For the rational study of the law the black-letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.¹

Title VII of the Civil Rights Act of 1964² makes it unlawful for an employer to discriminate on the basis of race, color, religion, sex, or national origin.³ The interpretation of the statute by the Supreme Court has led to the development of two distinct types of hiring discrimination claims: disparate treatment⁴ and disparate impact. The disparate impact claim evolved out of the need to eradicate the less obvious forms of discrimination, perpetrated by "employment practices that are facially neutral in their treatment of different groups but that in fact fall more harshly on one group than another and cannot be justified by business necessity."⁵ In

---

¹ O.W. Holmes, The Path of the Law, in COLLECTED LEGAL PAPERS 187 (1920).
³ See id. The statute provides in part:

It shall be an unlawful employment practice for an employer—
(1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin; or
(2) to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin.

Id. § 2000e-2(a).
⁴ See infra note 16.
disparate impact cases, the plaintiff must focus her efforts on showing the discriminatory impact of a hiring procedure. Once the plaintiff has made a prima facie case,⁶ the burden of proof shifts to the defendant to show that the hiring procedure is justified by business necessity.⁷ Through the years, statistical analysis has been the method of choice to show disparate impact, mainly because it is the only proof available in many cases of employment discrimination, but also because it makes "[i]nequality between blacks and whites, men and women, young and old ... stand out like figures carved in a mountainside."⁸

Yet the increased use of statistical methods in the disparate impact courtroom has forced the Title VII judge to become an expert in statistical analysis. Judges have struggled to adapt to this quantitative framework, but the development of disparate impact doctrine indicates that their attempts have not been very successful. They have misunderstood basic statistical principles and misinterpreted the relationship between statistical conclusions and legal constructs such as the prima facie case and the burden of proof.

The cost of this confusion has been borne by the disparate impact plaintiff, who has had to present statistical analyses that satisfy a strict scientific standard that is inconsistent with the traditional legal standards of proof. The existence of this scientific standard is evident in the long line of cases originating with Griggs v. Duke Power Co.,⁹ the Supreme Court case that defined the disparate impact framework. The recent Supreme Court case of Wards Cove Packing Co. v. Antonio¹⁰ has added fuel to the fire by requiring that the plaintiff present an air-tight statistical case, thus essentially insulating the defendant employer from disparate impact claims.

This Comment explores the traditional and recent statistical fallacies to which judges have adhered in their interpretations of the statistical requirements necessary to satisfy the prima facie stage of the disparate impact case. It also examines the possibility of developing an improved disparate impact framework through

---

⁶ See infra notes 22-51 and accompanying text (discussing what constitutes a disparate impact prima facie case).
⁷ See infra note 69 and accompanying text (discussing the business necessity justification).
Bayesian probability theory, a mathematical model that compels a judge to give the appropriate weight to statistical evidence introduced by the disparate impact plaintiff. Part I examines the statistical misunderstandings that arose under the traditional disparate impact framework outlined in Griggs and its progeny. Part II shows how the recent Supreme Court case, Wards Cove, has held plaintiffs to a near-scientific standard of proof by devising strict causation requirements for the plaintiff. Part III assesses the potential for creating a more coherent disparate impact framework, and concludes that use of the Bayesian probability model to analyze disparate impact claims will result in both a more equitable trial for the plaintiff and a more approachable framework for the judge.

I. THE TRADITIONAL DISPARATE IMPACT FRAMEWORK

The enactment of Title VII over twenty-five years ago marked the first statutory attempt to eradicate employment discrimination. The statute explicitly outlawed intentional discrimination, and the Supreme Court in Griggs interpreted Title VII to extend to cases in which the employer enforced facially neutral job requirements that discriminated against protected classes. Subsequently, the Griggs disparate impact framework has been modified by Wards Cove, which stands as current law.\(^1\) The traditional framework is reviewed first, however, in order to show how courts misinterpreted statistical requirements at the prima facie stage. Next, the impact of Wards Cove on the prima facie case will be analyzed to show how the Supreme Court built on these misinterpretations to create an even less coherent disparate impact theory.

Griggs represents the cornerstone of the disparate impact doctrine, which "prohibits the use of selection criteria and procedures which substantially burden minority group members, unless they can be justified in a manner that would not be required in the absence of unequal results."\(^2\) The Griggs majority recognized that to provide for the equal opportunity mandate of Title VII, it is necessary to outlaw neutral employment standards and procedures that "operate to 'freeze' the status quo of prior discriminatory employment practices."\(^3\) Griggs sent to employers the message

\(\text{1}\) See infra notes 108-12 and accompanying text (discussing the possible legislative overruling of Wards Cove).
\(\text{2}\) D. BALDUS & J. COLE, STATISTICAL PROOF OF DISCRIMINATION 44 (1980).
\(\text{3}\) Griggs, 401 U.S. at 430.
that all hiring practices with harmful effects on minorities must be justified by business necessity, regardless of the good faith with which they are instituted.

This new class of disparate impact employment discrimination cases arose from the Court's statutory construction of Title VII, and stood in contrast to the traditional disparate treatment cases that had been brought under Title VII. Those cases require proof of the employer's discriminatory intent and are therefore usually successful in instances of identifiable prejudice. The disparate impact case provides plaintiffs with the opportunity to present a claim of discrimination when bias is latent or non-existent. The plaintiff's case is met by a showing that she and

14 See infra note 69.

15 See Griggs, 401 U.S. at 434-36. The disparate impact theory is applied today in other areas. See, e.g., Voting Rights Amendments of 1982, 42 U.S.C. §§ 1973, 1973b (1982) (stating that no voting qualifications, tests, or procedures may be imposed or applied by any State or political subdivision in a manner which results in a denial or abridgement of the right of any citizen of the United States to vote on account of race or color); Note, Business Necessity in Title VII: Importing an Employment Discrimination Doctrine into the Fair Housing Act, 54 FORDHAM L. REV. 563, 564 (1986) (analyzing the use of disparate impact analysis to prove discrimination in housing); Note, Disparate Impact Analysis and the Age Discrimination in Employment Act, 68 MINN. L. REV. 1038, 1040 (1984) (examining the use of disparate impact analysis in age discrimination cases arising under the Age Discrimination in Employment Act); see also Wards Cove, 109 S. Ct. at 2129 n.9 (Stevens, J., dissenting).

16 The disparate treatment model is applied when the plaintiff alleges that the employer perpetrated intentional, purposeful discrimination in applying a hiring procedure. See, e.g., International Bhd. of Teamsters v. United States, 431 U.S. 324, 355 n.15 (1977) (defining disparate treatment actions as cases in which "[t]he employer simply treats some people less favorably than others because of their race, color, religion, sex, or national origin"). Proof of discrimination in a disparate treatment case focuses on direct evidence of the employer's discriminatory motive and intent. Direct testimony is pivotal in establishing intent, although "[p]roof of discriminatory motive ... can in some situations be inferred from the mere fact of differences in treatment." Id. The disparate treatment case is therefore most often used when the plaintiff may prove the employer's discriminatory intent to some degree. See also D. BALDUS & J. COLE, supra note 12, at 44-46. Baldus and Cole state that disparate impact differs from disparate treatment in three major ways. First, the defendant's state of mind is irrelevant in a disparate impact case, while discriminatory purpose is required under disparate treatment. Second, disparate impact doctrine allows the defendant to rebut the prima facie case with a showing of business justification, while the defendant's justification is irrelevant in disparate treatment cases. Third, while disparate impact can be used to challenge facially neutral selection procedures, disparate treatment can only apply to situations in which the defendant has sufficient discretion to allow an illegitimate factor to influence her decisions. See id.

