a prevalence for a particular heart disease. There is a test to detect this disease. The test is 100% accurate for people who have the disease and is 95% accurate for those who don't (this means that 5% of the people who do not have the disease will be wrongly diagnosed as having it). If a randomly selected person tests positive what is the probability that the person actually has the disease?<sup>4</sup>

Almost half of the respondents incorrectly answered 95%, which strongly suggests that those respondents succumbed to the transposition fallacy.<sup>5</sup> The correct answer, when calculated using Bayes's Theorem to account for the

---

3. The common error identified in the *Nature* article is referred to by several alternate names, including the transposition fallacy, the fallacy of the transposed conditional, and the prosecutor's fallacy (in reference to cases involving DNA matching evidence). See, e.g., Stephen J. Ceci & Richard D. Friedman, *The Suggestibility of Children: Scientific Research and Legal Implications*, 86 CORNELL L. REV. 33, 52 (2000). This fallacy is discussed at length in Jason R. Bent, *Hidden Priors: Toward a Unifying Theory of Systemic Disparate Treatment Law*, 91 DENV. U. L. REV. 807 (forthcoming 2015) [hereinafter Bent, *Hidden Priors*]. It may be thought of as a specific instance of what is often called the base rate fallacy in the field of cognitive psychology, referring to the tendency of subjects to underweight the relevance of background knowledge. See generally Louis Kaplow, *Multistage Adjudication*, 126 HARV. L. REV. 1179, 1265 n.210 (2013) (citing, among others, Maya Bar-Hillel, *The Base-Rate Fallacy in Probability Judgments*, 44 ACTA PSYCHOLOGICA 211 (1980), and Daniel Kahneman & Amos Tversky, *On the Psychology of Prediction*, 80 PSYCHOL. REV. 237 (1973)).

4. Ward Casscells et al., *Interpretation by Physicians of Clinical Laboratory Results*, 299 NEW ENG. J. MED. 999, 999 (1978); see also Norman Fenton & Martin Neil, *Comparing Risks of Alternative Medical Diagnosis Using Bayesian Arguments*, 43 J. BIOMED. INFORMATICS 485, 485 (2010) [hereinafter Fenton & Neil, *Comparing Risks*]; Bailey Kuklin, *Probability Misestimates in Medical Care*, 59 ARK. L. REV. 527, 528-29 (2006); James S. Liebman et al., *The Evidence of Things Not Seen: Non-Matches as Evidence of Innocence*, 98 IOWA L. REV. 577, 675 (2013).

5. See Fenton & Neil, *Comparing Risks*, *supra* note 4, at 486 ("When people give a high answer, like 95%, they are falling victim to a very common fallacy known as the 'base-rate neglect' fallacy; people neglect to take into consideration the very low probability (of having the disease) that forms the vital starting point.").

base rate prevalence of the disease, is approximately 2%.<sup>6</sup> Only 18% of the Harvard Medical School study respondents correctly provided this answer.<sup>7</sup>

As the *Nature* article points out, statistically significant research results with  $p$ -values of .01 or .05 do not tell the whole story, because they ignore the critical factor of “how plausible the hypothesis is in the first place.”<sup>8</sup> Even if the *Nature* article is overly pessimistic on the proportion of scientists that would fall victim to the transposition fallacy, the article still raises very troubling questions for the legal profession: if scientists or doctors have difficulty properly interpreting  $p$ -values, what would most *judges* or *lawyers* say about the meaning of a  $p$ -value of .01?<sup>9</sup> How would most *jurors* interpret a  $p$ -value of .01?

These are crucial questions for cases involving statistical evidence, including systemic employment discrimination cases. A fundamental misunderstanding of the meaning of  $p$ -values and statistical significance has infected employment discrimination law since the late 1970s, when statistical evidence of employment disparities first came to be accepted

---

6. See, e.g., Casscells et al., *supra* note 4, at 999-1000; Fenton & Neil, *Comparing Risks*, *supra* note 4, at 485; Kuklin, *supra* note 4, at 528 n.1. This result can be easily shown with a visual representation of the problem, *infra* Part III, Question 3, sec. D, fig. A.

7. See, e.g., Casscells et al., *supra* note 4, at 1000; Kuklin, *supra* note 4, at 529 n.1.

8. Nuzzo, *supra* note 1, at 151.

9. In his review of relevant cases, Professor Kingsley Browne identified examples from “virtually all the circuits” involving the commission of the transposition fallacy, suggesting that many (probably most) judges and jurors would be susceptible to the error identified in the *Nature* article and illustrated in the Harvard Medical School study. See Kingsley R. Browne, *The Strangely Persistent “Transposition Fallacy”: Why “Statistically Significant” Evidence of Discrimination May Not Be Significant*, 14 LAB. LAW. 437, 447 (1998) [hereinafter Browne, *Strangely Persistent*]; Kingsley R. Browne, *Statistical Proof of Discrimination: Beyond “Damned Lies,”* 68 WASH. L. REV. 477, 490-93 (1993) [hereinafter Browne, *Statistical Proof*]. Lawyers as a group are notoriously bad with mathematics. Lisa Milot, *Illuminating Innumeracy*, 63 CASE W. RES. L. REV. 769, 769 (2013) (“It is an open secret that lawyers don’t like math. Tales of lawyers who chose the profession over business or medicine at least in part because of discomfort with math are legion, as are reports of math avoidance by lawyers once in the profession.”).

as a means of establishing a prima facie case of systemic employment discrimination by the Supreme Court.<sup>10</sup> This misunderstanding also infected the Supreme Court's most recent systemic employment discrimination decision in *Wal-Mart Stores, Inc. v. Dukes*.<sup>11</sup>

The Court's systemic disparate treatment doctrine condones the use of  $p$ -values to test for statistical significance without adequately considering "how plausible the hypothesis [of unlawful discrimination] is in the first place."<sup>12</sup> This misuse of traditional hypothesis testing and  $p$ -values has, for years, covered up the opaque operation of hidden, unstated, and unexamined "priors"—assumptions about the prevalence, or base rate, of unlawful discrimination.<sup>13</sup> It is time to bring the unstated assumptions into the open in systemic discrimination cases. Just as many scientists and statisticians have argued that the time has come for a Bayesian revolution in scientific research,<sup>14</sup> so too has the

---

10. See *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307-12 (1977); *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 339-41 (1977); Browne, *Statistical Proof*, *supra* note 9, at 482-83, 503-505; Browne, *Strangely Persistent*, *supra* note 9, at 437, 441-42; see also 1 CHARLES A. SULLIVAN ET AL., *EMPLOYMENT DISCRIMINATION: LAW AND PRACTICE* § 3.01 (3d ed. 2002).

11. 131 S. Ct. 2541 (2011).

12. See *supra* note 8 and accompanying text.

13. See Bent, *Hidden Priors*, *supra* note 3, at 840-47.

14. See, e.g., David Leonhardt, *Subconsciously, Athletes May Play Like Statisticians*, N.Y. TIMES, Jan. 20, 2004, at F1 ("In academia, the Bayesian revolution is on the verge of becoming the majority viewpoint, which would have been unthinkable ten years ago," said Bradley P. Carlin, a professor of public health at the University of Minnesota and a Bayesian specialist."); see also John K. Kruschke, *An Open Letter to Editors of Journals, Chairs of Departments, Directors of Funding Programs, Directors of Graduate Training, Reviewers of Grants and Manuscripts, Researchers, Teachers, and Students*, 2, 5, <http://www.indiana.edu/~kruschke/AnOpenLetter.pdf> (last accessed Oct. 20, 2014) ("The point is simple: Bayesian methods are being adopted across the disciplines of science. We should not be laggards in utilizing Bayesian methods in our science, or in teaching Bayesian methods in our classrooms. . . . Science is moving to Bayesian methods because of their many advantages, both practical and intellectual, over 20th century [null hypothesis significance testing]. It is time that we convert our research and educational practices to Bayesian data analysis."); see generally Mark A. Beaumont & Bruce Rannala, *The Bayesian Revolution in Genetics*, 5 NATURE REV. GENETICS 251 (2004); Stephen P. Brooks, *Bayesian Computation: A Statistical Revolution*, 361 PHIL. TRANSACTIONS OF THE

time finally come for a Bayesian revolution in employment discrimination law.

This is the second paper in a two-part series advancing a Bayesian approach to systemic disparate treatment law. The first paper, *Hidden Priors*,<sup>15</sup> argued that reliance on traditional hypothesis testing statistics in systemic disparate treatment cases does not remove the influence of hidden prior assumptions about the prevalence of discrimination, but instead just conceals the influence of unstated assumptions beneath a veneer of doctrinal reasoning, statistical significance tests, and *p*-values.<sup>16</sup> *Hidden Priors* advocated for the use of Bayesian statistical analysis to bring the influence of priors to light, and challenged “courts and scholars to openly acknowledge the importance of priors in evaluating systemic discrimination cases, so that discussion of the difficult challenges we face in managing priors can begin.”<sup>17</sup>

At the conclusion of *Hidden Priors*, I acknowledged those “difficult challenges” in managing priors that seem to leave courts and legal scholars reluctant to embrace Bayesian analysis in systemic discrimination cases. I then laid out an agenda for tackling those challenges by way of a scholarly discussion of several second-order questions:

1. Whose priors matter? 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?
4. What are legitimate or illegitimate sources of priors?

---

ROYAL SOC'Y OF LONDON, SERIES A: MATHEMATICAL, PHYSICAL & ENGINEERING SCI. 2681 (2003).

15. See Bent, *Hidden Priors*, *supra* note 3, at 840-42.

16. *Id.* at 810.

17. *Id.* at 851.

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?<sup>18</sup>

This second paper offers my views on the best available—though admittedly imperfect—answers to these questions. I contend that our answers to these questions should be guided by the institutional strengths of the various entities involved in our civil litigation system. Legislatures, trial judges, appellate judges, and juries each play different roles in our civil justice system, and those roles are designed to capitalize on their institutional strengths and minimize the effects of their institutional weaknesses. Bearing those institutional characteristics in mind, sensible and functional decisions can be made about the assignment of responsibilities for estimating and constraining priors. Our system of civil justice and our rules of civil procedure are generally designed to assign responsibilities to those institutions in the best position to act on questions of law or fact, and they can similarly assign responsibilities for the estimation of priors. This concept of incorporating Bayesian priors into the ordinary civil litigation process is not fanciful or unprecedented. Rather, employment discrimination scholars can draw upon a body of legal and social science scholarship and judicial precedent in other areas of substantive law. Specifically, courts and scholars have long considered the implications of Bayesian statistical analysis in making paternity determinations and interpreting DNA evidence in criminal prosecutions.

Part I of this Article briefly outlines the current legal framework of systemic disparate treatment antidiscrimination law. Part II sketches the problem of acknowledging and estimating priors in the context of systemic employment discrimination cases. In doing so, Part II contends that imperfect solutions to prior estimation are nonetheless preferable to the status quo. Even imperfect, unsatisfying Bayesian solutions make the operation of priors

---

18. *Id.* at 849-50.

transparent when evaluating statistical evidence. Having set out the problem of priors, Part III turns to the specific second-order questions posed above. Tackling each of the four second-order questions, Part III argues that a sensible solution to prior management must capitalize on the institutional strengths of each component of the civil justice system. Finally, Part IV considers some pragmatic limitations that may counsel against application of express Bayesian analysis in antidiscrimination cases beyond systemic disparate treatment cases. Part IV also briefly charts a course for further scholarly discourse on the integration of prior probabilities in systemic discrimination law.

#### I. THE LEGAL FRAMEWORK: SYSTEMIC DISPARATE TREATMENT DOCTRINE

Title VII of the Civil Rights Act of 1964 prohibits discrimination “because of . . . race, color, religion, sex, or national origin.”<sup>19</sup> In the late 1970s, the Supreme Court, recognizing that direct proof of intentional discrimination is often unavailable to plaintiffs, identified the “pattern or practice”—or systemic disparate treatment—method of proving unlawful discrimination.<sup>20</sup> This systemic disparate treatment theory is not spelled out in the statute, but instead was formed by the Supreme Court’s interpretation of Title VII.<sup>21</sup> The systemic disparate treatment theory is distinct

---

19. See Pub. L. No. 88-352, § 703(a), 78 Stat. 255 (codified as amended in scattered portions of 42 U.S.C., primarily 42 U.S.C. § 2000e to 2000e-17 (2010)).

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

21. 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(e) (2010). This subsection originally granted the Attorney General the authority to investigate and act upon a charge of a “pattern or practice of discrimination.” *Id.* That authority was subsequently transferred to the Equal Employment Opportunity Commission (EEOC) in 1972. See Equal Employment Opportunity Act of 1972, Pub. L. No. 92-261, 86 Stat. 103, 107, § 5 (codified at 42 U.S.C. § 2000e-6(c) (2010)).

from the individual disparate treatment theory, epitomized by *McDonnell Douglas Corp. v. Green*<sup>22</sup> and the disparate impact theory, first recognized in *Griggs v. Duke Power Co.*<sup>23</sup>

The two cases that first recognized and defined the systemic disparate treatment theory were *International Brotherhood of Teamsters v. United States*<sup>24</sup> and *Hazelwood School District v. United States*.<sup>25</sup> Both cases permitted proof of a prima facie case of unlawful discrimination through statistical evidence, relying on this famous central premise:

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.<sup>26</sup>

The *Teamsters* Court set out a two-phase proof framework for systemic disparate treatment cases. In Phase I, plaintiffs may make out a prima facie case of discrimination by showing “that unlawful discrimination has been a regular procedure or policy followed by an employer . . . .”<sup>27</sup> The Court made clear that statistical evidence could be used to make this Phase I showing.<sup>28</sup> This statistical evidence may be combined with the anecdotal testimony of some affected individuals to bring the “cold

---

22. 411 U.S. 792 (1973).

23. 401 U.S. 424 (1971).

24. 431 U.S. 324 (1977).

25. 433 U.S. 299 (1977).

26. *Teamsters*, 431 U.S. at 339-40 n.20 (emphasis added); see also Browne, *Statistical Proof*, supra note 9, at 503-05; Paul Meier et al., *What Happened in Hazelwood: Statistics, Employment Discrimination, and the 80% Rule*, in STATISTICS AND THE LAW 20-21 (Morris H. DeGroot et. al. eds., 1986) (arguing that this central premise, taken literally, is not tenable, but may be viewed as an aspirational statement of impartiality).

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

28. *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 Philadelphia v. Educ. Equal. League*, 415 U.S. 605, 620 (1974)).

numbers convincingly to life.”<sup>29</sup> Upon a successful prima facie showing by the plaintiffs, “[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.”<sup>30</sup> If the defendant fails to defeat the prima facie showing of a pattern or practice of discrimination in Phase I, then the plaintiffs are entitled to prospective relief, such as an injunctive order.<sup>31</sup> In Phase II, plaintiffs may seek individual relief. A successful showing by plaintiffs in Phase I creates a rebuttable presumption that the employer’s pattern or practice of discrimination affected each individual claimant for purposes of determining individual relief in Phase II.<sup>32</sup>

In *Hazelwood*, the Court’s second systemic disparate treatment case, the Court reiterated its approval of the use of statistical evidence.<sup>33</sup> The Court, citing *Teamsters*, 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.”<sup>34</sup>

---

29. *Id.* at 339.

30. *Id.* at 360.

31. *Id.* at 361.

32. 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.”).