17 See, e.g., D. BALDUS & J. COLE, supra note 12, at 26 ("The [disparate treatment] claim alleges intentional, purposeful discrimination in either a rule-making or a rule-applying context.").

18 Cf. id. at 46 (stating that no clear standard has yet been developed by the courts
members of her class are in effect being “closed out” of job opportunities by certain identifiable neutral employment practices.\textsuperscript{19}

A. Statistical Analysis and the Prima Facie Case

The prerequisites for bringing a disparate impact case under \textit{Griggs} and its progeny are twofold. First, the plaintiff must satisfy the requirements of the prima facie case, by showing that the employer’s hiring practice has a disparate impact on protected minorities. Once the plaintiff satisfies this burden, the employer is given the opportunity to rebut the prima facie case by showing that the hiring practice was justified by business necessity.\textsuperscript{20} The defendant employer’s failure to prove business necessity results in a judgment for the plaintiff. Finally, if the employer establishes that the rule or procedure is necessary, it will violate Title VII if there exists a selection that achieves the same result but has a lesser disparate impact on the plaintiff’s class.\textsuperscript{21} The prima facie case represents the plaintiff’s “first hurdle.” She must present enough evidence to establish an inference of discrimination. The defendant’s opportunity to show business necessity represents the plaintiff’s “second hurdle.” The plaintiff’s case must be strong enough to withstand any allegations by the defendant employer that the practice was a business necessity. The plaintiff must surmount both these hurdles to prevail.

In a disparate impact case, a plaintiff satisfies the prima facie rule\textsuperscript{22} when she introduces sufficient evidence to establish a

\textsuperscript{19} See, e.g., \textit{Griggs}, 401 U.S. at 430 n.6 (finding that although census data indicated that 34\% of North Carolina’s white males and 12\% of its black males had completed high school, EEOC findings showed that the employer’s transfer qualification tests resulted in 58\% of whites passing and only 6\% of blacks).

\textsuperscript{20} See \textit{infra} note 69 and accompanying text.

\textsuperscript{21} See \textit{Albemarle Paper Co. v. Moody}, 422 U.S. 405, 425 (1975).

\textsuperscript{22} There are two interpretations of the requirements of a prima facie case. One views the rule as being satisfied when the plaintiff has submitted sufficient evidence to send the case to the jury, while the other considers it satisfied when the plaintiff has not only removed by sufficient evidence the duty of producing evidence to get \ldots to the jury, but has gone further, and, either by means of a presumption or by a general mass of strong evidence, has entitled himself to a ruling that the opponent should fail if he does nothing more in the way of producing evidence.

\textsuperscript{9} \textit{J. Wigmore, EVIDENCE} § 2494, at 379 (1981). The latter rule is applied in disparate impact cases, such that where the plaintiff establishes an inference of discrimination
significant statistical under-representation or harmful effect on her class. If the statistical evidence satisfies the court's disparate impact guidelines, an inference of disparate treatment will arise. Because disparate impact cases center on the harmful effects of certain quantifiable characteristics of the employer's selection process, proof of disparate impact will almost always be in quantitative form. The reliance on quantitative evidence is so great that "[u]nder a disparate impact theory, statistical evidence of impact is not merely circumstantial (as it is under disparate treatment theory), but is direct evidence of the results which trigger the demand for additional justification." Because plaintiffs develop statistical analyses to gauge the effect of an employer's hiring procedures on members of the plaintiff class, satisfying the prima facie rule will often hinge on the accuracy and credibility of the plaintiff's statistical analysis. For that reason, parties in disparate impact cases spend a considerable amount of time preparing statistical studies that will buttress their arguments in court.

A wide variety of statistical methods are available to assist the plaintiff in preparing the prima facie case. Indeed, the issue of in the prima facie case, this operates as a presumption in her favor. If the defendant fails to satisfy her burden of rebuttal, the plaintiff will prevail. See Texas Dep't of Community Affairs v. Burdine, 450 U.S. 248, 254 (1981) ("Establishment of the prima facie case in effect creates a presumption that the employer unlawfully discriminated against the employee."); Furnco Constr. Corp. v. Waters, 438 U.S. 567, 577 (1978) ("A prima facie case... raises an inference of discrimination only because we presume these acts, if otherwise unexplained, are more likely than not based on the consideration of impermissible factors."); see also Smalls, The Burden of Proof in Title VII Cases, 25 How. L.J. 247, 252-53 nn.31-34 (1982) (citing cases and relevant dicta).


For example, Baldus and Cole state that the plaintiff's evidence may provide evidence of how the system actually operates (e.g., comparative black and white pass rates for actual applicants) or... evidence of how it would operate if applied to a population of potential applicants (e.g., since fewer blacks than whites have a high school diploma, the eligibility requirement of a high school diploma has an adverse impact on blacks).

Id. at 47.

For a primer on statistical methods, see generally W. CURTIS, STATISTICAL
selecting and correctly applying the proper statistical method in a disparate impact case has received a considerable amount of attention from scholars.\textsuperscript{28} The variety of statistical methods permits the plaintiff to tailor her statistical study to her legal argument.\textsuperscript{29} The plaintiff's choice of a statistical method is thus a crucial one since the inference of discrimination arises only if the plaintiff has constructed a coherent analysis.

A preferred method of statistical analysis is hypothesis testing, a statistical procedure used to settle "disputed questions on which statistical evidence is brought to bear."\textsuperscript{30} The hypothesis being tested is termed the null hypothesis, or "the characteristic \ldots postulated or hypothesized to exist in a universe \ldots."\textsuperscript{31} In the context of employment discrimination litigation, the null hypothesis is that "race or gender is not involved in the employer's employment decisions."\textsuperscript{32} The statistical test of this hypothesis will eventually lead to its acceptance or rejection.

The test is conducted by compiling a sample or set of observations—that consist namely of employment decisions—and using that
data to confirm or reject the null hypothesis. The probative value of the hypothesis test will increase with sample size. From a disparate impact standpoint, this means that the smaller the number of observations, the more likely that a conclusion of discrimination is tenuous. For the test of statistical inference to be valid, the sample also must be a random selection from a relevant class of data. Although data selection is, in a sense, a preliminary step to conducting the statistical analysis, it is a crucial one. Indeed, the validity of the analysis will hinge on the requirements of randomness and relevance of the data that was selected. If the sample size is small in relation to the total number of observations, due to limited data, for instance, the defendant may challenge the plaintiff’s data selection as flawed. She will argue that the data does not represent a valid cross-section of employment decisions in the workplace, so that any statistical conclusions drawn from the data are biased. The plaintiff must, therefore, be ready to defend her methods of data selection in court.