33. See *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 308-09.

34. *Id.* at 307-08 (citing *Teamsters*, 431 U.S. at 339); see also Michael Selmi, *Theorizing Systemic Disparate Treatment Law: After Wal-Mart v. Dukes*, 32 BERK. J. EMP. & LAB. L. 477, 480 (2011) (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., *E.E.O.C. 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) (“Plaintiffs may sustain their burden at the prima facie stage exclusively on statistical evidence, for ‘no sound policy reason exists for subjecting the plaintiff

The *Hazelwood* Court endorsed the use of a statistical technique known as binomial distribution analysis (or standard deviation analysis) to show the required “gross statistical disparities.”<sup>35</sup> A binomial distribution analysis compares the difference between the expected and observed number of hires from a relevant labor pool. If the observed number of people hired in the protected category (race, sex, etc.) is “greater than two or three standard deviations” from the expected number of hires from that protected group, then the hypothesis that applicants were hired without regard to the protected characteristic can be treated as suspect.<sup>36</sup>

Following *Teamsters* and *Hazelwood*, “a flood of statistical tests of significance, confidence intervals, and multiple regressions thundered forth from the lower courts.”<sup>37</sup> Several years later, in 1986, the Court first approved the use of multiple regression analysis, a more sophisticated statistical methodology, to establish a prima facie case of discrimination.<sup>38</sup> The Court’s most recent systemic disparate treatment case was its 2011 decision in *Wal-Mart Stores, Inc. v. Dukes*, which involved the use of regression statistics in an attempt to establish a nationwide class of female victims of discrimination in pay and promotions at Wal-Mart stores.<sup>39</sup>

---

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)).

35. *Hazelwood*, 433 U.S. at 307-08.

36. *Id.* at 308 n.14 (quoting *Castaneda v. Partida*, 430 U.S. 482, 496 n.17 (1977)).

37. Meier et al., *supra* note 26, at 3. Systemic theories are a high priority in the EEOC’s enforcement strategy. The EEOC’s Strategic Plan for Fiscal Years 2012-2016 emphasizes the importance of systemic cases and contains performance measures based on the number of systemic cases. *See generally* EEOC, STRATEGIC PLAN FOR FISCAL YEARS 2012-2016, available at [http://www.eeoc.gov/eeoc/plan/strategic\\_plan\\_12to16.cfm](http://www.eeoc.gov/eeoc/plan/strategic_plan_12to16.cfm).

38. *Bazemore v. Friday*, 478 U.S. 385, 398-403 (1986) (involving a claim of salary discrimination).

39. 131 S. Ct. 2541, 2547 (2011); *see also* Angela D. Morrison, *Duke-ing Out Pattern or Practice After Wal-Mart: The EEOC as Fist*, 63 AM. U. L. REV. 87, 107-08 (2013) (describing plaintiffs’ claims in *Wal-Mart*); Joseph A. Seiner, *Weathering Wal-Mart*, 89 NOTRE DAME L. REV. 1343, 1344 (2014) (acknowledging the general perception among academics that *Wal-Mart* weakened the

Statistical evidence, particularly binomial distribution and multiple regression analyses, remains the cornerstone of most systemic discrimination cases.<sup>40</sup> Basing liability decisions on these traditional statistical methods (without acknowledging the influence of priors) is problematic, for the reasons illustrated by the *Nature* article. These traditional statistical methods can generate a statistically significant *p*-value, but without information about how likely or unlikely any given hypothesis is *before* observing the statistical evidence, these traditional tests of statistical significance cannot tell the court much about the actual likelihood of the hypothesis—unlawful discrimination. The *Nature* article puts it this way:

Most scientists would look at [a] *P* value of 0.01 and say that there was just a 1% chance of [the study's] result being a false alarm. But they would be wrong. The *P* value cannot say this: all it can do is summarize the data assuming a specific null hypothesis. It cannot work backwards and make statements about the underlying reality. That requires another piece of information: the odds that a real effect was there in the first place. To ignore this would be like waking up with a headache and concluding that you have a rare brain tumour—possible, but so unlikely that it requires a lot more evidence to supersede an everyday explanation such as an allergic reaction.<sup>41</sup>

The other “piece of information” described in the *Nature* article is the prior probability. For purposes of evaluating statistical evidence in employment discrimination cases, it is the prior probability that the employer engaged in discrimination. Traditional statistical analyses of employment outcome disparities can provide the *p*-value for

---

substantive civil rights protections of Title VII and exploring potential ways to minimize *Wal-Mart's* impact); Stephanie S. Silk, Note, *More Decentralization, Less Liability: The Future of Systemic Disparate Treatment Claims in the Wake of Wal-Mart v. Dukes*, 67 U. MIAMI L. REV. 637, 641 (2013) (describing the statistical evidence offered by plaintiffs in *Wal-Mart*).

40. MICHAEL J. ZIMMER ET AL., CASES AND MATERIALS ON EMPLOYMENT DISCRIMINATION 145-55 (7th ed. 2008); Roberto L. Corrada, Ricci's *Dicta*: Signaling a New Standard for Affirmative Action Under Title VII?, 46 WAKE FOREST L. REV. 241, 249 n.59 (identifying binomial distribution and multiple regression as the two types of statistical analysis approved by the Supreme Court); Meier et. al., *supra* note 26, at 3; Selmi, *supra* note 34, at 485.

41. Nuzzo, *supra* note 1, at 151.

the observed disparities, when testing the null hypothesis of completely random employment decisions, but they cannot “work backwards” to indicate just how likely it is that the employer in question actually discriminated.<sup>42</sup> The next Part considers the problem of estimating or specifying a prior probability in the employment discrimination context.

## II. THE PRIOR PROBLEM: WHY IMPERFECT BAYESIAN SOLUTIONS ARE SUPERIOR TO THE STATUS QUO

*Hidden Priors*, along with works by other scholars, including Ramona Paetzold<sup>43</sup> and Deborah Weiss,<sup>44</sup> make the case for a Bayesian approach to the use and interpretation of statistical evidence in systemic disparate treatment law. That case will not be repeated here, and the reader’s general familiarity with that argument will be presumed. Nevertheless, one point bears reemphasizing before considering the second-order questions to be addressed in Part III: the superiority of imperfect Bayesian solutions over the status quo.

As others have noted, the primary argument leveled against a Bayesian approach to statistical evidence in the litigation context is the difficulty of estimating or specifying prior probabilities.<sup>45</sup> But the status quo approach to systemic

---

42. See Browne, *Statistical Proof*, *supra* note 9, at 488 (“[T]he probability that an employer’s work-force disparities are a consequence of chance is completely dependent upon a statistic which the courts never have: the likelihood of discrimination prior to making the employment decision. Although one might attempt some estimate of the percentage of employers who engage in systematic discrimination, the estimate could be no more than the crudest approximation.”).

43. See, e.g., Ramona L. Paetzold, *Problems with Statistical Significance in Employment Discrimination Litigation*, 26 NEW ENG. L. REV. 395 (1991).

44. See, e.g., Deborah M. Weiss, *The Impossibility of Agnostic Discrimination Law*, 2011 UTAH L. REV. 1677 (2011).

45. See, e.g., David E. Adelman, *Scientific Activism and Restraint: The Interplay of Statistics, Judgment, and Procedure in Environmental Law*, 79 NOTRE DAME L. REV. 497, 572 (2004) (“For a variety of reasons, Bayesians are criticized for their reliance on such uniform [prior] distributions as a justifiable starting point for Bayesian analysis.”); Norman Fenton & Martin Neil, *Avoiding Probabilistic Reasoning Fallacies in Legal Practice using Bayesian Networks*, AUSTL. J. LEGAL PHIL. (forthcoming 2011) (draft at 9-10), available at [http://www.eecs.qmul.ac.uk/~norman/papers/fenton\\_neil\\_prob\\_fallacies\\_June2011web.pdf](http://www.eecs.qmul.ac.uk/~norman/papers/fenton_neil_prob_fallacies_June2011web.pdf)

disparate treatment law—using traditional hypothesis testing “statistical significance” to impose liability and shift burdens of proof onto the defendant<sup>46</sup>—involves a misleading application of *p*-values in a context for which they were never intended,<sup>47</sup> and obscures hidden assumptions about the prevalence of discrimination.<sup>48</sup> Decision rules turning on statistically significant results from traditional hypothesis testing necessarily involve a built-in, unstated, and unexamined assumption about the prevalence, or base rate, of discrimination—a hidden prior.<sup>49</sup> Even imperfect Bayesian solutions have the benefit of making the specification of the

---

[hereinafter Fenton & Neil, *Probabilistic Reasoning*] (referring to “the obligation to consider a prior probability for a hypothesis like ‘guilty’” as “one of the most commonly cited objections to Bayes Theorem”); Kristy L. Fields, *Toward a Bayesian Analysis of Recanted Eyewitness Identification Testimony*, 88 N.Y.U. L. REV. 1769, 1792 (2013) (“The choice of prior and the feasibility of using probabilistic statistics in jury trials have been hotly contested topics in both Bayesian and Bayesioskeptic camps.”); Kruschke, *supra* note 14, at 3 (“Some people may have the mistaken impression that the advantages of Bayesian methods are negated by the need to specify a prior distribution.”); Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439, 463 (1986) (“What then is the proper prior? This is, of course, the sticking point in all attempts to apply a Bayesian approach to real world decisionmaking.”); Marcel C. Garaud, Comment, *Legal Standards and Statistical Proof in Title VII Litigation: In Search of a Coherent Disparate Impact Model*, 139 U. PA. L. REV. 455, 497, 501-03 (1990) (noting the difficulty of quantifying a prior distribution, but urging a Bayesian approach to disparate impact liability).

46. See *Bazemore v. Friday*, 478 U.S. 385, 398-400 (1986); *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 307-09 (1977); *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 360-61 (1977).

47. See Nuzzo, *supra* note 1, at 150-51 (“The irony is that when UK statistician Ronald Fisher introduced the *P* value in the 1920s, he did not mean it to be a definitive test. He intended it simply as an informal way to judge whether evidence was significant in the old-fashioned sense: worthy of a second look. The idea was to run an experiment, then see if the results were consistent with what random chance might produce.”).

48. See generally Bent, *Hidden Priors*, *supra* note 3, at 840-42.

49. See Bent, *Hidden Priors*, *supra* note 3, at 840 (quoting RAMONA L. PAETZOLD & STEVEN L. WILLBORN, *THE STATISTICS OF DISCRIMINATION: USING STATISTICAL EVIDENCE IN DISCRIMINATION CASES* § 12.05, at 12-13 (Thomson West 2009)); see also Weiss, *supra* note 44, at 1689-93.

prior transparent, a feature that makes such imperfect solutions superior to the status quo.<sup>50</sup>

Still, one might ask whether the project of trying to specify, incorporate, and manage prior probabilities in systemic disparate treatment litigation is feasible or advisable. After all, as Professor Kingsley Browne argues, the true base rate of discrimination is something that courts “never have” and, in his view, estimates of the base rate “could be no more than the crudest approximation.”<sup>51</sup>

Professor Browne is correct that, at least given current data and technology, estimates of base rates of discrimination will be crude approximations. But, as I have previously argued elsewhere, there are at least *some* possible sources for developing base rate estimates.<sup>52</sup> Admittedly, none of the potential sources for estimating prior probabilities are perfect methods for identifying true base rates. But consider the only two other alternatives. The traditional (or frequentist) approach used today, under the *Teamsters* and *Hazelwood* framework, operates with a hidden, built-in, unexamined prior.<sup>53</sup> There is no reason to suppose that this unexamined prior is any more accurate than a prior probability distribution that might be imperfectly selected by a court, an expert witness, or a fact-finder that affirmatively focused on the question of specifying

---

50. See Kruschke, *supra* note 14, at 3 (“In fact, the use of a prior is both appropriate for rational inference and advantageous in practical applications. . . . It is *inappropriate not* to use a prior.”); see also Stephen Charest, *Bayesian Approaches to the Precautionary Principle*, 12 DUKE ENVTL. L & POL’Y F. 265, 291 (2002) (“[T]he Bayesian approach has characteristics that make it a more appropriate means of assessing even truly uncertain risks than its methodological rivals. Bayesian techniques are more transparent than classical hypothesis testing, and are consistent with well-established scientific principles such as falsifiability and simplicity when used to compare rival scientific hypotheses.”).

51. See Browne, *Statistical Proof*, *supra* note 9, at 488.

52. 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, 834 (2011) [hereinafter Bent, *Telltale Sign*] (pointing to empirical studies of employment discrimination litigation outcomes as well as various theoretical bases as possible sources from which prior probabilities might be estimated).

53. See *supra* note 49 and accompanying text.

a prior. Avoiding the prior probability question and maintaining the status quo approach to systemic disparate treatment cases only creates the illusion of objectivity by ignoring the base rate problem altogether.<sup>54</sup> An admittedly crude but deliberate approximation surely beats an unexamined and unspoken assumption.

The second alternative would be to completely discard the idea of proving systemic discrimination with statistical evidence of disparities in employment outcomes. This appears to be Professor Browne's favored approach,<sup>55</sup> but it has not been widely accepted, as demonstrated by courts' continuing acknowledgement of the importance of statistical evidence.<sup>56</sup> Courts have good reason not to completely reject statistical evidence of employment outcome disparities as irrelevant. Statistical analyses, including binomial distributions, multiple regressions, and perhaps someday an emerging technique known as potential outcomes (sometimes called "matching" or "causal inference"),<sup>57</sup> all tell us

---

54. See *supra* note 49 and accompanying text.

55. See Browne, *Statistical Proof*, *supra* note 9, at 553-54 ("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."). Elsewhere, 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 542, 549, 554. Professor Browne concludes: "If statistical proof of discrimination is still to be acceptable at all in court—which is perhaps 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)).

56. See, e.g., *E.E.O.C. v. Olson's Dairy Queens, Inc.*, 989 F.2d 165, 168 (5th Cir. 1993); *Satchell v. FedEx Express*, No. C 03-2659 SI, 2006 WL 3507913, at \*1-2 (N.D. Cal. Dec. 5, 2006); *McReynolds v. Sodexo Marriott Servs.*, 349 F. Supp. 2d 1, 17 (D.D.C. 2004).

57. Each of these statistical methodologies is discussed in detail in *Hidden Priors*, *supra* note 3, at 818-20, 824-28.

*something*. Traditional statistical evidence is, strictly speaking, “relevant” under the requirements of the Federal Rules of Evidence because it has a “tendency to make a fact [e.g., the employer discriminated] more or less probable than it would be without the evidence . . . .”<sup>58</sup> Indeed, the central premise of the Bayesian philosophy of understanding statistical evidence is that new evidence permits an observer (or fact-finder) to update prior knowledge, information, or estimates about a given hypothesis—by definition making the hypothesis “more or less probable than it would be without the evidence.”<sup>59</sup>

The *Nature* article drives this point home with a relatively simple diagram, reproduced for reference in the Appendix herein. The diagram illustrates the effect of statistically significant study results on three different hypotheses, with three different prior likelihoods: the “Long Shot” hypothesis, with a 5% prior; the “Toss-Up” hypothesis, with a 50% prior; and the “Good Bet” hypothesis, with a 90% prior probability.<sup>60</sup> The statistical evidence, with *p*-values of .01 and .05 respectively, make each of the three hypotheses more likely than they were before observing the statistical results.<sup>61</sup> In the legal sense, this makes the statistical results relevant under Federal Rule of Evidence 401(a).<sup>62</sup> Hence, statistically significant evidence of employment outcome disparities should be admissible in systemic disparate treatment cases because it tends to make the hypothesis of unlawful discrimination more probable than it would have been in the absence of such evidence. But, importantly, the statistical results do not make each hypothesis *more likely than not*. For some hypotheses—in the *Nature* example, the Long Shot hypothesis—the posterior probability remains well below 50%, even when considering statistical evidence

---

58. FED. R. EVID. 401(a).

59. *See id.*; *see also* Fields, *supra* note 45, at 1782 (describing Bayesian updating in the context of the probability of a criminal defendant’s guilt).

60. Nuzzo, *supra* note 1, at 151.

61. *Id.*

62. *See* FED. R. EVID. 401(a).

with a  $p$ -value of .01.<sup>63</sup> As illustrated by the *Nature* examples, the prior probability (long shot, toss-up, or good bet) plays a critical role in the determination of the posterior probability.<sup>64</sup> That role has been ignored in systemic disparate treatment law since its inception.<sup>65</sup>

Completely prohibiting the use of statistical evidence would go too far, eliminating an important source of relevant information in a field of law where direct evidence of wrongdoing is often unavailable.<sup>66</sup> If courts are going to continue admitting statistical evidence, however, then the choices essentially boil down to two: (1) continue to ignore the question of prior probabilities altogether and allow them to operate *sub silentio*; or (2) address prior probabilities head on, using admittedly flawed estimations and crude approximations.<sup>67</sup> The latter is the better choice, despite the difficulties involved in estimating prior distributions. Engaging in the effort to make sense of the statistical evidence that is available—flawed though that effort will be—ultimately serves the remedial goals of Title VII.