Once the null hypothesis and the sample have been selected, a level of statistical significance must be chosen. Statistical significance measures the probability of rejecting the null hypothesis when it is in fact true. The choice of a significance level is a matter of judgment and depends on the ultimate purpose of the statistical test. If, for example, the disparate impact plaintiff selects a significance level of 5%, the null hypothesis of no discrimination will not be rejected unless the probability that the results occurred by chance is less than 5%.

The complement of the significance level is the “confidence coefficient,” equal to one minus the significance level. The confidence coefficient in the above example would, for instance, be

---

35 See D. BALDUS & J. COLE, supra note 12, at 32, 134-36; M. FINKELSTEIN, QUANTITATIVE METHODS IN LAW 5 (1978). From a practical standpoint, if the defendant in a disparate impact case has a small number of employees, it will be more likely that the plaintiff’s analysis encompasses the entire sample of relevant observations. If, however, the employer is a corporation with thousands of employees, the sample size will probably be smaller than the total number of relevant observations, since the plaintiff will often be unable to compile the multitude of relevant employment decisions.

34 See D. BALDUS & J. COLE, supra note 12, at 313.

36 See W. CURTIS, supra note 27, at 119-20.

37 See Kaye, supra note 30, at 1347.
95%. It must be stressed that the confidence coefficient does not necessarily represent the probability that the findings of the analysis are correct. The reason for this is that the statistical test generating the confidence coefficient does not incorporate all the probabilities that are required to make this a valid statement. Although the statistical reasoning behind this assertion is not entirely intuitive, it may be summarized as follows: the hypothesis test that generates the confidence coefficient calculates the probability of the observed evidence given the null hypothesis of no discrimination. In probabilistic terms, however, the burden of persuasion relates to the probability that the disparity is associated with discrimination given the statistical and non-statistical evidence. These two probabilities are not the same, so that probabilities generated through hypothesis testing may not be viewed as indicators of certainty, and thus do not gauge the degree to which legal standards of proof have been satisfied. For the hypothesis test to model accurately the burden of persuasion, it must also look to "the probability of the evidence given the alternative hypothesis and the probability of the hypothesis itself."

Thus, this probability must exceed the value attributed to the burden of persuasion for the burden to be satisfied. For a discussion of the probabilistic interpretation of the burden of persuasion, see infra notes 156-62 and accompanying text.

Kaye, Statistical Significance and the Burden of Persuasion, 46 LAW & CONTEMPPROBS. 13, 22-23 (1983). Kaye has constructed a good example that illustrates the above reasoning. Assume that a gambler represents that certain dice have a 50% chance of showing an even or odd number, although in fact, the gambler also carries a second pair of loaded dice that have an 80% probability of showing an odd number. Assume that the gambler entices an unsuspecting stranger into playing a game in which they alternate rolling the dice and predicting whether the outcome will be even or odd. They play ten times, and the dice come out odd eight of those times. The innocent bystander loses a large sum of money, and later discovering that the gambler always carries two sets of dice with her, brings suit against the gambler. How does the jury decide whether the dice were loaded? If a hypothesis test is used, the null hypothesis is that the defendant used the fair dice, while the alternative hypothesis is that she used the loaded dice. The test generates the probability of the evidence (eight odd outcomes, two even) given the null hypothesis that the fair dice were used, and determines that it is 5.5%. As 5.5% is deemed the level of significance, the confidence coefficient is equal to 94.5%.

Kaye argues that the analysis does not indicate that there is a 94.5% chance that the loaded dice were used. First, the test ignores the probability of coming up with eight odd outcomes if the alternative dice are used. Second, the probabilities generated by the test do not change if instead of being a "crooked gambler," the defendant is an "honest law professor" who, not knowing that one of the two sets of dice were loaded, selected one of the sets at random. This second point may be expressed in probabilistic terms. The hypothesis test assumes that the null
The third statistical concept crucial to any disparate impact analysis is that of statistical power. Statistical power is the probability that the statistical test will correctly reject the null hypothesis.\textsuperscript{40} "It is the complement of the probability that the test will accept the null hypothesis when that hypothesis is false."\textsuperscript{41} If the power of the test is high—and, thus, its complement low—the test will be considerably more accurate. If, however, the power of the test is low, its complement will be high, so that the test will be less reliable.

The concepts of statistical significance and power may also be defined in terms of what is known as Type I and Type II errors. Type II error is the risk that the hypothesis test fails to reject the null hypothesis when it is false.\textsuperscript{42} In the employment discrimination context, this corresponds to false exculpation, that is concluding falsely that there is no discrimination where discrimination exists. Type I error represents the risk of rejecting the null hypothesis when it is true. This corresponds to false inculpation, namely finding incorrectly that there has been discrimination where no discrimination has occurred. Type I risk is another way of expressing the concept of statistical significance, so that the probability of falsely rejecting the null hypothesis increases with the level of significance.\textsuperscript{43} Power, on the other hand, "is the probability of \textit{not} making a Type II error," as it "is the probability of correctly rejecting the null hypothesis."\textsuperscript{44} Finally, although Type I and Type II risks are not complementary in the sense that they do not add up to one, they are inversely related—increasing one decreases the other.\textsuperscript{45}

---

hypothesis—that the fair dice were used—and the alternative hypothesis—that the loaded dice were used—are equally likely at the outset. This assumption, however, is accurate only if, as in the case of the honest law professor, the dice are selected randomly. Thus, the results of the hypothesis test will be inaccurate if the defendant is a confirmed crooked gambler, as the prior probability that she will use the fair dice will be less than 50%. One problem with hypothesis testing, then, is that it ignores prior probabilities. \textit{See id.} at 17-23. Kaye’s analysis underscores one of the difficulties of using hypothesis testing in the litigation process.

\textsuperscript{40} \textit{See Kaye, supra} note 35, at 844 n.41.

\textsuperscript{41} \textit{Id.}

\textsuperscript{42} \textit{See id.}


\textsuperscript{44} \textit{See} Goldstein, \textit{supra} note 43, at 34.

The way in which Type I and Type II errors operate in the courtroom may be described in terms of whether a defendant is found guilty or innocent. If the defendant is found guilty of discrimination and the null hypothesis is thus rejected, a Type II error is impossible, since the probability of falsely failing to reject the null hypothesis is 0.\textsuperscript{46} If, on the other hand, the defendant is found not guilty and the hypothesis is accepted, then the probability of falsely rejecting the null hypothesis is equal to 0, so that Type I error is impossible. Typically, the defendant will want to point out the degree of Type I error made by the plaintiff, while the plaintiff will focus on the degree of Type II error made by the defendant.\textsuperscript{47}

The next step in hypothesis testing involves selection of a test statistic with which to test the hypothesis. The plaintiff chooses a test statistic from a number of available test statistics\textsuperscript{48} depending on the probabilistic process that generated the data. The test statistic is computed and compared to a chart listing the significance levels that correspond to various values of the test statistic. From this chart, the P value is determined, which indicates the "level of statistical significance to be ascribed to a statistical result."\textsuperscript{49} Typically, the P value decreases as the population analyzed increases and as the population/workforce disparity increases.\textsuperscript{50}

Finally, by comparing the P value to the cut-off level selected earlier in the analysis, the decision to accept or reject the null hypothesis is made. If the significance level generated by the analysis is lower than that required to reject the null hypothesis, the null hypothesis is rejected and, in a disparate impact case, the analysis will lead to a finding of discrimination. Otherwise, the null hypothesis of no discrimination will be accepted. Ultimately, the

\textsuperscript{46} Remember that the null hypothesis is that the employer is not guilty of discrimination, see supra text accompanying note 32, so that in this case the null hypothesis is rejected, see Goldstein, supra note 43, at 33-34.

\textsuperscript{47} See Goldstein, supra note 43, at 34-35.

\textsuperscript{48} There are four main categories of test statistics: the Z statistic, chi-square statistic, t statistic, and F statistic. See D. BALDUS & J. COLE, supra note 12, at 360.

\textsuperscript{49} Id. at 356.

\textsuperscript{50} The magnitude of the disparity measured in percentage differences may be distinguished from the probability that the disparity could have occurred by chance. These concepts are nonetheless related in the sense that "the larger the difference, the lower the probability that [the difference] occurred by chance. . . .

\textsuperscript{Castaneda v. Partida, 430 U.S. 482, 494 n.13 (1977) ("If a disparity is sufficiently large, then it is unlikely that it is due solely to chance or accident, and, in the absence of evidence to the contrary, one must conclude that racial . . . factors entered into the selection process"); Boardman & Vining, The Role of Probative Statistics in Employment Discrimination Cases, 46 LAW & CONTEMP. PROBS. 189, 204-05 (1983).
lower the significance level and thus the greater the confidence coefficient of the analysis, the more credence a court will lend to the hypothesis test.\textsuperscript{51}

### B. Legal Standards and Statistical Proof

Two major statistical misunderstandings arise when statistical findings are introduced into the employment discrimination courtroom. The main cause for the problem is that the judge must decide whether the analysis satisfies the traditional common law standards that govern civil cases. Thus, the ultimate question for the court is whether the plaintiff has met the requirements of the prima facie case, namely whether she has sustained her burdens of production and persuasion which constitute the burden of proof.\textsuperscript{52} The statistical analysis will thus be examined to determine whether that burden has been satisfied.\textsuperscript{53}

The court's decision that a disparate impact plaintiff has made a prima facie showing\textsuperscript{54} generally rests on the conclusion that the

\textsuperscript{51} That is not to say that the greater the confidence coefficient, the more the court may be "confident" that the results of the hypothesis test are correct. \textit{See supra} notes 37-39 and accompanying text.

\textsuperscript{52} \textit{See} C. McCormick, \textit{The Law of Evidence} § 336 (3d ed. 1984). The burden of production is satisfied if the evidence is such that a reasonable man could infer the existence of the fact to be proved. A "scintilla" of evidence will not suffice. \textit{See id.} § 338. The burden of persuasion is satisfied in civil cases when the jury finds "proof by a preponderance" that the existence of the contested fact is more probable than its nonexistence. \textit{See id.} § 339.

\textsuperscript{53} At the prima facie stage, the employer may and often will present statistical studies that question the conclusions of the plaintiff's statistical analysis. The court's decision as to whether the plaintiff has satisfied her burden of proof may therefore be made on the basis of both the plaintiff's and the employer's analysis. EEOC guidelines, however, impose a heavy burden on employers' statistical analyses, requiring that there be no more than a one-in-twenty possibility that the correlation between the test results and job performance could have been by chance. This indicates that the confidence coefficient generated in the employer's statistical analysis must be at least equal to 95\%. \textit{See} Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.14(B)(5) (1990).

\textsuperscript{54} In practice, the plaintiff's prima facie showing implies that her statistical analysis shows a sufficient degree of population/workforce disparity to infer discrimination, that the results are statistically significant, and that the analysis presents no flaws such as inaccurate or insufficient data, poor choice of relevant population, or inappropriate selection or use of a statistical method. \textit{See}, e.g., Kaye, \textit{Statistical Evidence of Discrimination}, 77 J. AM. STATISTICAL A. 773, 776-82 (1982) (examining the role of statistics in the prima facie case). Throughout this discussion it is assumed that the plaintiff's analysis satisfies these requirements, so that the focus is on the courts' interpretation of the requirement of statistical significance through the burden of proof. Statistical significance is not always dispositive in employment
statistical significance of the plaintiff's study matched or exceeded the corresponding statistical significance that meets the burden of persuasion. Thus, in disparate impact cases, the traditional claim that the plaintiff has proved discrimination "by a preponderance of the evidence" has become a term to express the conclusion that the plaintiff's statistical analysis satisfies the court's statistical significance cut-off rules. In fact, courts have applied a 95% confidence coefficient corresponding to a 5% significance level cut-off in disparate impact cases.

The first problem with this interpretation is that courts have generally assumed that the confidence coefficient may be equated with the burden of proof, so that a hypothesis test with a low level of significance—and thus a high confidence coefficient—is one in which the plaintiff has proved her case. The level of statistical significance, however, merely represents the probability of rejecting the null hypothesis when it is true, and the confidence coefficient, its complement, does not necessarily represent the confidence one has in the accuracy of the hypothesis test. Focusing on statistical significance and the confidence coefficient, therefore, distorts the standards of proof that the plaintiff must meet. The court must focus on power as well, as it represents the probability of correctly rejecting the null hypothesis. In fact, this distortion is an

discrimination cases. See, e.g., Boardman & Vining, supra note 50, at 205 (concluding that the 95% degree of statistical significance of evidence admitted in Albemarle became one of several criteria for establishing the prima facie case). It is, however, an issue to which an increasing number of courts have paid attention. See infra notes 56-64 and accompanying text.

55 C. MCCORMICK, supra note 52.

56 See Palmer v. Schultz, 815 F.2d 84, 96 (D.C. Cir. 1987) ("Statistical evidence must meet the 5% level . . . for it alone to establish a prima facie case under Title VII."); see also Barnett, An Underestimated Threat to Multiple Regression Analyses Used in Job Discrimination Cases, 5 INDUS. REL. LJ. 156, 168 (1982) ("The most common rule is that evidence is compelling if and only if the probability the pattern obtained would have arisen by chance alone does not exceed five percent."); Kaye, supra note 39, at 20 ("[T]here is an understandable tendency on the part of litigants and judges to demand that quantitative scientific evidence meet some unambiguous test for acceptance in the scientific community.").

57 See supra note 37 and accompanying text.

58 The confusion between confidence coefficients and accuracy has pervaded the legal field, in both court opinions and scholarly work. See, e.g., Smith & Abram, supra note 28, at 41 n.31 (stating that the significance level is a benchmark "level of confidence that must be met to reject the null hypothesis" of no discrimination"); see also Moultrie v. Martin, 690 F.2d 1078, 1083 n.7 (4th Cir 1982) (confusing confidence coefficients with degrees of certainty); National Lime Ass'n v. EPA, 627 F.2d 416, 453 n.139 (D.C. Cir. 1980) (misstating the 95% confidence coefficient used in scientific analyses as a standard of certainty); Kaye, supra note 39, at 21 n.41.
argument for discarding hypothesis testing altogether and using a
more intuitively understandable probabilistic model.  