### III. PRIORS AND PROCEDURE: THE SECOND-ORDER QUESTIONS

Once courts and scholars recognize that an imperfect, but transparent, treatment of prior probabilities is preferable to either: (a) continuing to misapply  $p$ -values and statistical significance tests using traditional statistical analysis with built-in, unexamined priors; or (b) ignoring statistical evidence of employment outcome disparities altogether, then the difficult second-order questions come into focus. This

---

63. Nuzzo, *supra* note 1, at 151.

64. See Nuzzo, *supra* note 1, at 151; Appendix, *infra*.

65. See Bent, *Hidden Priors*, *supra* note 3, at 821-23; *supra* Part I.

66. See, e.g., Sean W. Colligan, *In Good Measure: Workforce Demographics and Statistical Proof of Discrimination*, 23 LAB. LAW. 59, 59 (2007) (“Courts have long recognized that plaintiffs in discrimination lawsuits rarely have the benefit of direct evidence of discrimination. Because overt expressions of discriminatory motives by managers are rare, indirect evidence of discriminatory intent is often crucial in employment cases. One commonly used form of indirect evidence of discrimination . . . is statistical analysis of the employer’s workforce.”) (footnotes omitted) (citing *Price Waterhouse v. Hopkins*, 490 U.S. 228, 271 (1989)).

67. See Browne, *Statistical Proof*, *supra* note 9, at 488.

Part considers each of the second-order questions in turn, and contends that our civil procedure devices are well-suited to the management of prior probabilities by capitalizing on institutional strengths.

#### QUESTION 1: WHOSE PRIORS MATTER?

The first question for prior management is: whose priors count? Professors Paetzold and Willborn acknowledge this question, but do not attempt an answer.<sup>68</sup> Possible answers to this question include trial judges, trial fact-finders, appellate judges, and legislatures. Perhaps the most logical answer is that it depends on the specific allegations and evidence, together with the procedural posture of the case. Indeed, perhaps the division of responsibilities between fact-finder, judge, and appellate court regarding priors should be consistent with that division of labor in all other civil litigation contexts. Our procedural rules are designed to permit judges to act as gatekeepers throughout the litigation process, and to provide standards of appellate review adjusted to the institutional strengths and weaknesses of appellate courts.

The specification of prior probabilities might be effectively managed in the same manner as other similar questions. The current procedural framework for judicial gatekeeping in civil litigation, including the dispositive motions provided under Federal Rules 12(b)(6)<sup>69</sup> and 56,<sup>70</sup> are increasingly interpreted to include explicitly probabilistic components,<sup>71</sup> and therefore appear well-suited to the task. Because the question of whose priors should matter is so closely intertwined with procedural devices, second-order Questions 1 and 2 turn out to be related. Both questions will be considered together in the following Sections, which

---

68. See PAETZOLD & WILLBORN, *supra* note 49, § 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.”).

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

70. See FED. R. CIV. P. 56.

71. See *infra* Part III, Question 2, secs. A–B.

consider the interplay of procedural gatekeeping devices and the specification of prior probabilities of discrimination.

QUESTION 2: HOW SHOULD EVALUATION OF PRIORS FIT INTO  
CIVIL LITIGATION PRETRIAL PROCEDURE?

A. *Motions to Dismiss*

1. *Plausibility Pleading and Priors.* The Court's recent decisions in *Bell Atlantic Corp. v. Twombly*<sup>72</sup> and *Ashcroft v. Iqbal*<sup>73</sup> have renewed interest in probability estimates and pleading standards. With these cases, the Court tightened the federal pleading standard by establishing a "plausibility" test.<sup>74</sup> The Court describes the standard as follows: "To survive a motion to dismiss, a complaint must contain sufficient factual matter, accepted as true, to 'state a claim to relief that is plausible on its face.'"<sup>75</sup> The Court prescribed a two-step analysis.<sup>76</sup> First, the court should separate nonconclusory factual allegations, which must be assumed true, from alleged legal conclusions, which need not be assumed true.<sup>77</sup> Second, the court should determine whether the plaintiff's nonconclusory allegations "plausibly give rise

---

72. 550 U.S. 544 (2007).

73. 129 S. Ct. 1937 (2009). Probabilities likely played some unstated, inchoate role in the federal pleading regime prior to the *Twombly* and *Iqbal* decisions as well. Even under the prior, more liberal notice pleading rule drawn from the Court's decision in *Conley v. Gibson*, 355 U.S. 41 (1957), courts could dismiss "claims about little green men, or the plaintiff's recent trip to Pluto, or experiences in time travel." See *Iqbal*, 129 S. Ct. at 1959 (Souter, J., dissenting); see also *Courie v. Alcoa Wheel & Forged Prods.*, 577 F.3d 625, 629 (6th Cir. 2009); Marc I. Steinberg & Diego E. Gomez-Cornejo, *Blurring the Lines Between Pleading Doctrines: The Enhanced Rule 8(a)(2) Plausibility Pleading Standard Converges with the Heightened Fraud Pleading Standards Under Rule 9(b) and the PSLRA*, 30 REV. LITIG. 1, 28-29 (2010).

74. Rule 8 provides, in relevant part: "A pleading that states a claim for relief must contain: . . . a short and plain statement of the claim showing that the pleader is entitled to relief . . ." FED. R. CIV. P. 8(a).

75. *Iqbal*, 129 S. Ct. at 1949 (quoting *Twombly*, 550 U.S. at 570).

76. *Id.* at 1950.

77. *Id.*

to an entitlement to relief.”<sup>78</sup> In making this determination, the court should draw reasonable inferences in favor of the plaintiff from the alleged nonconclusory facts,<sup>79</sup> but the court should also consider any “obvious alternative explanation” for the alleged facts.<sup>80</sup> If the facts alleged are “more likely explained by” the obvious alternative explanation than by plaintiff’s allegation of a legal violation, then plaintiff’s legal claim is not plausible and should be dismissed.<sup>81</sup> In making this determination, courts are instructed to rely on “judicial experience and common sense.”<sup>82</sup>

The Court denied that the plausibility standard requires an assessment of probabilities:

The plausibility standard is not akin to a “probability requirement,” but it asks for more than a sheer possibility that a defendant has acted unlawfully. Where a complaint pleads facts that are “merely consistent with” a defendant’s liability, it “stops short of the line between possibility and plausibility of ‘entitlement to relief.’”<sup>83</sup>

Despite this protestation, the plausibility standard articulated in *Iqbal* is unquestionably some form of probability requirement. The line between “mere possibility” and “plausibility” may not be quantifiable,<sup>84</sup> but that line must exist, and it must be located at some probability greater

---

78. *Id.*

79. *Id.* at 1949 (“A claim has facial plausibility when the plaintiff pleads factual content that allows the court to draw the reasonable inference that the defendant is liable for the misconduct alleged.”).

80. *Id.* at 1951 (quoting *Twombly*, 550 U.S. at 567).

81. *Id.* at 1951-52.

82. *Id.* at 1950.

83. *Id.* at 1949 (citations omitted). The *Twombly* Court likewise rejected any suggestion that it was imposing a probability requirement. *See Twombly*, 550 U.S. at 556 (“Asking for plausible grounds to infer an agreement does not impose a probability requirement at the pleading stage; it simply calls for enough fact to raise a reasonable expectation that discovery will reveal evidence of illegal agreement.”).

84. *See, e.g.*, David L. Noll, *The Indeterminacy of Iqbal*, 99 GEO. L.J. 117, 134-35 (2010) (noting that the Court did not explain the “threshold of plausibility” a complaint must cross, but suggesting that it could be compared to “something like probable cause to believe the defendant breached a legal duty owed to the plaintiff.”).

than 0 and less than 1. This is true even if the plausibility threshold is allowed to fluctuate based on the type of case or other factors.<sup>85</sup> By comparing the plaintiff's explanation to "obvious alternative explanations" and deciding which is "more likely," the Court requires a comparative probability assessment. The phrase "more likely," is simply another way of saying "more probably true," making the analysis a probability analysis, as others have noted.<sup>86</sup>

The probabilistic nature of the analysis is revealed by the competing explanations at issue in *Twombly* and *Iqbal*. *Twombly* involved allegations of an unlawful conspiracy among defendants to refrain from competing against each other in certain designated geographic regions, in violation of Section 1 of the Sherman Act.<sup>87</sup> Parallel conduct, even conscious parallelism, is not itself a violation of Section 1.<sup>88</sup> Section 1 prohibits "only restraints [of trade] effected by a contract, combination, or conspiracy."<sup>89</sup> The complaint alleged that defendants engaged in a "parallel course of conduct," and that, on information and belief, defendants had "entered into a contract, combination, or conspiracy" to prevent competition.<sup>90</sup> The Court held that the complaint

---

85. See Ronald J. Allen & Alan E. Guy, *Conley as a Special Case of Twombly and Iqbal: Exploring the Intersection of Evidence and Procedure and the Nature of Rules*, 115 PENN ST. L. REV. 1, 36 (2010) (contending that the plausibility pleading rule could be interpreted as a non-static rule that imposes a burden that fluctuates according to the particular characteristics of the claims at issue).

86. See, e.g., Allen & Guy, *supra* note 85, at 36 ("Strikingly, except for the obviously erroneous assertion that probabilism plays no part in plausibility, the Court applied the concept of plausibility in a straight forward inference to the best explanation fashion."); Rory Bahadur, *The Scientific Impossibility of Plausibility*, 90 NEB. L. REV. 435, 456-57 (2011) ("The Court's use of the term 'possibility,' however, belies the assertion that plausibility is not a probability analysis because possibility is an expression of probability. . . . Plausibility is therefore achieved when the complaint reaches a threshold level of probability, but the Court does not define what this threshold level of probability is."). Whether or not relative probabilities *should* be part of the federal pleading standard is a question that is beyond the scope of this Article.

87. *Twombly*, 550 U.S. at 548.

88. *Id.* at 553.

89. *Id.* (quoting *Copperweld Corp. v. Independence Tube Corp.*, 467 U.S. 752, 775 (1984)).

90. *Id.* at 551.

failed to state a claim.<sup>91</sup> The allegation that defendants “entered into a contract, combination, or conspiracy” was a conclusory allegation unsupported by nonconclusory facts, and was not entitled to a presumption of truth.<sup>92</sup> The alleged parallel conduct was “merely consistent” with conspiracy.<sup>93</sup> It was also consistent with an “obvious alternative explanation”—namely, “the former Government-sanctioned monopolists [defendants] were sitting tight, expecting their neighbors to do the same thing.”<sup>94</sup> This obvious alternative explanation was based on the Court’s understanding of how rational firms would behave in the free market. The Court had “concluded that [the *Twombly* complaint] did not plausibly suggest an illicit accord because it was not only compatible with, but indeed was more likely explained by, lawful, unchoreographed free-market behavior.”<sup>95</sup>

In *Iqbal*, the plaintiff brought claims against former Attorney General John Ashcroft and former FBI director Robert Mueller, alleging that the defendants violated plaintiff’s constitutional rights by subjecting plaintiff (a Pakistani Muslim) “to harsh conditions of confinement on account of his race, religion, or national origin.”<sup>96</sup> Plaintiff alleged that “under the direction of Defendant [Mueller], [the FBI] arrested and detained thousands of Arab Muslim men” during its investigation into the September 11, 2001 attacks.<sup>97</sup> Plaintiff alleged that defendants Ashcroft and Mueller approved a “policy of holding post-September-11th detainees in highly restrictive conditions of confinement until they were ‘cleared’ by the FBI . . . .”<sup>98</sup> Finally, plaintiff alleged that defendants “knew of, condoned, and willfully and maliciously agreed to subject [plaintiff] to harsh conditions of confinement “as a matter of policy, solely on account of [his]

---

91. *Id.* at 570.

92. *Id.* at 564-67.

93. *Id.* at 557.

94. *Id.* at 567-68.

95. *See* *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1950 (2009).

96. *Id.* at 1942.

97. *Id.* at 1944.

98. *Id.*

religion, race, and/or national origin and for no legitimate penological interest.”<sup>99</sup>

The Court found this last allegation to be conclusory, and therefore not entitled to an assumption of truth.<sup>100</sup> The Court then considered the nonconclusory factual allegations of a policy of holding detainees in harsh conditions of confinement until they were cleared by the FBI. The Court noted that these allegations were “consistent with” plaintiff’s assertion that defendants “purposefully designat[ed] detainees ‘of high interest’ because of their race, religion, or national origin.”<sup>101</sup> But, according to the Court, there was an obvious, alternative, and “more likely” explanation: “[T]he Nation’s top law enforcement officers, in the aftermath of a devastating terrorist attack, sought to keep suspected terrorists in the most secure conditions available until the suspects could be cleared of terrorist activity.”<sup>102</sup> Given this more likely explanation, the complaint did not “plausibly suggest [defendants’] discriminatory state of mind.”<sup>103</sup>

In both *Twombly* and *Iqbal*, the Court drew upon non-case-specific information to inform its analysis of relative probabilities. In *Twombly*, the Court relied on microeconomic theory to determine that innocent parallel conduct was a more likely explanation for defendants’ behavior than unlawful conspiracy. In *Iqbal*, the Court relied on its assessment of the likely motivations of the nation’s top law enforcement officers to conclude that a lawful concern for safety was a more likely explanation for the alleged treatment than unlawful discrimination. In both cases, the Court applied its own understanding of background probabilities, or priors.

---

99. *Id.*

100. *Id.* at 1951 (“These bare assertions, much like the pleading of conspiracy in *Twombly*, amount to nothing more than a ‘formulaic recitation of the elements’ of a constitutional discrimination claim . . .”) (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 555 (2007)).

101. *Id.*

102. *Id.* at 1952.

103. *Id.*

2. *Competing Explanations in Systemic Disparate Treatment Cases.* In systemic disparate treatment cases the probabilistic nature of the analysis is more salient. Statistics—and therefore probabilities—are the crux of a systemic disparate treatment case.<sup>104</sup> The critical allegations in a typical systemic disparate treatment complaint will be: (1) there is an observed, statistically significant disparity in employment outcomes; and (2) the observed disparity is the result of unlawful discrimination. The former allegation is a nonconclusory allegation of fact and is assumed to be true at the pleading stage. The second allegation appears to be a conclusion of law under *Iqbal*, and the court need not accept it as true.<sup>105</sup>

A statistically significant disparity *could* be explained by unlawful discrimination. But the observed disparity could also be explained by other obvious alternative explanations. For example, the statistical disparity could be explained by the choice of an unrepresentative reference class in a binomial distribution, or by the omission of a critical explanatory variable in a regression analysis. Or, the observed disparity—even if statistically significant—could be explained by chance. To focus the discussion on the analysis of priors, I will focus only on two possible explanations for an observed disparity: (1) unlawful discrimination; and (2) chance.<sup>106</sup> Applying *Iqbal*, the question for the court is whether an alleged statistical disparity is *more likely* explained by unlawful discrimination or instead *more likely* explained by the obvious alternative explanation: chance.

As demonstrated in *Hidden Priors* and the discussion above, the answer to this question is a mathematical function

---

104. See Natalie Bucciarelli Pedersen, *The Hazards of Dukes: The Substantive Consequences of a Procedural Decision*, 44 U. TOL. L. REV. 123, 135 (2012).

105. See *Iqbal*, 129 S. Ct. at 1950.

106. Another candidate for an “obvious alternative explanation” might be that the observed statistical disparities are explained by differences in other independent variables, such as qualifications, education, experience, or levels of interest in the job. Multiple regression analyses attempt to isolate and control for the effects of such independent variables. See generally Meier et al., *supra* note 26, at 7-15. However, even where a multiple regression analysis is used, the statistician’s choice of independent variables to include in the analysis can be challenged by the opposing party. See *id.* at 19.

of the prior probability of discrimination and the  $p$ -value. In the simplified example offered in *Hidden Priors*, an assumed prior of less than 5% would mean that chance is a more likely explanation than discrimination for an observed disparity that has a  $p$ -value equal to .05.<sup>107</sup> Conversely, an assumed base rate of more than 5% would mean that unlawful discrimination is a more likely explanation for that same observed disparity.<sup>108</sup> The choice of prior can be determinative, even at the motion to dismiss stage.

How should the court determine what prior to use? Should the district judge use his or her own preconceptions about the prevalence of discrimination—i.e., “judicial experience and common sense”?<sup>109</sup> Or should the judge consider that a jury might reasonably arrive at a different prior? Given the early procedural stage, the judge should allow for a wide range of potential priors, recognizing that jurors may hold vastly different notions about the background rate of discrimination, at least given the limited empirical evidence available to date.<sup>110</sup>

To return once again to the simplified example provided in *Hidden Priors*, if the district judge believes that a jury could reasonably assign a prior probability of employment discrimination higher than 5%, and if the plaintiff alleges in the complaint an observed statistical disparity that is

---

107. See Bent, *Hidden Priors*, *supra* note 3, at 822. This example is simplified in a number of ways, including that it assumes there are no false negatives (or Type II errors), i.e., the statistical test always catches all true discriminators. See *id.* at 821-24. The diagram in the *Nature* article, in which the Long Shot (5%) prior is considered in light of  $p$ -values of .01 and .05, drew upon a more careful calibration of  $p$ -values for testing *precise* null hypotheses (e.g., the effect of the independent variable gender on the dependent variable salary is exactly = 0), and it also included consideration of the probability of false negatives. See Nuzzo, *supra* note 1, at 151 (citing Thomas Sellke et al., *Calibration of p Values for Testing Precise Null Hypotheses*, 55 AM. STATISTICIAN 62 (2001)); see also Appendix, *infra*.

108. See Bent, *Hidden Priors*, *supra* note 3, at 822-23.

109. See *Iqbal*, 129 S. Ct. at 1950; see also *supra* text accompanying notes 71-77.

110. For a discussion of why deference to the jury may be institutionally preferable where empirical evidence or academic theory offers little basis for specifying a prior, see *infra* Part III, Question 4, sec. E.

statistically significant at the .05 level (to be later established by admissible evidence), then the district court should find the complaint sufficient under the *Iqbal* plausibility standard.

The foregoing considers cases where observed statistical disparity is the only nonconclusory factual allegation supporting the legal conclusion of discrimination. As part of the court's plausibility analysis, however, the court should also consider other nonconclusory factual allegations. For example, if plaintiffs allege that every member of the class was fired by the same rogue store manager who refused to follow company nondiscrimination policies, then the court might find unlawful discrimination a more likely explanation for an observed disparity than chance, even if the court might have reached the opposite conclusion in the absence of such nonstatistical allegations.<sup>111</sup>

3. *Beyond Probabilities: Access to Information.* The pleading standard analysis should not be limited to only the comparative probability analysis described above. Plaintiffs have not yet had an opportunity to conduct discovery and obtain access to evidence in defendants' possession that may substantiate their claims. The *Twombly* Court apparently recognized this information access issue, stating that the plausibility standard "simply calls for enough fact to raise a reasonable expectation that discovery will reveal evidence of illegal agreement."<sup>112</sup> This suggests that, along with a relative probability analysis, trial courts should also consider the parties' relative access to evidence pre-discovery.

Neither *Twombly* nor *Iqbal* provide further guidance on how to incorporate the information access problem.<sup>113</sup> The

---

111. This type of plausibility analysis at the motion to dismiss stage appears consistent with Professor Michael Selmi's general notion that, as compared to courts in the late 1970s, courts are now more likely to require plaintiffs to offer an explanation of the story or narrative that the statistics are telling. See Selmi, *supra* note 34, at 481 (arguing that a statistical showing alone is no longer sufficient to establish a systemic disparate treatment claim, but that it is incumbent upon plaintiffs to also "explain the story the statistical presentation is telling . . .").

112. *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 556 (2007).

113. The *Iqbal* Court rejected the argument that tight controls on discovery can effectively prevent unwarranted discovery from disrupting the work of

*Twombly* passage quoted above, however, leaves room for district judges to assess the parties' relative access to information. If the district judge compares the competing explanations for an observed statistical disparity, taking into account the wide range of priors that a reasonable jury might adopt, and concludes that chance is a more likely explanation for an observed disparity in employment outcomes than the employer's unlawful discrimination, then the district court should next consider whether discovery can be reasonably expected to change that assessment.<sup>114</sup> If so, the court should deny the motion to dismiss. At the very least, the court should permit limited, targeted, pleading-stage discovery to permit a more careful assessment of the relative probabilities of the competing theories of chance and discrimination.<sup>115</sup>

### B. Summary Judgment

How should the district court consider priors at the summary judgment stage? At this point, the parties have had an opportunity to obtain through the discovery process any information relevant to the probability determination. The court should therefore focus only on the question of whether a "justifiable inference[ ]"<sup>116</sup> of unlawful discrimination is warranted by the evidence presented, when viewed in the light most favorable to the non-moving party (typically the plaintiff in discrimination cases).<sup>117</sup>

---

governmental officials. See *Iqbal*, 129 S. Ct. at 1953 (citing *Twombly*, 550 U.S. at 559).

114. See *Twombly*, 550 U.S. at 556; see also Arthur R. Miller, *From Conley to Twombly to Iqbal: A Double Play on the Federal Rules of Civil Procedure*, 60 DUKE L.J. 1, 105 n.405 (2010) (noting that more robust EEOC pre-suit investigation could alleviate problems of relative access to information in discrimination cases).

115. See generally Edward A. Hartnett, *Taming Twombly, Even After Iqbal*, 158 U. PA. L. REV. 473, 508-14 (2010) (considering ways in which district courts might permit discovery to proceed during pendency of motion to dismiss); Bahadur, *supra* note 86, at 480 n.295.

116. *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 255 (1986).

117. See *Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574, 587-88 (1986) (quoting *United States v. Diebold, Inc.*, 369 U.S. 654, 655 (1962)); see also FED. R. CIV. P. 56.

The relative probability analysis will be essentially the same as outlined above, except that it will now be conducted with the documentary record.<sup>118</sup> This evidence must be “viewed in the light most favorable to the party opposing the motion.”<sup>119</sup> The similarity of the relative probability analysis for summary judgment purposes to the plausibility pleading standard is highlighted by comparing *Twombly* to one in the trilogy of cases that redefined summary judgment in 1986, *Matsushita Electric Industrial Co. v. Zenith Radio Corp.*<sup>120</sup> In *Matsushita*, the Court considered an antitrust case quite similar to *Twombly*. The plaintiffs were American television manufacturers who brought claims under Section 1 of the Sherman Act against Japanese manufacturers, alleging a price-fixing conspiracy.<sup>121</sup> Defendants moved for summary judgment, arguing that there was insufficient evidence to establish an illegal conspiracy.<sup>122</sup> The Court held that summary judgment would be appropriate where there was insufficient evidence to support an inference of a conspiracy, in light of more likely—more *plausible*—explanations for the evidence.<sup>123</sup> The Court reasoned: “Respondents in this case, in other words, must show that the inference of conspiracy is reasonable *in light of the competing inferences* of independent action or collusive action that could not have harmed respondents.”<sup>124</sup>

For systemic disparate treatment employment discrimination cases, the analysis would be similar: the court should consider whether an inference of discrimination is a

---

118. Federal Rule 56 provides that the record for summary judgment may consist of “depositions, documents, electronically stored information, affidavits or declarations, stipulations (including those made for purposes of the motion only), admissions, interrogatory answers, or other materials . . . .” FED. R. CIV. P. 56(c)(1)(A).

119. *See Matsushita*, 475 U.S. at 587-88 (quoting *Diebold*, 369 U.S. at 655).

120. 475 U.S. 574.

121. *See id.* at 577-78.

122. *Id.* at 578.

123. *See id.* at 587, 597-98.

124. *Id.* at 588 (emphasis added); *see also* Michael J. Kaufman, *Summary Pre-Judgment: The Supreme Court’s Profound, Pervasive, and Problematic Presumption about Human Behavior*, 43 LOY. U. CHI. L.J. 593, 600 (2012).

reasonable explanation for the observed statistical disparity (now shown through record evidence, rather than just allegations) in light of the alternative explanation: chance. In making this determination, the court should again consider the entire range of prior probabilities that a reasonable jury could adopt. The relevant prior probabilities will be informed by the estimated background rate of discrimination in the relevant labor market, as well as all the nonstatistical evidence in the record.<sup>125</sup>

At summary judgment, a court may be more tempted to substitute its own preconceptions about base rates of discrimination for the jury's.<sup>126</sup> When priors are not transparent, the temptation to engage in this type of substitution is heightened. The majority opinion in *Wal-Mart*, the Court's most recent systemic disparate treatment case, is illustrative.<sup>127</sup> Although *Wal-Mart* actually involved a motion for class certification, rather than summary judgment, the Court acknowledged that class certification decisions must sometimes overlap with inquiries into the merits.<sup>128</sup> As such, class certification motions will involve similar concerns about courts substituting their priors for those of a reasonable jury. In *Wal-Mart*, plaintiffs sought to certify a nationwide class of female current and former Wal-Mart employees, alleging pay and promotion discrimination.<sup>129</sup> The plaintiffs' central theory was that Wal-Mart's decentralized promotion policy, which left promotion decisions largely to the discretion of store managers, resulted in systematic discrimination companywide against women.<sup>130</sup> The majority refused to permit certification of the nationwide

---

125. See Bent, *Hidden Priors*, *supra* note 3, at 810 n.8.

126. See generally Suja A. Thomas, *The Fallacy of Dispositive Procedure*, 50 B.C. L. REV. 759, 769-73 (2009) (arguing that "judges decid[e] dispositive motions based on their own views of the facts[.]" despite the supposed requirement that judges consider "whether a reasonable jury could find" for the nonmovant).

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

128. See *id.* at 2552 ("In this case, proof of commonality necessarily overlaps with respondents' merits contention that Wal-Mart engages in a *pattern or practice* of discrimination.").

129. See *id.* at 2547-48.

130. See *id.*

class, finding there was not sufficient commonality under Federal Rule 23.<sup>131</sup> In reaching this conclusion, the majority made an empirically testable (or falsifiable) claim about background rates of discrimination. The majority reasoned: “[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.”<sup>132</sup> What the majority did in this sentence, quite evidently, was to substitute its own priors on the background rate of promotion discrimination against women for the range of potential priors that might be held by a reasonable jury.

The opaqueness of the operation of priors in current systemic disparate treatment law allowed the majority to include this empirical claim, without appearing to usurp the role of the jury. Were systemic disparate treatment law to undergo a Bayesian revolution, and were prior probabilities openly acknowledged and managed through the usual civil procedure devices, then the import of the *Wal-Mart* majority’s statement would have been obvious—it substituted its own priors for that of a reasonable jury. Whether all reasonable juries would necessarily agree with the majority’s statement about what “most managers” would do when making promotion decisions is a separate question, but it is one that is never even broached under current systemic disparate treatment doctrine because it is not recognized as a question about priors.<sup>133</sup>

Unstated assumptions held by a particular trial judge about the prevalence (or scarcity) of discrimination could lead that judge to conclude that proffered statistical evidence is more (or less) likely to be the result of unlawful discrimination. This unstated use of the judge’s own prior, rather than a range of priors that a jury might reasonably hold, should be guarded against in dispositive pretrial

---

131. See *id.* at 2554; see also FED. R. CIV. P. 23(a)(2).

132. *Wal-Mart*, 131 S. Ct. at 2554; see also Weiss, *supra* note 44, at 1687 (referring to this statement as the “*Wal-Mart* presumption”).

133. See generally Bent, *Hidden Priors*, *supra* note 3, at 843-47 (discussing in further detail the underlying influence of prior probabilities in *Wal-Mart*).

motions. Prior probability assessments are necessarily somewhat subjective and will differ from person to person. A judge's priors will be shaped by his or her own experiences, and will not match those of the individual jury members. For summary judgment purposes, the court should consider that range of priors that a reasonable jury could hold, viewing the nonstatistical evidence in plaintiff's favor.<sup>134</sup>

### C. Appellate Review

What standard should be used to review prior probability determinations on appeal? Motions to dismiss and for summary judgment have been characterized as questions of law, and appellate review of these procedures has traditionally been *de novo*.<sup>135</sup> Once the role of specified prior probabilities in pretrial dispositive motions is recognized, however, this traditional standard of review might fairly be questioned. Questions of fact are generally reviewed under a more deferential standard than questions of law; a trial court's factual findings are generally not set aside by an appellate court unless they are "clearly erroneous."<sup>136</sup> Which standard should apply to a trial court's determination of priors in pretrial dispositive motions? Which court is in a better position to evaluate the reasonableness of a given range of priors for purposes of motions to dismiss or motions for summary judgment?

*Iqbal's* importation of "judicial experience and common sense" into the pleading standard has already raised questions about whether *de novo* review remains appropriate

---

134. Professor Suja Thomas might question whether a judge is even capable of performing such a metaphysical analysis. See Thomas, *supra* note 126, at 779 ("[I]t should not be assumed that judges could determine who would sit on a jury and could consider all viewpoints of those jurors in their decision of whether a reasonable jury could find for the plaintiff.").

135. On dismissals under Federal Rule 12(b)(6), see, e.g., *Levy v. Ohl*, 477 F.3d 988, 991 (8th Cir. 2007); *Blackstone Realty LLC v. F.D.I.C.*, 244 F.3d 193, 197 (1st Cir. 2001); *Shepherd v. Sanchez*, 27 F. App'x. 31, 33 (2d Cir. 2001). On summary judgment under Federal Rule 56, see, e.g., *McCoy v. Harrison*, 341 F.3d 600, 604 (7th Cir. 2003); *White v. ABCO Eng'g Corp.*, 221 F.3d 293, 300 (2d Cir. 2000).

136. FED. R. CIV. P. 52(a)(6).

for Rule 12(b)(6) dismissals: “[District] Judge [Sidney H.] Stein ‘questioned what the scope of review on appeal of rulings on such motions will be, observing that he does not know how closely appeals courts will be regulating the application of judicial experience and common sense in reviewing motions to dismiss.’”<sup>137</sup>

If “judicial experience and common sense” includes a comparison of relative probabilities in light of the judge’s assessment of reasonable prior probabilities, then perhaps a more discretionary standard of review is warranted. Prior probabilities, which may be thought of as “foundational facts” (sometimes called “legislative facts”<sup>138</sup> or “social facts”<sup>139</sup>), occupy an uneasy middle ground between fact and law. They are, in the scientific sense of the word, “facts,” because they

---

137. Henry S. Noyes, *The Rise of the Common Law of Federal Pleading: Iqbal, Twombly, and the Application of Judicial Experience*, 56 VILL. L. REV. 857, 882 (2012) (quoting the statements of Judge Sidney H. Stein, Southern District of New York, as recounted in Pamela Atkins, Twombly, Iqbal *Introduce More Subjectivity to Rulings on Dismissal Motions, Judge Says*, 78 U.S.L.W. 2667, 2667 (2010)); see also Edward H. Cooper, *King Arthur Confronts TwIqy Pleading*, 90 OR. L. REV. 955, 983 (2012) (reviewing Miller’s *Double Play* article, and noting the question raised by Professor Miller: “Need the rules be amended to ensure continued de novo review of dismissals for failure to state a claim?”); Miller, *supra* note 114, at 35 (“And has the traditional de novo standard used on appellate review of a grant of a Rule 12(b)(6) motion been compromised by the subjective appraisals the Court has authorized?”).