The second problem relates to the courts' statistical significance
cut-off to satisfy the burden of proof. The 95% confidence
coefficient cut-off that courts have adopted is a stringent criterion
developed to test scientific conclusions reached on the basis of
imperfect or incomplete data. It has been adopted for use in
empirical research by the scientific community to test formal
experiments and establish or refute scientific propositions. The
effect of setting the statistical significance cut-off, and thus Type I
error, at 5% is that courts will accept a high risk of Type II error.
Indeed, since Type I and II errors are inversely correlated, seeking a smaller Type I error will require accepting a larger Type II error. The effect of importing this purely scientific standard into
a legal context is mostly felt by disparate impact plaintiffs; they are
required to assemble an analysis in which the risk of falsely
inculpating a defendant, that is, Type I risk, is no more than 5%. At this juncture then, review of such statistical misconstructions
points to the conclusion that an improved quantification of the
burden of proof in civil trials should "express . . . the legal value
that, if an error is made, it is no worse if a defendant is wrongly
found liable than if it is mistakenly exonerated." This weighty burden has been criticized extensively by scholars. In some instances, courts have applied a less taxing stan-

---

59 See infra notes 153-66 and accompanying text (discussing probabilistic interpretation of the burden of proof).
60 See Dawson, supra note 35, at 2.
61 See id.
62 See supra text accompanying note 45.
63 See M. ZIMMER, C. SULLIVAN & R. RICHARDS, supra note 28, at 120-21. The point here is not that the 95% standard requires that the plaintiff prove her case to 95% certainty, as the above discussion has indicated that confidence coefficient is not equivalent to the burden of persuasion, see supra notes 57-39 and accompanying text. Rather, the point is that the choice of such a standard "represents a value judgment within the context of [physical and social science] research as to the relative costs of incorrectly proclaiming a result on one hand and incorrectly deeming a result not to have been demonstrated on the other." Cohen, supra note 45, at 412. Indeed, by adopting this standard which minimizes Type I error and thus increases Type II error, courts have implicitly acknowledged that it is more important to protect against Type I "false inculpation" risk than it is to protect against Type II "false exculpation" risk. See id.
64 M. ZIMMER, C. SULLIVAN & R. RICHARDS, supra note 28, at 121 (stating that this standard means that "no important social value difference has been established between finding an innocent defendant liable and finding a guilty defendant not liable").
65 See, e.g., Cohen, supra note 45, at 394 (adopting a 50% standard); Dawson, supra
A COHERENT DISPARATE IMPACT MODEL

... implicitly acknowledging the difference between the concept of legal significance and that of statistical significance. Overall, however, very few courts have adopted less stringent guidelines, thus supporting the argument that courts are ill-equipped to handle statistical evidence. By admitting statistical evidence in the form of hypothesis testing into their courtrooms, judges have boxed themselves into a corner. When they apply complicated and limited statistical results to a scale of legal requirements, they misinterpret the burden of proof and hinder the plaintiff to the benefit of the defendant.

The foregoing analysis indicates that plaintiffs have remained at a disadvantage in employment discrimination trials. Satisfying the courts' traditional statistical interpretation of the prima facie rules is difficult, particularly for a plaintiff with neither the means nor the capacity to assemble and analyze large amounts of complex data necessary for a "scientific" analysis. Furthermore, the employer defendant benefits from greater access to data, employment records, and in the case of large corporations, a substantial corporate treasury. Finally, the problem of statistical misinterpretation exacerbates the structural advantages of the employer in the disparate impact trial process. Indeed, once the plaintiff has fulfilled the requirements of a prima facie case, the employer is

---

note 35, at 2 (arguing against a 5% significance standard in civil litigation); Goldstein, supra note 43, at 47 (arguing that fairness requires that courts take Type II errors into account); Henkel & McKeown, Unlawful Discrimination and Statistical Proof: An Analysis, 22 JURIMETRICS J. 34, 57 (1981) (arguing against the 95% standard); Kaye, supra note 35, at 840-41 (criticizing the Court's handling of statistics in Hazelwood); Kaye, supra note 39, at 20-21 (comparing legal and scientific standards of proof). But see Hallock, The Numbers Game—The Use and Misuse of Statistics in Civil Rights Litigation, 23 VILL. L. REV 5, 13 (1977) (arguing in favor of the 5% standard).

See, e.g., Bazemore v. Friday, 478 U.S. 385, 400 (1986) ("A plaintiff in a Title VII suit need not prove discrimination with scientific certainty; rather, his or her burden is to prove discrimination by a preponderance of the evidence."); Addington v. Texas, 441 U.S. 418, 423 (1979) (stating that the preponderance of the evidence standard requires litigants to share the risk of error roughly equally); Ethyl Corp. v. EPA, 541 F.2d 1, 28 n.58 (D.C. Cir. 1976) (The certainty which a scientist must have to certify evidence "has never characterized the judicial or the administrative process.... [T]he standard of ordinary civil litigation, a preponderance of the evidence, demands only 51% certainty.")., cert. denied, 426 U.S. 941 (1976); see also Kaye, supra note 39, at 21 (arguing that the Ethyl court "was on the right track").

Cf. United States v. Test, 550 F.2d 577, 582-84 (10th Cir. 1976) (explicitly recognizing the distinction between legal and statistical significance). The Test court concluded that although evidence of discrimination in jury selection was statistically significant it was still too low to meet the court's test of legal significance. See id. at 587.

See infra note 122.
offered the opportunity to rebut the inference of discrimination by showing that the disputed practice or procedure is justified by business necessity.69

The courts' misunderstanding of Type I and Type II errors has resulted in overly strict requirements for the plaintiff at the prima facie stage. These statistical requirements do not correspond to the preponderance of the evidence standard imposed in employment discrimination trials, due to the judiciary's obsessive focus on statistical significance as a proxy for the burden of proof.

II. THE WARDS COVE FRAMEWORK

The Supreme Court's traditional disparate impact doctrine suffered a blow last year, resulting in an even greater reliance on complex statistical analysis and stricter statistical standards for aggrieved plaintiffs. In *Wards Cove*, a mixed disparate impact/disparate treatment case, the Court held in a five to four decision that statistical evidence of racial imbalance within a work force did not alone establish a prima facie case of discrimination. The plaintiff was required to identify the particular employment practice that caused the disparity.70 The Court further ruled that on the issue of burden allocation, the plaintiff bears the burden of persuasion throughout the trial with respect to the Title VII claim. The employer, on the other hand, only bears the burden of production with respect to the business necessity defense.71 In so doing, the *Wards Cove* Court reversed the longstanding disparate impact theory of *Griggs* and its progeny that required the defendant to carry the burden of persuasion with respect to her defense.72

69 See *Griggs* v. Duke Power Co., 401 U.S. 424, 431 (1971) ("The touchstone is business necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited."). Although the extent of the employer's burden was not described in *Griggs*, until recently it had been understood that the burden was one of persuasion, and, therefore, a significant one. See, e.g., Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975) (construing *Griggs* as mandating a shift of evidentiary burdens similar to those established in disparate treatment cases). The Court's decision in *Wards Cove*, reviewed in the next part of this Comment, interpreted business necessity as a more lenient standard lying somewhere between a "mere insubstantial justification" and an "essential" or "indispensable" one. See *Wards Cove Packing Co. v. Antonio*, 109 S. Ct. 2115, 2126 (1989).