138. See Caitlin E. Borgmann, *Appellate Review of Social Facts in Constitutional Rights Cases*, 101 CALIF. L. REV. 1185, 1188 (2013) [hereinafter Borgmann, *Appellate Review*] (considering the question of appellate deference to trial courts on findings of social fact in constitutional rights cases); Tristin K. Green, *“It’s Not You, It’s Me”: Assessing an Emerging Relationship Between Law and Social Science*, 46 CONN. L. REV. 287, 291 n.8 (2013) (discussing social framework evidence in discrimination cases as legislative fact); Weiss, *supra* note 44, at 1715 (applying the classifications of social authority and social fact described in Laurens Walker & John Monahan, *Social Facts: Scientific Methodology as Legal Precedent*, 76 CALIF. L. REV. 877, 880-82 (1988) and John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. PA. L. REV. 477, 488-89 (1986)); see generally Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364, 402 (1942) (distinguishing between legislative facts and adjudicative facts).

139. See Borgmann, *Appellate Review*, *supra* note 138, at 1187 (“The facts relevant to these kinds of cases are ‘social’ facts, also commonly referred to as ‘legislative’ facts.”).

are falsifiable.<sup>140</sup> They are, at least in the abstract, empirically testable notions about the state of the world.<sup>141</sup> The background rate of discrimination on the basis of race by all employers in the United States may be practically unknowable, given our current constraints on data sources and polling abilities, but theoretically it could be an empirically measured fact. But these are not case-specific, adjudicative (or, in Professor Suzanna Sherry's terminology, "decisional") facts.<sup>142</sup> They are generalized notions about the state of the world *before* taking into account the specific evidence offered in litigation about the specific defendant employer before the court.

Given that priors are not easily categorized as fact or law, what standard of review should apply to priors? This question closely mirrors the question of the appropriate scope of appellate review of "social facts" or "legislative facts" in constitutional cases.<sup>143</sup> The question of appellate review of priors, I posit, should be answered by taking a functional approach. In deciding the standard of appellate review, courts should ask the following question: is the district court in any better position than the reviewing appellate court to delimit the reasonable range of prior probabilities for the proposition? In some cases, the appellate courts will be in just as good a position to delimit the reasonable range of priors. In others, the trial courts are in a better position, for reasons

---

140. See Suzanna Sherry, *Foundational Facts and Doctrinal Change*, 2011 U. ILL. L. REV. 145, 150 (2011) ("I rely on the basic scientific definition of facts as (at least in theory) falsifiable. This is what distinguishes them from values or policy in the context of judicial decision making.").

141. See *id.* at 146.

142. See *id.*

143. See Borgmann, *Appellate Review*, *supra* note 138, at 1187. As an example of a social fact relevant to a constitutional claim, Professor Borgmann offers: "[W]hether violent video games cause aggression in children." See *id.* at 1187. In the context of social facts in constitutional cases, Professor Borgmann argues that appellate courts should review findings of social fact under the same "clearly erroneous" standard that they apply to questions of adjudicative fact. See *id.* at 1191, 1247.

recognized by scholars in other contexts.<sup>144</sup> This will often depend on the source of the prior.

To illustrate, consider the source of the prior in *Twombly*. The Supreme Court, reviewing a dismissal under Rule 12(b)(6), applied its predictions about how rational firms would act under certain circumstances, drawn from the Court's notions of microeconomic theory about rational firm behavior. The Court did not cite any expert evidence on microeconomics offered in the case, but nonetheless found innocent parallel conduct a more likely explanation for the alleged facts than unlawful conspiracy or combination.<sup>145</sup> For this prior source (an academic understanding of microeconomic theory, not based on expert testimony), reviewing courts may be in just as good a position as the district court to assess the range of reasonable priors. Here, the reliability of the prior source gained little, if anything, from the adversarial process. Different judges may take different views about what a rational firm will do under various conditions, and the trial judge is in no better position than the appellate judge to make such a determination. Where academic theory is the best available source of a prior, the goal of consistency may counsel in favor of more searching appellate review.<sup>146</sup> This would provide guidance to lower courts within the appellate court's binding jurisdiction, and increase the likelihood of reaching similar results in similar cases.

But consider an alternative source of priors—expert testimony from an economist about how rational firms act under certain conditions. Here, the source of the prior is not

---

144. See, e.g., *id.* at 1210-20 (citing efficiency, institutional competence, and the role of the judiciary as rights protector to argue in favor of a clearly erroneous standard of review for social fact determinations).

145. See *supra* Part III, Question 2, sec. A.

146. See generally Borgmann, *Appellate Review*, *supra* note 138, at 1244-48 (recognizing the consistency claim, but contending that *de novo* appellate review of findings of social fact “presents a cure that is worse than the disease, inviting unpoliced appellate-level fact-finding for all cases . . . .”); Allison Orr Larsen, *Factual Precedents*, 162 U. PA. L. REV. 59, 99-101 (2013) (considering efficiency and institutional competence as reasons for a strong conception of *stare decisis* on determinations of generalized questions of fact, but questioning those reasons because factual precedents may not be the product of careful deliberation).

just academic theory. In this situation, the fact-finder observing the testimony may be in a better position to assess prior probabilities because it can better assess the expert's credibility, command of the subject, and responses to cross-examination. In this situation, a more deferential standard of review, such as abuse of discretion, may be preferable. Indeed, the Federal Rules seem to contemplate that such factual determinations will be upheld unless "clearly erroneous."<sup>147</sup> Professor Borgmann argues that trial courts are in a better position to make all such "social fact" determinations, and that appellate courts are susceptible to making factual determinations based on questionable and untested sources, such as scientific or empirical claims advanced in amicus briefs or independent research by judicial clerks or the judges themselves.<sup>148</sup>

When specifying or estimating prior probabilities, courts should take a functional approach. The standard of appellate review for priors should turn on the pragmatic question of whether the fact-finder has an institutional advantage over the appellate court in assessing the prior probabilities, which will in turn depend on the sources of the prior, discussed below.

#### D. *Legislation*

Finally, specifications of priors for purposes of systemic disparate treatment cases might be established by the legislature. Congress could, if there were sufficient empirical evidence and political support, amend Title VII to indicate a range of reasonable prior probabilities for systemic

---

147. See FED. R. CIV. P. 52(a)(6) ("Findings of fact, whether based on oral or other evidence, must not be set aside unless clearly erroneous, and the reviewing court must give due regard to the trial court's opportunity to judge the witnesses' credibility."); see also Borgmann, *Appellate Review*, *supra* note 138, at 1199-1202.

148. See Borgmann, *Appellate Review*, *supra* note 138, at 1215-20; see also Brianne J. Gorod, *The Adversarial Myth: Appellate Court Extra-Record Factfinding*, 61 DUKE L.J. 1, 25-37 (2011) (citing *Citizens United v. FEC*, 130 S. Ct. 876 (2010) as one example of extra-record fact-finding by appellate courts, and noting the importance of factual claims in amicus briefs).

discrimination.<sup>149</sup> At present, with a general lack of reliable empirical studies on the prevalence of employment discrimination and disagreements about the magnitude of the problem of unconscious or structural bias in workplaces, this possibility seems an unlikely scenario. Nonetheless, advances in the developing social science on discrimination could make such legislation a possibility in the future.

There is no formal requirement that Congress have empirical data or engage in legislative fact-finding before passing legislation.<sup>150</sup> However, Professor Wendy Rogovin outlines a number of informal political checks on Congress that compel it to engage in fact finding, “rang[ing] from the self-interested desire for re-election to the civic-minded concern for serving one’s constituents by ascertaining their needs and how best to serve them, to the need to use facts to persuade other lawmakers.”<sup>151</sup> Should the legislature make such a factual determination as a premise to a Title VII amendment, then trial courts and appellate courts would generally not be permitted to disregard such legislative factual determinations.<sup>152</sup> Traditionally, the legislature has been viewed as holding an institutional advantage over

---

149. Regarding congressional fact-finding, and judicial review thereof, see generally Wendy M. Rogovin, *The Politics of Facts: “The Illusion of Certainty,”* 46 HASTINGS. L.J. 1723 (1995). Professor Rogovin distinguishes between political facts (e.g., “the public wants guns kept away from school yards”) and scientific facts (e.g., “the extent to which the presence of guns within 1000 feet of a school yard has a quantifiably significant impact on interstate commerce”). *Id.* at 1741.

150. *See id.* at 1741-42.

151. *Id.* at 1743. Whether courts should defer to such legislative fact-finding is a separate question, addressed in Caitlin E. Borgmann, *Rethinking Judicial Deference to Legislative Fact-Finding*, 84 IND. L.J. 1 (2009) [hereinafter Borgmann, *Judicial Deference*] (arguing that courts should independently review legislative fact-finding where individual rights are affected by the legislation in question).

152. *See* Robert E. Keeton, *Legislative Facts and Similar Things: Deciding Disputed Premise Facts*, 73 MINN. L. REV. 1, 43 (1988) (“Neither a trial court nor an appellate court may set aside or disregard legislative determinations of fact that were the basis for a statute’s enactment, except to the extent that the court determines that, as a matter of constitutional law, the legislature’s reliance on those asserted facts cannot withstand scrutiny.”).

courts in making findings of social fact.<sup>153</sup> Whether this traditional institutional advantage argument holds in every case is the subject of considerable debate well beyond the scope of this Article.<sup>154</sup> The broader point is that the legislature might, at some point in the future, be capable of marshalling sufficient empirical evidence on the prevalence of employment discrimination to make factual determinations that would be useful for interpreting traditional statistical evidence of employment outcome disparities in systemic disparate treatment litigation, and could amend Title VII accordingly.

QUESTION 3: HOW SHOULD COURTS CONVEY THE  
RELATIONSHIP BETWEEN PRIOR PROBABILITIES AND  
STATISTICAL EVIDENCE OF DISPARITY?

A frequent and long-standing challenge to the use of Bayesian statistical analysis in the courtroom is that it will be difficult to convey or confusing to the fact-finder.<sup>155</sup>

---

153. See, e.g., F. Andrew Hessick, *Rethinking the Presumption of Constitutionality*, 85 NOTRE DAME L. REV. 1447, 1473 (2010) (“[L]egislatures are better equipped than the courts to make the sorts of empirical findings relevant to legislation. Legislatures have more resources than courts to gather information—they have large staffs, general subpoena power, and large institutions such as the Congressional Research Service to facilitate their factfinding—and members of the legislature are more likely to be aware of local issues than judges because of the electoral process.”); Kate T. Spelman, *Revising Judicial Review of Legislative Findings of Scientific and Medical “Fact”: A Modified Due Process Approach*, 64 N.Y.U. ANN. SURV. AM. L. 837 (2009) (providing an overview of the traditional view); Note, *Deference to Legislative Fact Determinations in First Amendment Cases After Turner Broadcasting*, 111 HARV. L. REV. 2312, 2315-16 (1998).

154. For interesting challenges to this traditional view, see Neal Devins, *Congressional Factfinding and the Scope of Judicial Review: A Preliminary Analysis*, 50 DUKE L.J. 1169, 1206 (2001) (contending that Congressional fact-finding deserves deference only under certain conditions in which the legislature has the proper incentives to be a reliable fact-finder) and Borgmann, *Judicial Deference*, *supra* note 151, at 35-46.

155. See, e.g., Michael O. Finkelstein & William B. Fairley, *The Continuing Debate Over Mathematics in the Law of Evidence: A Comment on “Trial By Mathematics,”* 84 HARV. L. REV. 1801, 1806 (1971) (recognizing importance of the confusion point and recommending further, controlled study of the question); Lempert, *supra* note 45, at 446 (discussing the confusion critique); Samuel Lindsey et al., *Communicating Statistical DNA Evidence*, 43 JURIMETRICS J. 147,

Incorporating Bayesian statistical reasoning into judicial proceedings is not a novel concept, and on the question of fact-finder confusion, employment discrimination scholars can learn from experiences in other fields of substantive law. Bayesian interpretations of traditional statistical evidence have been urged, and, to some extent accepted by, courts in other legal fields.<sup>156</sup> For example, in paternity disputes and criminal cases involving DNA evidence, experts have advanced, and some courts have allowed, Bayesian interpretations of traditional statistical evidence in order to avoid the transposition fallacy.<sup>157</sup> Drawing upon experience in these other fields, this Question considers possible

---

155-56, 161 (2003) (finding that expressing statistical results in the form of natural frequencies can reduce confusion); Dale A. Nance & Scott B. Morris, *An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Large and Quantifiable Random Match Probability*, 42 JURIMETRICS J. 403, 444-45 (2002) (finding that careful use of Bayesian statistical analysis “can indeed assist the jury in reaching more accurate verdicts, and that it can do so without manipulating or deceiving the jury.”); Laurence H. Tribe, *A Further Critique of Mathematical Proof*, 84 HARV. L. REV. 1810, 1814 (1971) (responding to Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970)); Laurence H. Tribe, *Trial By Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971).

156. References to Bayes’s Theorem or the concept of prior probabilities are rare in reported decisions involving systemic employment discrimination claims. The following search in the ALLCASES Westlaw database returned only 7 results: (BAYES! or “PRIOR PROBABILIT!” or “PRIOR ODDS”) & (“SYSTEMIC DISPARATE TREATMENT” or “DISPARATE IMPACT” or “PATTERN OR PRACTICE”). None of the results were cases in which courts had rigorously applied the Bayesian view articulated herein to a systemic disparate treatment case. By contrast, a significant number of courts have considered the implications of Bayes’s Theorem and prior probabilities in the context of criminal DNA match results and paternity testing. The following search in the ALLCASES Westlaw database returned 114 results: (BAYES! or “PRIOR PROBABILIT!” or “PRIOR ODDS”) & (“DNA MATCH” or “PATERNITY”). See also DAVID H. KAYE ET AL., *THE NEW WIGMORE: A TREATISE ON EVIDENCE: EXPERT EVIDENCE*, § 14.3.1-14.3.2 (2d ed. 2010); Liebman et al., *supra* note 4, at 616.

157. See Liebman et al., *supra* note 4, at 616 (“Nor is it fanciful any longer to contemplate the routine use of Bayes[s] Theorem in criminal cases. As we note above, a combination of DNA and Bayesian analysis presented by experts has revolutionized proof of paternity, including proof of identity in rape cases involving minors or severely disabled nursing-home patients who give birth to a child.”).

methods for conveying statistical evidence in ways that make the operation of priors transparent.

#### A. *Prior Probability Charts*

One possibility for incorporating prior probabilities into systemic disparate treatment litigation is to provide juries with charts that include a range of prior probabilities, along with accompanying posterior probabilities reflecting a proper interpretation of the statistical evidence. Using this method, juries would be encouraged to estimate for themselves the prior likelihood that the defendant engaged in discrimination, based on their assessment of all the nonstatistical evidence in the case (which will necessarily be affected by their intuitions about background rates of employment discrimination). The chart would then indicate the correct posterior probability estimate using Bayes's Theorem to update their prior probability with the new information learned from the statistical evidence. An example of such a chart is available in a study performed by Dale Nance and Scott Morris. The chart used by Nance and Morris demonstrated the effect of a piece of evidence, having a likelihood ratio of 25, on a series of prior probabilities starting from 0% and increasing in 5% increments up to 100%.<sup>158</sup>

This idea is not new. Professor Kaye describes this possibility in the context of DNA match evidence in his evidence treatise, *The New Wigmore*:

[The] expert might present the jury with a table or graph showing how the posterior probability changes as a function of the prior probability. Each juror could consider the other evidence in the case to arrive at prior odds that are personally satisfactory, then use the table to determine the posterior odds or probability.<sup>159</sup>

A variant on this idea is to present a chart or graph that demonstrates the relationship between various prior probabilities and their resulting posterior probabilities as a

---

158. See Nance & Morris, *supra* note 155, at 448 app. B.

159. KAYE ET AL., *supra* note 156, § 14.3.1 (footnotes omitted).

heuristic device, without encouraging jurors to specifically select a particular prior probability.<sup>160</sup>

The introduction of Bayesian concepts into employment discrimination litigation in this way, including prior and posterior probability estimates, may seem somewhat far-fetched, but the use of Bayes's Theorem is not new in other contexts. The presentation of posterior probabilities is actually common in paternity litigation.<sup>161</sup> Unfortunately, courts frequently (and in most cases erroneously) allow experts to assume a prior probability of .5.<sup>162</sup>

Some courts have properly recognized the problem with a fixed .5 prior assumption, and have required experts to provide a range of possible prior probabilities, with their accompanying posterior probabilities. For example, in *Plemel v. Walter*, a paternity case, the Supreme Court of Oregon instructed:

If the expert uses various assumptions [of priors] and makes these assumptions known, the fact[-]finder's attention will be directed to the other evidence in the case, and it will not be misled into adopting the expert's assumption as the correct weight to be assigned to the other evidence. The expert should present calculations based on assumed prior probabilities of 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 percent.<sup>163</sup>

*Plemel* demonstrates that courts, in other contexts, have recognized the limitations of traditional statistical evidence

---

160. *See id.*

161. *See id.* § 12.8.5 n.101 ("The exception is paternity litigation; when genetic tests are indicative of paternity, testimony as to a posterior 'probability of paternity' is common.") (citing 1 MCCORMICK ON EVIDENCE § 211, at 922-23 (Kenneth S. Broun ed., 6th ed. 2006)).