70 See *Wards Cove*, 109 S. Ct. at 2124.

71 Id. at 2126.

Although pending legislation may overturn the case, Wards Cove provides an excellent example of courts' statistical misunderstandings in employment discrimination cases.

The plaintiffs in Wards Cove were nonwhite workers employed by a number of Alaskan salmon canneries. Unskilled cannery jobs were filled mainly by nonwhites, while higher paying, skilled noncannery jobs were filled predominately by white workers. The plaintiffs, a class of nonwhite cannery workers, brought disparate treatment and disparate impact Title VII actions alleging that a number of the employers' hiring practices "were responsible for the racial stratification of the work force, and had denied them and other nonwhites employment as noncannery workers on the basis of race."

The District Court in Wards Cove entered 172 findings of fact, rejecting all of the plaintiff's disparate treatment claims, including the housing and dining discrimination claims. It also rejected the disparate impact challenges involving subjective employment criteria to fill noncannery positions on the ground that they were not subject to attack under disparate impact theory. Although the employer's objective practices were held to be subject to disparate impact theory, the claims were rejected for failure of proof.

On appeal, the Court of Appeals for the Ninth Circuit initially affirmed the court's decision, then withdrew the decision and

---

73 See infra notes 108-12 and accompanying text.
74 Among the alleged hiring/promotion practices were: "nepotism, a rehire preference, a lack of objective hiring criteria, separate hiring channels, [and] a practice of not promoting from within." Wards Cove, 109 S. Ct. at 2120. The employees also complained about the employer's racially segregated housing and dining facilities. See id.
75 Id.
76 See id. at 2120; see also Antonio v. Wards Cove Packing Co., 54 Empl. Prac. Dec. (CCH) ¶ 34,437 (W.D. Wash. 1983). The district court found with respect to housing that Hardy Parrish, a foreman at one of the canneries, had responded to a written inquiry about employment by writing:

We are not in a position to take many young fellows to our Bristol Bay canneries as they do not have the background for our type of employees.

Our cannery labor is either Eskimo or Filipino and we do not have the facilities to mix others with these groups.

Id. at ¶ 33,836 (Finding No. 148). Although the district court concluded that the plaintiffs' evidence of segregated housing showed by a preponderance of the evidence facts sufficient to establish a prima facie case, it held that the defendants' evidence dispelled the inference of discriminatory animus. See id. at ¶¶ 33,843-44.
77 See Wards Cove, 109 S. Ct. at 2120.
78 See id.
ordered the case reheard by the court en banc to settle an intracircuit conflict over whether disparate impact theory may be applied to subjective employment practices. The Ninth Circuit panel held that plaintiffs may challenge subjective hiring practices using disparate impact theory, and remanded the case to the appeals court. On remand, the court of appeals held that the plaintiffs had made a prima facie case of disparate impact in hiring for both skilled and unskilled noncannery positions. The holding was based solely on the plaintiffs' statistics that showed "racial stratification by job category." The appeals court then remanded the case to the district court with the instructions that the defendant canneries have the burden of showing that the challenged practices were justified by business necessity. The Supreme Court reversed in an opinion by Justice White that modified the requirements of the prima facie and rebuttal stages of disparate impact cases.

A. Prima Facie Case Revisited: The Quest for Accuracy

With respect to the prima facie case, Justice White wrote that "the cannery work force in no way reflected 'the pool of qualified job applicants' or the 'qualified population in the labor force,'" and that "[t]he 'proper comparison [was] between the racial composition of [the at-issue jobs] and the racial composition of the qualified . . . population in the relevant labor market.' Thus, the majority held that the court of appeals' approach was flawed, since it compared the racial composition of two segments of the employer's work force, not the racial composition of qualified

---

79 See Antonio v. Wards Cove Packing Co., 768 F.2d 1120 (9th Cir.), withdrawn, 787 F.2d 462 (9th Cir. 1985).
80 See Antonio v. Wards Cove Packing Co., 810 F.2d 1477, 1486 (9th Cir. 1987) (en banc).
81 See Antonio v. Wards Cove Packing Co., 827 F.2d 439, 444 (9th Cir. 1987).
82 See id. at 445.
84 Id. at 2122. According to the majority opinion, using the cannery workers as the "potential 'labor force' for unskilled noncannery positions" was "[t]oo broad because the vast majority of these cannery workers did not seek jobs in unskilled noncannery positions . . . . [and] too narrow because there are obviously many qualified persons in the labor market for noncannery jobs who are not cannery workers." Id. at 2123.
85 Id. at 2121 (quoting Hazelwood School District v. United States, 433 U.S. 299, 308 (1977)).
86 See id. at 2122 ("Racial imbalance in one segment of an employer's work force
applicants and that of the segment of the employer's business to which they applied. Under this theory, the Court reasoned, the employer "could be haled into court and forced to engage in the expensive and time-consuming task" of defending the employment practices. The employer's only escape hatch would be "to adopt racial quotas, insuring that no portion of his work force deviates in racial composition," an alternative, the Court stated, that was expressly rejected in drafting Title VII.

The Court also maintained a strict causation requirement for plaintiffs at the prima facie stage, ruling that they must demonstrate that "specific elements of the [employers'] hiring process have a significantly disparate impact on nonwhites." Thus, the plaintiffs were required on remand to show specifically "that each challenged practice has a significantly disparate impact on employment opportunities for whites and nonwhites." The Court stated that this specific causation requirement was justified by the need to avoid the adverse impact of "employers being potentially liable for 'the myriad of innocent causes that may lead to statistical imbalances in the composition of their work forces.'" According to the majority, the requirement was minimally onerous to plaintiffs since liberal discovery rules give employees broad access to employers' records.

The Court illustrated the problem inherent in the lower court's approach with a number of examples. The opinion cited the plaintiffs' statistics, indicating that 15% of new office worker hires and 17% of new medical hires at the cannery were nonwhite. According to the appeals court, the majority concluded, plaintiffs could establish a prima facie case of discrimination because 52% of the cannery workers were nonwhite, even though less than 15-17% of the applicants might be nonwhite and nonwhites might make up a lower percentage of the qualified labor market. See id. The flaws of this approach were further illustrated with a reference to the district court's finding that nonwhites were over-represented among cannery workers, because the employer contracted with a predominately nonwhite union. Under the lower court's methodology, the majority explained, the employer would make the prima facie case "disappear" if it ceased to use Local 37 as its source of cannery hires, even though its hiring practices remained unchanged. See id. at 2123.

The Court relied on both the actual statute, 42 U.S.C. § 2000e-2(j) (1988), and Albemarle Paper Co. v. Moody, 422 U.S. 405, 449 (1975) (Blackmun, J., concurring) (stating that a quota system is far from the intent of Title VII).