162. *See id.* § 14.3.2.b (discussing paternity cases that invoke the "equally likely random man assumption" to justify use a prior probability of .5). Sometimes this .5 prior probability assumption is even left unstated, on the basis that it is a "neutral" starting point. *See id.* As Professor Kaye notes, and as some courts have recognized, the .5 prior probability assumption in a paternity case is not a "neutral" starting point at all, but rather suggests that the individual before the court is just as likely to be the father as all the other men in the relevant population combined. *See id.* ("For example, if there are 100,000 other men in the jurisdiction, then *all of them combined* have a probability of one-half of being the father.").

163. *Plemel v. Walter*, 735 P.2d 1209, 1219 (Or. 1987).

and have avoided the transposition fallacy by insisting on the conveyance of statistical evidence in a way that highlights the role of prior probabilities.<sup>164</sup>

### B. *Prior Probability Decision Thresholds*

Other possibilities exist. Rather than providing jurors with a chart of priors and accompanying posterior probabilities, an expert might simply describe a threshold cut-off point estimate for the prior, using as the decision rule any posterior probability  $>.5$ . In other words, a statistical expert could conceivably use Bayes's Theorem to work backwards from the preponderance of the evidence standard (in mathematical terms, the posterior  $\text{Prob}(A|B) >.5$ ), to arrive at the prior probability needed, when combined with the statistical evidence, to cross the burden of persuasion threshold.<sup>165</sup> In the simplified example presented in Part III, Question 2, Section A.2, a statistical analysis found a  $p$ -value of exactly .05. Using the preponderance of the evidence ( $>.5$ ) decision rule, the statistical expert could work backwards and explain to the jury that if it believed, based on its assessment of only the nonstatistical evidence in the case (importantly, including any social framework evidence or similar evidence bearing on background rates), that the likelihood of discrimination was some number exceeding 5%, then the jury should interpret the statistical evidence as the observance of an outcome disparity sufficient to make it more likely than not that the employer discriminated. On the other

---

164. See D.H. Kaye, Plemel as a Primer on Proving Paternity, 24 WILLAMETTE L. REV. 867, 867-69 (1988).

165. Here, the implementation of a Bayesian view of statistical evidence in civil cases has a distinct advantage over the use of such techniques in criminal cases. While there is general agreement that the "preponderance of the evidence" burden of persuasion in a typical civil case can be translated into probability terms as  $>.5$ , the same cannot be said for quantifying the "beyond a reasonable doubt" standard in criminal cases. See generally Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385, 394 (1985) (noting general agreement regarding the quantification of the "preponderance of the evidence" standard at  $>.5$ , but arguing that a confidence interval approach is preferable to use of single point estimates); Andrea Roth, *Safety in Numbers? Deciding When DNA Alone Is Enough to Convict*, 85 N.Y.U. L. REV. 1130, 1157 (2010) (describing various concerns with attempts to quantify the "beyond a reasonable doubt" criminal standard).

hand, if, considering all the nonstatistical evidence, the jury believed that the likelihood of discrimination was exactly 5% or less, then the jury should interpret that statistical evidence as insufficient to conclude that it is more likely than not that the employer discriminated.

### C. *Likelihood Ratios*

An alternative presentation of probabilities that can help avoid the transposition fallacy is a simple presentation of the likelihood ratio, rather than a presentation of *p*-values. The likelihood ratio is an expression of the relationship between the likelihood of seeing particular evidence (here, the statistical disparity observed in employment outcomes) in the case of discrimination, as compared to the likelihood of seeing that same evidence in the case where the defendant did not discriminate.

In the criminal context, Norman Fenton and Martin Neil described the likelihood ratio as follows:

For any piece of evidence *E*, the likelihood ratio of *E* is the probability of seeing that evidence if the defendant is guilty divided by the probability of seeing that evidence if the defendant is not guilty. It follows directly from Bayes Theorem that if the likelihood ratio is bigger than 1 then the evidence increases the probability of guilt (with higher values leading to higher probability of guilt) while if it is less than 1 it decreases the probability of guilt (and the closer it gets to zero the lower the probability of guilt).<sup>166</sup>

To relate this to the systemic discrimination context, a statistical expert could present the statistical evidence of employment outcome disparities in the form of a likelihood ratio. If the employment outcome disparity is such that we would be equally as likely to observe the disparity if the employer had been discriminating as if the employer had not been discriminating, then the likelihood ratio would equal 1. This would mean that the statistical evidence offered no real value in helping to decide whether the employer unlawfully discriminated.<sup>167</sup> On the other hand, if the probability of observing a given disparity if the employer had hired at

---

166. Fenton & Neil, *Probabilistic Reasoning*, *supra* note 45, at 9.

167. *See id.* ("If the likelihood ratio is equal to or close to 1 then *E* offers no real value at all since it neither increases nor decreases the probability of guilt.").

random were 1000 to 1, and the probability of observing that same disparity if the employer had discriminated were 1, then we would know that the statistical evidence has made it 1000 times more likely that the employer was a discriminator than it had been before considering the statistical evidence.<sup>168</sup> And, importantly, this would be true *regardless* of the prior selected by a jury or judge, and yet it still manages to convey the statistical information in a way that prevents misinterpretation of *p*-values.<sup>169</sup>

Fenton and Neil cite a specific example of likelihood ratios being used to avoid the transposition fallacy in a criminal case, *R. v. Barry George*.<sup>170</sup> There, the defendant had been convicted of murdering TV presenter Jill Dando, based in part on forensic gunpowder evidence.<sup>171</sup> At the original trial, the only probability evidence offered was the probability of finding the evidence, given an assumption that the defendant was not guilty (.01).<sup>172</sup> On appeal in *Barry George*, expert statistical testimony was offered, showing that the likelihood ratio of the gunpowder evidence was approximately equal to 1. That is, the likelihood of observing the gunpowder evidence assuming the defendant's guilt was also equal to .01.<sup>173</sup> The gunpowder evidence actually offered no real probative value.

Fenton and Neil note a concern about the use of likelihood ratios—they tend to be confusing to lawyers, judges, and jurors for the same reason that a full explication

---

168. *See id.* (using a similar example in the criminal context).

169. *See id.* at 9-10 (“Another advantage of using the likelihood ratio is that it removes one of the most commonly cited objections to Bayes Theorem, namely the obligation to consider a prior probability for a hypothesis like ‘guilty’ . . .”).

170. *See id.* at 9 (citing *R. v. George*, [2007] EWCA (Crim) 2722, [1] (Eng.)).

171. *See id.*; *see also George Not Guilty of Dando Murder*, BBC NEWS (Aug. 1, 2008, 10:47 PM), [http://news.bbc.co.uk/2/hi/uk\\_news/7536815.stm](http://news.bbc.co.uk/2/hi/uk_news/7536815.stm).

172. *R. v. Barry George*, [2007] EWCA (Crim) 2722, [23], [44] (Eng.). This probability evidence is analogous to the *p*-value evidence offered in systemic discrimination cases: the probability of observing a statistical disparity as large as was observed, given an assumption of hiring at random.

173. *See Fenton & Neil, Probabilistic Reasoning, supra* note 45, at 9 (“This is because both  $P(E \mid \text{Guilty})$  and  $P(E \mid \text{not Guilty})$  were approximately equal to 0.01.”).

of the mathematical formula underpinning Bayes Theorem tends to be confusing.<sup>174</sup> While likelihood ratios hold promise, due to their ability to avoid the express selection of a prior distribution, they may be difficult for judges and juries to digest.<sup>175</sup>

#### D. *Visual Representations*

In some circumstances, visual representations of frequencies may work better than charts or likelihood ratios to convey the probative value of statistical evidence in a way that avoids the transposition fallacy. Fenton and Neil explain that even “highly intelligent barristers, judges and surgeons” have a tendency to “simply switch-off at the sight of a formula,” making it impossible for them to follow explanations of Bayes Theorem or likelihood ratios.<sup>176</sup> Fenton and Neil argue that the underlying formulaic workings of Bayes’s Theorem, or the “first principles” of Bayesian reasoning, need not be presented to juries.<sup>177</sup> Rather, Fenton and Neil propose showing jurors visual explanations or event trees that illustrate frequencies. They offer a relatively simple event tree that illustrates the application of Bayes’s Theorem in a DNA matching case.<sup>178</sup> A similar event tree could be constructed for the classic Harvard Medical School example, set out in the Introduction of this Article. That example is presented below as Diagram A.

---

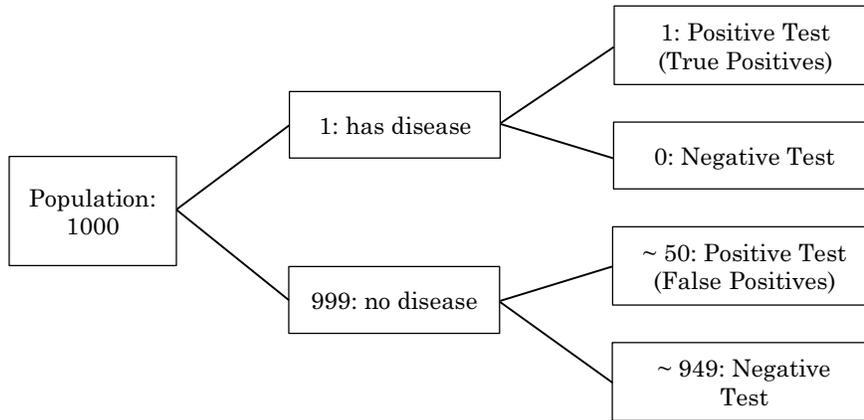
174. *See id.* at 10.

175. Others have similarly questioned whether jurors can understand likelihood ratios. *See* Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859, 877 (1996) (“Even when likelihood ratios are properly conveyed, there is little reason to believe that jurors will understand what they mean and how they should be used. Although they have scientific merit, likelihood ratios—which are the ratios of conditional probabilities—are not easy to understand.”).

176. Fenton & Neil, *Probabilistic Reasoning*, *supra* note 45, at 11.

177. *See id.* at 13-14.

178. *See id.* at 6 fig. 2. Fenton and Neil also construct more elaborate visual representations attempting to show three-dimensional depth, as well as visual representations that permit multi-step Bayesian updating. *See id.*



**Diagram A**

For every 1 true positive test, there are approximately 50 false positive tests. For any randomly selected individual that tests positive, the probability that the individual actually has the disease is approximately 2%.<sup>179</sup>

As is readily apparent from Diagram A, the use of visual representations of natural frequencies may be useful in helping even judges and juries with little mathematical aptitude avoid the transposition fallacy.<sup>180</sup> A similar visual representation of statistical evidence could be prepared by statistical experts in systemic employment discrimination cases. Such visual representations could avoid the misinterpretation of *p*-values, and could assist the court and the fact-finder in understanding the probative value of statistical evidence of employment outcome disparities.

Of course, in the Harvard Medical School study the prior is assumed, and can be quantified, as 1 in 1000.<sup>181</sup> In the employment discrimination context, the prior is obviously not

---

179. Based on the research question presented to subjects in Casscells et al., *supra* note 4, at 999-1000.

180. This is generally consistent with the findings of Lindsey et al., *supra* note 155, at 156, regarding the presentation of natural frequencies to avoid juror confusion.

181. See Casscells et al., *supra* note 4, at 999.

known. But this does not undermine the usefulness of visual representations of frequencies. As Fenton and Neil point out, the jury can be presented with a variety of assumptions regarding the prior probabilities, where there is genuine and acknowledged uncertainty, and experts can run a variety of such visual models for various assumptions.<sup>182</sup> Fenton and Neil liken this approach to an electronic calculator performing long division for the jurors, without the jurors necessarily needing to understand all of the inner processes of the calculator.<sup>183</sup>

#### QUESTION 4: WHAT ARE LEGITIMATE SOURCES OF PRIORS?

Where should priors come from? Priors are necessarily subjective, at least to some degree. But not all priors are created equally. In the Bayesian view, prior beliefs are revised over time as more information is added to the decision-maker's base of knowledge. In this sense, some priors may be more well-informed or more "objective"<sup>184</sup> than others. This Question briefly examines some possible sources of priors. The point is not to develop a definitive ordinal ranking of sources in terms of objectiveness, but rather to show that exposing the existence of priors will allow judges, reviewing courts, and the public to examine the reasonableness of priors. Making prior sources transparent will allow for more effective structural and political checks on fact-finders and courts.

---

182. See Fenton & Neil, *Probabilistic Reasoning*, *supra* note 45, at 20-21 (explaining that Bayesian visual models can be run with different sets of assumptions, and that the jury would decide whether the assumptions were reasonable).

183. *Id.* at 13.

184. A subset of Bayesian statistical thinking, known as "objective Bayesian analysis" emphasizes the selection of prior distributions by using available objective, empirical data. See generally James Berger, *The Case for Objective Bayesian Analysis*, 1 BAYESIAN ANALYSIS 385 (2006) (addressing the debate as to the value of objective versus subjective Bayesian analysis).

### A. *Judicial Notice*

One source of priors is information from which a court would be permitted to take judicial notice of the fact's existence. Federal Rule of Evidence 201(b) provides a workable definition of when priors could be adopted by judicial notice. Under Rule 201(b), a court "may judicially notice a fact that is not subject to reasonable dispute because it: (1) is generally known within the trial court's territorial jurisdiction; or (2) can be accurately and readily determined from sources whose accuracy cannot reasonably be questioned."<sup>185</sup>

Scenarios involving judicial notice of a prior might include the extreme examples, offered by Justice Souter in his *Iqbal* dissent, of "little green men, or the plaintiff's recent trip to Pluto, or experiences in time travel."<sup>186</sup> In those cases, the range of reasonable prior probabilities might include only 0, or some number approaching 0.<sup>187</sup> At present, in the discrimination context, reasonable people can hold vastly different preconceptions about the prior likelihood of discrimination,<sup>188</sup> making judicial notice generally improper as a source of prior probabilities for systemic disparate treatment cases.

### B. *Empirical Evidence*

Another potential source for developing a prior is empirical evidence. For example, in the employment discrimination context, the court or fact-finder could consider any non-case-specific empirical studies that have attempted to discern the base rate of employment discrimination. Elsewhere, I have discussed some potential empirical studies that could shed light on base rates of employment

---

185. FED. R. EVID. 201(b).

186. *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1959 (2009).

187. For potential shortcomings of the Federal Rule of Evidence's description of what may be judicially noticed, see Christopher Onstott, *Judicial Notice and the Law's "Scientific" Search for Truth*, 40 AKRON L. REV. 465, 483-87 (2007).

188. See Weiss, *supra* note 44, at 1691.

discrimination.<sup>189</sup> These might include studies of litigant success in employment discrimination litigation.<sup>190</sup>

For some priors, empirical studies may provide a robust and relatively reliable source of information about the reasonable range of potential prior probabilities. Robust empirical data would perhaps represent the ideal source of priors for objective Bayesian analysis. For other questions, the nature of the proposition at issue may mean that empirical studies are rare, deeply flawed, or non-existent—at least given currently available methodologies and technology. In these cases, available empirical studies are likely to be of limited or questionable value in setting a range of reasonable priors.