The Court addressed the issue of availability of employer records and concluded that employers falling within the scope of the Uniform Guidelines on
Although both of these holdings make it more difficult for a plaintiff to satisfy the prima facie rule, they are not completely unfounded. Indeed, the Court's first holding correctly recognizes that it would be unjust to hold the employer liable for having hired a predominately white work force if none of the applicants for positions were nonwhite. Yet the choice of a relevant qualified labor supply is a painstaking process that constitutes a crucial portion of the plaintiff's analysis, since the disparate impact claim will turn on whether qualified nonwhites applied for the position. The plaintiff must select a geographical area that corresponds to the employer's hiring area and isolate actual and potential applicants for the jobs at-issue. The plaintiff must then show that the employer's hiring procedures limited the number of nonwhite hires by comparing the qualified applicant pool, that is the number of actual and potential nonwhite applicants, to the number of nonwhite hires. Demarcating the qualified applicant pool is particularly difficult, as an accurate study must count the number of potential or "discouraged" qualified applicants who would have applied but for their knowledge of the employer's discriminatory practices. This must be achieved by engaging in a complex economic and demographic analysis to determine the geographical scope of the employer's hiring.

The process of determining the qualified applicant pool will result in a set of figures that is nothing more than a "guesstimate" of the number of qualified applicants. Because in most cases both the plaintiff and defendant can adequately justify their choice of the relevant labor market, it is nearly impossible for the court to endorse one analysis over another. Indeed, a considerable

Employee Selection Procedures are required to maintain records on the impact of tests and selection procedures on employment by racial, sexual, or ethnic composition. See id.  
94 Smith and Abram state that:

[i]he labor market area used in deriving the minority . . . percentage, for comparison with the employer's actual work force data, often controls the degree of disparity between minority . . . employment data and the comparison statistic. Egregious displays of gerrymandering characterize litigation and compliance proceedings. Plaintiffs . . . commonly argue that geographic boundaries encompassing the highest concentrations of minorities . . . should apply, while defendants construct geographic areas for comparison purposes which have relatively low concentrations of minorities . . . . No general agreement exists, however, on the standards which should govern the geographic scope of the labor market area.

Smith & Abram, supra note 28, at 59-60 (footnote omitted).
number of courts have struggled with the need to compare the racial composition of the employer's labor force to that of an accurate set of potential applicants. Thus, the Court's quest for scientific accuracy in data selection is misguided, as the analysis will often hinge on too many statistical and economic assumptions to make it accurate. Ultimately, the prevailing party in a disparate impact action will have proven to the Court that her data selection was more reasonable than that of her opponent, not that her statistical results were factually correct.

The majority's pursuit of accuracy is also apparent in its discussion of the prima facie case causation requirements. Stating that it is insufficient to claim "bottom line" discrimination, the Court held that plaintiffs must show how specific elements of the hiring process have a significant disparate impact on nonwhites. The degree of specificity required by the Court is unclear. The Wards Cove plaintiffs had in fact claimed that certain identifiable hiring practices violated Title VII; they did not make an amorphous claim of discrimination.

Certainly, the fact that the plaintiffs were able to isolate identifiable factors that were at the root of the employer's alleged discrimination should have weighed in their favor. It is generally difficult to isolate the causes of racial disparity in an employer's work force. Plaintiffs must first act on their intuition and attempt to identify the major types of hiring procedures that caused the disparity. The next step is confirming their intuition with a coherent statistical study. If the study fails to verify intuition, the plaintiffs must seek out other possible causes. In some instances, the employee may have no intuitive indication of the source of the disparity. This may be the case, for example, if the employer is an institution or a large corporation whose rules and procedures appear fair on their face, but are in fact implemented collectively in a complex regulatory web that results in hiring discrimination.

The majority's imposition of strict causation requirements is a second attempt at inducing the plaintiff to present an air-tight prima facie case. This lofty standard of causation is nearly impossible for plaintiffs to meet. Indeed, civil case law generally has been based on the principle that "[a]lthough the causal link must have

---

95 See Wards Cove, 109 S. Ct. at 2124 (citing Connecticut v. Teal, 457 U.S. 440, 450 (1982)).
96 See id.
97 See supra note 74.
substance, the act need not constitute the sole or primary cause of the harm." This notion is consistent with the preponderance of the evidence standard of proof applied in civil cases. Courts have loosened this causation standard in instances where the burden of proving causation is particularly heavy, such as in large-scale product liability actions. It would therefore appear that the Court has imposed stricter causation requirements for disparate impact plaintiffs than for other types of plaintiffs.

The *Wards Cove* majority is striving for a disparate impact framework in which the plaintiff will present the most accurate statistical statement of disparate impact to prove her prima facie case. This puts an unfair burden on the plaintiff by requiring her to bear the risk of error, and strains the legal system by denying the reality that statistics can be manipulated to state a conclusion based on a set of possibly unwarranted assumptions. Statistical analyses do not result in a conclusion that is "right or wrong," but rather a set of conclusions whose accuracy is based on the validity of the assumption underlying the statistical analysis, the soundness of the analysis, and the "quality" of the set of observations that is run through the statistical model. For that reason, strict statistical guidelines on data and causation will not necessarily generate accurate analyses. They may lead plaintiffs and defendants to engage in a war of statistics which will confuse the trier of fact. Alternatively, they may simply deter plaintiffs with genuine causes of action from bringing suit, for fear that their statistical analysis does not justify a finding of discrimination.

Taken as a whole, *Wards Cove* is quite troubling. Indeed, the *Wards Cove* majority also redefined the burden the employer must meet in the business justification phase of a disparate impact case. The court held that "the employer carries the burden of producing evidence of a business justification for his employment practice." In so doing, the Court relied on Justice O'Connor's plurality opinion in *Watson*, which introduced a more lenient interpretation of the employer's burden in the business justification phase. Thus, the majority explicitly rejected prior case

98 *Wards Cove*, 109 S. Ct. at 2132 (Stevens, J., dissenting) (citing RESTATEMENT (SECOND) OF TORTS §§ 431-433 (1965)).
99 See infra notes 144-52 and accompanying text.
100 See supra notes 62-64 and accompanying text.
101 *Wards Cove*, 109 S. Ct. at 2126 (emphasis added).
102 Justice O'Connor's framework in the *Watson* plurality opinion provided that
law, stating that "to the extent that those cases speak of an employer's 'burden of proof' with respect to a legitimate business justification defense . . . they should have been understood to mean an employer's production—but not persuasion—burden." Thus, Wards Cove now requires the plaintiff to isolate accurately the qualified applicant pool in the geographic and economic perimeter of the employer, identify the specific causes for the disparity, and show that those causes are the source of racial disparity through a statistical study that satisfies strict scientific statistical testing standards. The defendant may then rebut the case by introducing business justification evidence that is more than "insubstantial." It is difficult to see why an individual claiming a Title VII disparate impact harm would ever file suit against the employer when the odds of her prevailing are, in theory, quite slim.

In practice, it is conceivable that Wards Cove has not eradicated disparate impact theory. Civil rights activists and attorneys disagree over the impact of the ruling on future employment discrimination cases. Still, the case has undoubtedly had an immediate impact when a plaintiff has made out a prima facie case of disparate impact, and when the defendant has met its burden of producing evidence that its employment practices are based on legitimate business reasons, the plaintiff must 'show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in efficient and trustworthy workmanship.'