### C. *Expert Testimony (Including Contextualist Evidence)*

Another source of information for setting priors would be expert testimony. This could include an expert's report on empirical research into the proposition directly at issue, which would place it in the preceding category. Alternatively, this could involve expert testimony about more generalized, non-case-specific social science evidence that can help describe the context in which the prior probability range should be assessed.

This type of expert testimony would not attempt to directly pinpoint an estimated base rate, but instead would attempt to “adjust” otherwise erroneous or likely mistaken prior preconceptions by identifying specific and systematic failings in our ability to accurately assess priors. One example is the type of social framework evidence that has been offered in systemic employment discrimination cases. Social framework evidence may include “general social science research [that] can provide a valuable context for

---

189. See Bent, *Telltale Sign*, *supra* note 52, at 834-36 (suggesting the courts could consider empirical studies of employment discrimination litigation, including Kevin M. Clermont & Stewart J. Schwab, *Employment Discrimination Plaintiffs in Federal Court: From Bad to Worse?*, 3 HARV. L. & POL'Y REV. 103 (2009) and Kevin M. Clermont & Stewart J. Schwab, *How Employment Discrimination Plaintiffs Fare in Federal Court*, 1 J. EMPIRICAL LEGAL STUD. 429 (2004)).

190. See *supra* note 181.

deciding case-specific factual issues.”<sup>191</sup> The purpose of social framework evidence is to “help[ ] educate fact-finders about the conditions under which gender stereotypes and prejudice are likely to influence impressions, evaluations, and behavior in social and organizational settings.”<sup>192</sup>

In *Wal-Mart*, Dr. William Bielby, a sociologist, testified that “social science research demonstrates that gender stereotypes are especially likely to influence personnel decisions when they are based on subjective factors, because substantial decision-maker discretion tends to allow people to seek out and retain stereotyping-confirming information and ignore or minimize information that defies stereotypes.”<sup>193</sup> Dr. Bielby also went one (controversial) step further, and compared the actual practices at Wal-Mart to this general social science backdrop, concluding that Wal-Mart’s “personnel policies and practices make pay and promotion decisions vulnerable to gender bias.”<sup>194</sup>

The Court’s opinion in *Wal-Mart* disregarded Dr. Bielby’s testimony, stating: “Bielby’s testimony does nothing to advance respondents’ case.”<sup>195</sup> The Court reached this conclusion because of Dr. Bielby’s inability to make case-specific conclusions. The Court noted that “Dr. Bielby conceded that he could not calculate whether 0.5 percent or 95 percent of the employment decisions at Wal-Mart might be determined by stereotyped thinking.”<sup>196</sup> Because Dr. Bielby had no answer to this “essential question,” the Court dismissed his testimony as “worlds away from ‘significant

---

191. John Monahan et al., *The Limits of Social Framework Evidence*, 8 L. PROBABILITY & RISK 307, 308 (2009) [hereinafter Monahan et al., *The Limits*].

192. Melissa Hart & Paul M. Secunda, *A Matter of Context: Social Framework Evidence in Employment Discrimination Class Actions*, 78 FORDHAM L. REV. 37, 53 (2009) (quoting Susan T. Fiske & Eugene Borgida, *Social Framework Analysis as Expert Testimony in Sexual Harassment Suits*, in SEXUAL HARASSMENT IN THE WORKPLACE: PROCEEDINGS OF NEW YORK UNIVERSITY 51ST ANNUAL CONFERENCE ON LABOR 575, 579 (Samuel Estreicher ed., 1999)).

193. *Dukes v. Wal-Mart Stores, Inc.*, 603 F.3d 571, 601 (9th Cir. 2010), *rev’d*, 131 S. Ct. 2541 (2011).

194. *See id.*

195. *Wal-Mart*, 131 S. Ct. at 2554.

196. *Id.* at 2553.

proof that Wal-Mart ‘operated under a general policy of discrimination.’”<sup>197</sup>

But the Court’s analysis of Dr. Bielby’s expert testimony overlooked the importance of prior probabilities and thereby misconceived the role of social framework testimony. Priors are crucial in interpreting statistical evidence of disparity, and a social framework expert’s *non-case-specific* testimony can inform or adjust a prior probability. Scholars critical of Dr. Bielby’s testimony in *Wal-Mart* object to his attempt to link general social science research to specific practices or procedures at Wal-Mart stores, without following a scientific methodology.<sup>198</sup> They are not generally opposed to the introduction and consideration of general social science evidence of background rates of discrimination or the circumstances under which stereotyping or discrimination is more or less likely to occur.<sup>199</sup> Acknowledgement of prior probabilities brings the proper role of social framework evidence into clear focus: it can help courts or fact-finders identify ranges of reasonable priors, or help them adjust misconceived, unsupported assumptions of prior probabilities. A social framework expert need not link background social science findings to case-specific information. Contextual information is still valuable because it can help set or adjust the range of reasonable priors.

#### D. *Academic Theory*

Another prior source is academic theory, as illustrated in *Twombly*. There, the Court predicted rational firm behavior

---

197. *Id.* at 2554.

198. See Tristin K. Green, *The Future of Systemic Disparate Treatment Law*, 32 BERKELEY J. EMP. & LAB. L. 395, 445 n.205 (2011); John Monahan et al., *Contextual Evidence of Gender Discrimination: The Ascendance of “Social Frameworks,”* 94 VA. L. REV. 1715, 1743 (2008) [hereinafter Monahan et al., *Contextual Evidence*]; compare Monahan et al., *The Limits*, *supra* note 191 (arguing that “general research findings cannot be linked by an expert witness to the facts of a specific case.”), with Hart & Secunda, *supra* note 192 (arguing that experts should be permitted to “link[ ] a field of knowledge to the case facts” at least in some circumstances).

199. See Monahan et al., *Contextual Evidence*, *supra* note 198, at 1743; Monahan et al., *The Limits*, *supra* note 191, at 308, 319.

based on its understanding of microeconomic theory. The *Twombly* Court “concluded that [the complaint] did not plausibly suggest an illicit accord because it was not only compatible with, but indeed was more likely explained by, lawful, unchoreographed free-market behavior.”<sup>200</sup> The *Twombly* Court did not cite any empirical or expert evidence to substantiate this claim, but rather relied on economic theory to predict expected wealth-maximizing behavior.<sup>201</sup>

A similar assumption based on economic theory was central to the summary judgment decision in *Matsushita*. As Professor Michael Kaufman describes *Matsushita*:

In reaching its result, the Supreme Court presumed that businesses naturally engage in rational wealth-maximizing behavior. The Court next concluded that no rational wealth-maximizing institution would enter into a long-term agreement to set below-cost pricing because no institution would expect to recover substantial short-term losses with uncertain long-term gains or would trust their coconspirators to maintain their agreement.<sup>202</sup>

In *Matsushita*, the Court cited to a textbook as the basis for the theory predicting rational behavior: Robert Bork’s *The Antitrust Paradox: A Policy at War with Itself*.<sup>203</sup> Whether stated or unstated, both cases involved an underlying assumption about how the world works, based on microeconomic theory and “a strong presumption that

200. See *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1950 (2009) (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 567 (2007)).

201. The Court did cite an Antitrust Law text for the proposition that conscious, interdependent parallelism does not establish an unlawful conspiracy in violation of Section 1 of the Sherman Act. See *Twombly*, 550 U.S. at 553-54 (citing 6 P. Areeda & H. Hovenkamp, *ANTITRUST LAW*, ¶ 1433a, at 236 (2d ed. 2003)). In dissent, Justice Stevens challenged the majority’s implicit assumption by quoting Adam Smith. See *Twombly*, 550 U.S. at 591 (Stevens, J., dissenting) (quoting Adam Smith, *An Inquiry Into the Nature and Causes of the Wealth of Nations*, in 39 GREAT BOOKS OF THE WESTERN WORLD 1, 55 (Robert Maynard Hutchins & Mortimer J. Adler eds., 1952)); see also Kaufman, *supra* note 124, at 603 n.47.

202. Kaufman, *supra* note 124, at 598-99.

203. *Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574, 589 (1986) (citing ROBERT H. BORK, *THE ANTITRUST PARADOX: A POLICY AT WAR WITH ITSELF* 144 (1978)).

persons and businesses make purely rational choices with a singular intent to maximize their wealth.”<sup>204</sup>

Whether the theoretical prediction of rational, wealth-maximizing behavior is consistent with what individuals and firms actually do in the real world is a question that could in theory be empirically tested. Firms either do or do not act according to this theoretical prediction. Recent critiques from the field of psychology cast serious doubt on whether individuals and firms act rationally. Developments in psychology, and the related field of behavioral economics, for example, call into question assumptions about purely rational, wealth-maximizing behavior.<sup>205</sup>

The point is not to debate the merits of classical economic assumptions versus behavioral economics. The point is that courts sometimes, as in *Matsushita* and *Twombly*, rely on prior assumptions rooted in academic theory.<sup>206</sup> If prior probabilities were made transparent, such predeterminations could be subjected to critique, evaluation, and challenge by litigants, by reviewing courts, and by the public.

#### E. *Otherwise Unsupported Estimates*

The last category of prior source needs little explanation. Prior probabilities could simply be unsupported guesses, estimates, or assertions about the state of the world, based on idiosyncratic beliefs held by judges or fact-finders, and not based on any evidence in the record. An arguable example is Justice Scalia’s prediction in *Wal-Mart* about how “most managers in any corporation” would make promotion decisions:

[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

---

204. Kaufman, *supra* note 124, at 595.

205. *See id.* at 611-20 (detailing critiques based in “neuroscience, psychology, behavioral economics, and hedonics”).

206. *See supra* Part III, Question 3, secs. A–B.

that produce disparate impact—such as scores on general aptitude tests or educational achievements . . . . And still other managers may be guilty of intentional 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.<sup>207</sup>

Justice Scalia cited no empirical support, no expert testimony or report, no academic theory, nor any other evidence inside or outside of the record on appeal to support this assertion.<sup>208</sup> Yet, this statement earned the votes of five Supreme Court justices.<sup>209</sup> It could be that the majority believed this assertion to be true because it is believed to be consistent with purely rational, wealth-maximizing behavior by managers. But the majority did not identify economic theory as the source for this presumption.<sup>210</sup>

A similarly unsupported prior appeared to be at work in *Iqbal*, where the Court assumed, without citing a source, that defendants’ desire to “keep suspected terrorists in the most secure conditions available until the suspects could be cleared of terrorist activity” was a more likely explanation for the observed events than unlawful discrimination.<sup>211</sup> Again, this represents a statement of fact—a statement about which of two factual conditions is more likely than the other.

To the extent that priors are based on unsupported estimates, guesses, or hunches, this is probably a task best left to juries rather than to trial judges, appellate judges, or even Supreme Court justices. Juries, unlike judges, are selected more or less randomly from the general population.<sup>212</sup> This gives them an institutional advantage over judges in estimating questions about the prevalence of discrimination in America’s workplaces. Empirical evidence

---

207. *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2554 (2011); *see also* Weiss, *supra* note 44, at 1687.

208. *See Wal-Mart*, 131 S. Ct. at 2554.

209. *See id.* at 2546.

210. *See id.*

211. *Ashcroft v. Iqbal*, 129 S. Ct. 1937, 1952 (2009).

212. *See* Douglas G. Smith, *Structural and Functional Aspects of the Jury: Comparative Analysis and Proposals for Reform*, 48 ALA. L. REV. 441, 479 (1997).

has shown that, in discrimination cases, the minority status of the presiding judge is related to the likelihood of the claim's success.<sup>213</sup> There is similar evidence that a judge's gender matters for outcomes in sexual harassment cases.<sup>214</sup> And there is little diversity currently on the federal bench.<sup>215</sup> Juries, on the other hand, are composed of more than a single individual, and therefore tend to bring a wider diversity of backgrounds. While any individual juror may be biased in the same way an individual judge might be, the jury as a group has the benefit of deliberation.<sup>216</sup> Moreover, the decisions of a diverse jury can bring increased public confidence in outcomes.<sup>217</sup> The institutional advantages and added

---

213. See Pat K. Chew & Robert E. Kelley, *Myth of the Color-Blind Judge: An Empirical Analysis of Racial Harassment Cases*, 86 WASH. U. L. REV. 1117, 1121, 1156 (2009); Pat K. Chew & Robert E. Kelley, *The Realism of Race in Judicial Decision Making: An Empirical Analysis of Plaintiffs' Race and Judges' Race*, 28 HARV. J. ON RACIAL & ETHNIC JUST. 91, 94, 103 (2012); Jill D. Weinberg & Laura Beth Nielsen, *Examining Empathy: Discrimination, Experience, and Judicial Decisionmaking*, 85 S. CAL. L. REV. 313, 349 (2012); see also Thomas B. Colby, *In Defense of Judicial Empathy*, 96 MINN. L. REV. 1944, 2001-05 (2012) (noting that diversity on the bench can help judges overcome "empathic blind spots").

214. See SALLY J. KENNEY, *GENDER AND JUSTICE: WHY WOMEN IN THE JUDICIARY REALLY MATTER* 28 (2013); Christina L. Boyd et al., *Untangling the Causal Effect of Sex on Judging*, 54 AM. J. POL. SCI. 389, 401 (2010); Pat K. Chew, *Judges' Gender and Employment Discrimination Cases: Emerging Evidence-Based Empirical Conclusions*, 14 J. GENDER RACE & JUST. 359, 366 (2011).

215. See Weinberg & Nielsen, *supra* note 213, at 347-48 ("According to the Federal Judicial Center, diversity within the federal judiciary is minimal. Only 136 of the 597 active federal district judges are a member of a racial/ethnic minority, roughly 22.8 percent. This figure is slightly larger for female judges; 174 federal district court judges are women—roughly 29 percent.").

216. See, e.g., Cheryl L. Wade, *When Judges are Gatekeepers: Democracy, Morality, Status, and Empathy in Duty Decisions (Help from Ordinary Citizens)*, 80 MARQ. L. REV. 1, 58-59 (1996) (noting the advantage of deliberation, but recognizing that such advantage may be limited, in light of empirical studies showing that jurors reach initial decisions before deliberation); Joshua Wilkenfeld, *Newly Compelling: Reexamining Judicial Construction of Juries in the Aftermath of Grutter v. Bollinger*, 104 COLUM. L. REV. 2291, 2311 (2004) ("Yet another set of improvements in functionality arises in juries from an improved climate for communication. Social science research has established that diverse groups may be more likely to communicate effectively and creatively.").

217. See generally Nancy J. King, *Racial Jurymandering: Cancer or Cure? A Contemporary Review of Affirmative Action in Jury Selection*, 68 N.Y.U. L. REV. 707, 767 (1993) (considering racial diversity in criminal cases); Wilkenfeld, *supra* note 216, at 2314 ("Communities view demographically balanced juries as more

legitimacy of juries suggests that, where guesses or estimates without empirical or academic support serve as the basis for priors, juries should be doing the guessing—not Supreme Court justices.

#### IV. A BAYESIAN REVOLUTION FOR ANTIDISCRIMINATION LAW?

*Hidden Priors* makes the case that Bayesian analysis *should be* adopted for systemic disparate treatment law,<sup>218</sup> and the foregoing makes the case that Bayesian analysis, including the estimation of prior probabilities, *can be* successfully implemented in systemic disparate treatment cases.<sup>219</sup> A further question is whether this Bayesian revolution should be limited to only systemic disparate treatment law, or whether it should instead apply to all employment discrimination theories. In *The Impossibility of Agnostic Discrimination Law*, Deborah Weiss argues that all of employment discrimination law is necessarily influenced by prior preconceptions about background rates of discrimination,<sup>220</sup> and therefore an expressly Bayesian approach should be adopted in all discrimination doctrine, including in the formulation of evidentiary devices used in individual disparate treatment cases.<sup>221</sup>

A form of Bayesian reasoning probably underlies most rational decision-making under uncertainty, regardless of the decision-maker. Judge Richard Posner, in his book *How Judges Think*, eloquently put the point this way: “Bayesian theory is a way of systematizing the elementary point that preconceptions play a role in rational thought.”<sup>222</sup> Jurors unquestionably take their views of the world with them into any deliberations. But this observation does not necessarily mean that juries in all discrimination cases, including in

---

legitimate than those that are unreflective. The presence of balanced juries, then, engenders greater respect for the justice system and its verdicts.”)

218. See Bent, *Hidden Priors*, *supra* note 3, at 810-11.

219. See *supra* Parts I-III.

220. See Weiss, *supra* note 44, at 1679, 1694-96 (explaining the role of background rates of discrimination in the evaluation of nonstatistical, informal evidence typically used in individual disparate treatment cases).

221. See *id.* at 1694-96, 1731-32, 1742.

222. RICHARD A. POSNER, *HOW JUDGES THINK* 67 (2008).

individual disparate treatment cases, should have Bayes's Theorem presented to them so that they can mathematically understand how to value their priors. Nor does it mean that appellate courts should, in the absence of reliable empirical evidence on background rates of discrimination, establish evidentiary devices based on the Supreme Court's (or the Circuit Court of Appeals') best guesses about background rates of discrimination. Rather, where unsupported guesses or other weak evidence on background rates are the best available evidence, juries should be allowed to reach inconsistent outcomes attributable to disparate, yet reasonable, estimates of priors.<sup>223</sup> Bayes's Theorem should be used sparingly, and only when necessary to guard against the misuse of traditional statistical evidence. Jurors intuitively factor in their own priors in a nonmathematical fashion when viewing evidence in individual cases,<sup>224</sup> and attorneys often have the opportunity to strike jurors who seem too tightly tethered to their priors.<sup>225</sup> No introduction of mathematical formulas or visual representations of Bayes's Theorem is necessary in these individual cases.<sup>226</sup>

---

223. Although I am in general agreement with Professor Weiss that background assumptions about discrimination do influence individual discrimination cases just as they do systemic cases, I part ways with her on the need for additional evidentiary devices prescribed by appellate courts to combat inconsistency in outcomes at the trial level. *See* Weiss, *supra* note 44, at 1742 (urging the Supreme Court to develop evidentiary devices based on social framework evidence, because decisions made by triers of fact in the absence of such evidentiary devices are "hopelessly inconsistent").

224. *See id.* at 1679 ("[T]o the extent that a finding of discrimination is left to the trier of fact, background assumptions are merely an unstated factor determining the trier's conclusions.").

225. *See generally* Valerie P. Hans & Alayna Jehle, *Avoid Bald Men and People with Green Socks? Other Ways to Improve the Voir Dire Process in Jury Selection*, 78 CHI.-KENT L. REV. 1179, 1183-84 (2003) (summarizing voir dire procedures in state and federal courts).

226. Professor Weiss appears to agree that the open litigation of background rates of discrimination in every individual disparate treatment case may not be appropriate, as that would "turn each and every discrimination trial into a forum for examining the pattern of discrimination in the United States today." Weiss, *supra* note 44, at 1679. Instead, she argues for reinvigorating the use of evidentiary devices set by appellate courts in individual disparate treatment discrimination cases, and basing those evidentiary devices on background rates

The problem starts with the first mention of *p*-values or statistical significance in a systemic case. Once traditional hypothesis testing statistics are offered into evidence, or are presented to the judge in connection with dispositive pretrial motions, the possibility for misuse and misunderstanding appear.<sup>227</sup> At this point, the transposition fallacy predicted in the *Nature* article becomes a very real risk.<sup>228</sup> When even scientists and doctors cannot necessarily be trusted to properly interpret *p*-values, judges and jurors have little chance. At this point, a Bayesian presentation—whether in the form of prior probability charts, a Bayesian decision threshold, a likelihood ratio, or a visual representation—becomes necessary to avoid the transposition fallacy. Thus, the Bayesian revolution that I advocate is a limited one. It calls for a Bayesian analysis only when statistical evidence of employment outcome disparities are offered to establish disparate treatment employment discrimination.<sup>229</sup>

There is much more for attorneys, scholars, and courts to do to before this Bayesian revolution in systemic disparate treatment law can become a reality. Scholars must continue to devote attention to the vexing second-order questions discussed above. This Article presented only one view on these questions, and nobody has yet advanced a fully satisfactory answer to the persistent challenge of devising

---

that are determined by reference to social framework evidence. *See id.* at 1731-32.

227. *See supra* Introduction; *see also* Bent, *Hidden Priors*, *supra* note 3, at 820-23.

228. *See* Nuzzo, *supra* note 1, at 151-52; *supra* note 9 and accompanying text.

229. This may extend to certain, limited individual disparate treatment cases where statistical evidence is admissible to establish an individual case. There is currently a split of authority on the question whether individual disparate treatment plaintiffs may rely on statistical evidence of disparities to establish a prima facie case of discrimination. *See* Bent, *Telltale Sign*, *supra* note 52, at 812-14 (discussing circuit split on this question and citing cases). Even where statistical evidence cannot be used to establish an individual's prima facie case of discrimination, a number of courts nonetheless permit the use of statistical evidence to help prove the pretext portion of an individual disparate treatment claim. *See id.* at 813 (citing *Mendelsohn v. Sprint/United Mgmt. Co.*, 466 F.3d 1223, 1227 n.2 (10th Cir. 2006), *Bell v. E.P.A.*, 232 F.3d 546, 553 (7th Cir. 2000), and *Lowery v. Circuit City Stores, Inc.*, 158 F.3d 742, 761 (4th Cir. 1998), *vacated on other grounds*, 527 U.S. 1031 (1999)).

the best method for specification of prior probabilities. Courts must be open to considering Bayesian arguments in discrimination cases, as some have been in the paternity testing and DNA matching contexts.<sup>230</sup>

Finally, attorneys must be willing to offer a Bayesian view of employment statistics in a systemic discrimination case. To my knowledge, this has not yet happened, despite numerous opportunities, including the several regional class actions against Wal-Mart that were filed following the Supreme Court's decision in *Wal-Mart*.<sup>231</sup> Why it has not happened remains a bit of a mystery, ripe for further exploration by scholars. One possibility is that both plaintiffs and defendants are satisfied with the predictable battle of experts that occurs in systemic discrimination cases involving traditional hypothesis testing statistics. A common scenario involves the plaintiffs defining the relevant population for study quite broadly, increasing the overall sample size, and thereby making a finding of statistical

---

230. See *supra* Part III, Question 3, secs. A–D.

231. See Plaintiffs' Fourth Amended Complaint ¶¶ 58-60, 70, *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137 (N.D. Cal. 2004) (No. C-01-2252-CRB); Plaintiffs' Opposition to Wal-Mart's Motion to Dismiss Fourth Amended Complaint at 3, 20-22, *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137 (N.D. Cal. 2004) (No. C-01-2252-CRB); Declaration of Dr. Richard Drogin in Support of Plaintiffs' Motion for Class Certification ¶ 74, *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137 (N.D. Cal. 2004) (No. C-01-2252-CRB) (describing regression analyses performed at various levels and reporting the conclusion: "The shortfall in female earnings, pay rates, and promotion rates have a high degree of statistical significance."); Declaration of Marc Bendick Jr., Ph.D. in Support of Plaintiffs' Motion for Class Certification ¶ 43, *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137 (N.D. Cal. 2004) (No. C-01-2252-CRB) (describing "statistically significant" results from a standard deviation analysis); Class Action Complaint ¶¶ 48, 58-59, *Phipps v. Wal-Mart Stores, Inc.*, 925 F. Supp. 2d 875 (M.D. Tenn. 2012) (No. 3:12-01009), *on appeal on other grounds*, No. 13-6194 (6th Cir. 2013); Plaintiffs' Opposition to Wal-Mart's Motion to Dismiss Fourth Amended Complaint at 21, *Phipps v. Wal-Mart Stores, Inc.*, 925 F. Supp. 2d 875 (M.D. Tenn. 2012) (No. 3:12-01009), *on appeal on other grounds*, No. 13-6194 (6th Cir. 2013) (describing plaintiffs' "refined" statistical analysis); Plaintiffs' Original Complaint and Jury Demand ¶¶ 39-41, *Odle v. Wal-Mart Stores, Inc.*, No. 3:11-CV-02954-O, 2012 WL 5292957 (N.D. Tex. Oct. 15, 2012); Class Action Complaint ¶¶ 56, 66-68, *Love v. Wal-Mart Stores, Inc.*, No. 0:12-cv-61959-RNS, 2013 WL 5434565 (S.D. Fla. Sept. 23, 2013); Class Action Complaint ¶¶ 61-63, *Ladik v. Wal-Mart Stores, Inc.*, 291 F.R.D. 263 (W.D. Wis. 2013) (No. 3:13-cv-00123-bbc).

significance more likely.<sup>232</sup> The defendant's expert typically responds by arguing that the plaintiffs' selected reference population is too broad. The defense expert then runs traditional statistical tests on a smaller population, often resulting in finding that the observed disparities are not statistically significant. This scenario played out in *Wal-Mart*, where the plaintiffs' expert argued that the relevant population for study was either the nationwide population of Wal-Mart employees or region-by-region populations (making the observed disparities statistically significant),<sup>233</sup> while the defense argued that separate statistical analyses should be performed for each store or even for each department within stores (which led to findings that the observed disparities were not statistically significant).<sup>234</sup> The parties to systemic discrimination cases are by now familiar with this dance, and its predictability may be valued by both

---

232. See PAETZOLD & WILLBORN, *supra* note 49, § 4.29-4.36; William T. Bielby & Pamela Coukos, "Statistical Dueling" with Unconventional Weapons: What Courts Should Know About Experts in Employment Discrimination Class Actions, 56 EMORY L.J. 1563, 1597-98 (2007) ("[M]ost importantly, power increases with sample size. A gender gap of a specific magnitude (say 15%) is more likely to show up as statistically significant in a regression analysis based on 1,000 observations than in one based on 100 observations."); Tristin K. Green, *Discrimination in Workplace Dynamics: Toward a Structural Account of Disparate Treatment Theory*, 38 HARV. C.R.-C.L. L. REV. 91, 120-21 (2003) ("Statistical significance becomes harder to attain as the sample size shrinks."); Paetzold, *supra* note 43, at 402 ("Third, p-values are highly influenced by sample size (here, the actual number of hiring decisions included), so that simply by increasing the sample size and holding other things constant, a statistician can eventually achieve statistical significance."); Ramona L. Paetzold & Rafael Gely, *Through the Looking Glass: Can Title VII Help Women and Minorities Shatter the Glass Ceiling?*, 31 HOUS. L. REV. 1517, 1527 n.50, 1540 (1995); see also Weiss, *supra* note 44, at 1710-13 (considering the reference class problem).

233. See *Wal-Mart Stores, Inc. v. Dukes*, 131 S. Ct. 2541, 2555 (2011) (discussing statistical evidence of regression analyses performed at national and regional levels).

234. See *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137, 156-57 (N.D. Cal. 2004), *aff'd in part, remanded in part by* 603 F.3d 571 (9th Cir. 2010), *rev'd*, 131 S. Ct. 2541 ("Defendant argues that because Dr. Drogin's regression analysis was conducted at the regional, rather than the store (or store department) level, his results are too generalized and fail to account for the significant differences in compensation practices that exist among the individual stores. In statistical language, Dr. Drogin's regional analyses suffer from 'aggregation bias,' and the only way to cure this bias, according to Defendant, would be to 'disaggregate' the data and test it at the store-by-store level.").

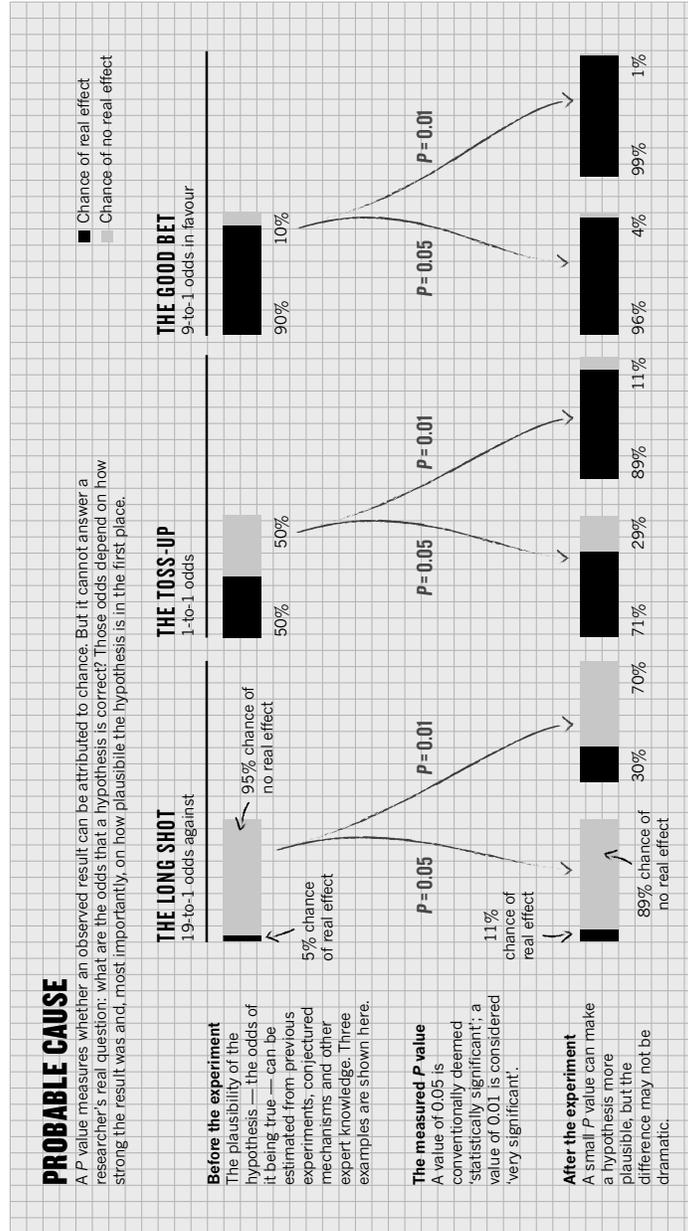
sides. Nevertheless, a creative attorney—whether representing employees or employers—may soon decide to venture into the untested waters of Bayesian analysis in discrimination cases. Until that happens, scholars can only continue to make the case that stubborn adherence to  $p$ -values, without acknowledgment of prior probabilities, is misguided and that a Bayesian revolution is both preferable and manageable.

### CONCLUSION

Systemic disparate treatment law has been infected with a misunderstanding and misuse of  $p$ -values since its inception, and the Court's *Wal-Mart* decision reflects the consequences of that infection. The *Nature* article highlights the pervasiveness of the misunderstanding of  $p$ -values, even in the scientific community, and it serves as a valuable reminder of the limitations of statistical significance testing. The misunderstanding of  $p$ -values is magnified in courtrooms.  $P$ -values were simply never intended to perform the tasks that they have been assigned by the Supreme Court in systemic disparate treatment law. Ronald Fisher would never have intended for a  $p$ -value of .05 in an employment discrimination case to satisfy the burden of proving that it is more likely than not that the defendant employer engaged in a pattern or practice of unlawful discrimination.

The time has come for a Bayesian revolution in systemic disparate treatment law. This Article has demonstrated that the Bayesian revolution is indeed possible, and that the hard questions raised by acknowledging the influence of prior probabilities can be functionally addressed through our civil justice system, our current rules of civil procedure, and a careful focus on institutional competence. Although imperfect estimations of prior probabilities based on flawed data or educated guesses about the prevalence of discrimination may be unsatisfying, they are preferable to the status quo and preferable to ignoring statistical evidence of employment outcome disparities altogether. This Article represents only an initial attempt to systematically address the hard questions that will be presented by switching to Bayesian analysis of employment statistics, but it is my fervent hope that it will not be the last such attempt.

APPENDIX\*



\* Copyright Regina Nuzzo (2014). Reprinted with permission from: Regina Nuzzo, *Scientific Method: Statistical Errors*, 506 NATURE 151 (Feb. 12, 2014), available at <http://www.nature.com/news/scientific-method-statistical-errors-1.14700> (citing Thomas Sellke et al., *Calibration of p Values for Testing Precise Null Hypotheses*, 55 AM. STATISTICIAN 62 (2001)).