Watson, 108 S. Ct. at 2790 (quoting Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975)).

See Wards Cove, 109 S. Ct. at 2132 (Stevens, J., dissenting) (stating that this new framework "casually—almost summarily" overruled a line of cases that "always have emphasized that in a disparate impact case the employer's burden is weighty").

Id. at 2126 (citation omitted). The Court described this burden as a "reasoned review of the employer's justification for his use of the challenged practice . . . . There is no requirement that the challenged practice be 'essential' or 'indispensable' to the employer's business for it to pass muster. . . ." Id.

See The Supreme Court, 1988 Term—Leading Cases, 103 Harv. L. Rev. 137, 356 n.37 (1989) ("[Wards Cove] 'does not make it impossible to prove [disparate impact] cases, but makes it more difficult. The important point about Griggs was that once you carried the prima facie case, the burden shifted and it was a tremendous litigation advantage.'" (quoting Penda Hair of the NAACP Legal Defense and Educational Fund, in Stewart, Civil Rights: Just a Trim?, A.B.A. J., Aug. 1989, at 44)); Speakers at Dallas Meeting Assess Impact of Supreme Court's 1988-89 EEO Decisions, 199 Daily Lab. Rep. (BNA) A-7 (Oct. 17, 1989) (quoting University of Michigan law professor, Theodore St. Antoine, as stating that Wards Cove will make it more difficult, but not impossible to prove Title VII disparate impact cases); Is the High Court Hiding Reversals on Rights?, Nat'l L.J., June 19, 1989, at 5, 38, col. 1 (quoting Isabelle Katz Pinzler of the ACLU as stating that "the doors of opportunity opened by Griggs have been slammed shut").
on recent disparate impact claims, as a number of courts have recently ruled on Title VII actions based on the *Wards Cove* framework.\(^{107}\)

In any case, the bearing of *Wards Cove* on disparate impact theory may be temporary, given the current legislative push to overturn the case.\(^{108}\) Both the House of Representatives\(^{109}\) and the Senate\(^{110}\) recently passed civil rights legislation that could have overturned the Court’s decision, had it not been successfully vetoed by the President.\(^{111}\) President Bush sent an alternative version of the civil rights bill to Congress in an effort to develop an acceptable version of the bill.\(^{112}\) Although enacting such legislation would rectify some of the inconsistencies of the *Wards Cove* opinion, it most likely would not correct courts’ general misunderstandings of statistical evidence in employment discrimination cases.

---

\(^{107}\) *Wards Cove* has caused a number of courts to deny relief on workplace discrimination claims. *See, e.g.*, Zamlen v. Cleveland, 906 F.2d 209 (6th Cir. 1990) (finding that a firefighter qualifying exam did not discriminate against female applicants); Walls v. Petersburg, 895 F.2d 188 (4th Cir. 1990) (upholding the discharge of a black employee who failed to complete an intrusive job questionnaire). *But see* Nash v. Jacksonville, 905 F.2d 355 (11th Cir. 1990) (holding that a fire department promotion test violated Title VII). A number of circuit courts have remanded cases for consideration under the *Wards Cove* doctrine. *See, e.g.*, EEOC v. Joint Apprenticeship Comm. of the Joint Indus. Bd. of the Elec. Indus., 895 F.2d 86 (2d Cir. 1990); Evans v. City of Evanston, 881 F.2d 382 (7th Cir. 1989); Allen v. Seidman, 881 F.2d 375 (7th Cir. 1989).

\(^{108}\) The Minnesota legislature has expanded its workplace discrimination laws partly in response to *Wards Cove*. The relevant portion of the legislation requires that employers "prove that an employment practice that has a significant adverse impact on a particular protected class is manifestly related to the job or significantly furthers an important business purpose." *Minnesota Law Reverses Impact of 1989 U.S. Supreme Court Civil Rights Rulings*, 95 Daily Lab. Rep. (BNA) A-17 (May 16, 1990); *see also* Human Rights—Discriminatory Practices Amendments, S.F. No. 1847, 1990 Minn. Sess. Law. Serv. 961, 967 (West) (to be codified at MINN. STAT. § 363B.03.11).


\(^{112}\) This compromise bill is, according to President Bush, intended to "strike a blow against racial bias, without institutionalizing quotas." *Bush Vetoes Bill, supra* note 111, at A-A.
B. The Wards Cove Dilemma: Integrating Law and Statistics

It is unwise to attempt to evaluate the logical reasoning of a Supreme Court opinion from a purely statistical standpoint. Indeed, the Court often tiptoes through issues, framing the logic of the opinion with subtleties and nuances that may not be translated statistically. Its opinions must balance considerations of public policy, national values, and stare decisis. In Wards Cove, the Court altered the prior disparate impact law of Griggs and its progeny by including specific causation requirements in its analysis of the plaintiff's burden of proof, justifying its departures on the ground that employers must be protected from frivolous claims. The logic and structure of this opinion is thus not uncharacteristic of Supreme Court opinions.

Wards Cove, then, might be read as a simple reversal of policy concerning civil rights, a decision resting on grounds of public policy rather than considerations of statistical theory. To be sure, the Court's opinion makes no specific claims with respect to such theory. Yet disparate impact actions are so thoroughly intertwined with statistical evidence that the opinion is in effect a pronouncement on the probative value of statistical analysis. As such, the Wards Cove majority has placed itself in the unenviable position of challenging accepted statistical norms. The Court's strict guidelines regarding the plaintiff's data selection and proof of specific causation, combined with its lax burden requirements for the employer are such that scientifically significant findings may be rebutted in court by reasonable business justifications. This blatant imbalance is troubling, as the Court's framework sends the message to protected classes and employers that even when discriminatory impact has been determined to exist in the workplace, it will be tolerated. Wards Cove thus makes a bad system worse.

113 To put it more eloquently, "[s]ometimes the Court paints with broad strokes, splashing color across the canvas as boldly as any Abstract Expressionist. Sometimes it draws delicately, the results as fragile and fine as a Persian miniature, precise as Pointillism, or filled with Renaissance shadings of light and shadow." Greenhouse, Supreme Court's Style: A Seeming Richness in Diversity, N.Y. Times, Mar. 7, 1990, at A18, col. 5.

114 See Wards Cove, 109 S. Ct. at 2125.

115 Indeed, the presence of statistical reasoning in the Wards Cove opinion is not an isolated occurrence. The rise in the prestige and authority of scientific theory is eroding the conception that law is an autonomous discipline. See Posner, The Decline of Law as an Autonomous Discipline: 1962-1987, 100 HARV. L. REV. 761, 772-73 (1987).

116 The traditional requirements for satisfying the disparate impact prima facie
Although the *Wards Cove* holding is statistically troubling, it is only the most recent of a number of cases that have misapprehended the role of statistics in the law.\textsuperscript{117} Courts can hardly be faulted for misapplying statistics in the legal context, however, since law and statistics do not mix well.\textsuperscript{118} The logic of legal reasoning is based on the use of terms such as the "burden of proof" and the "prima facie case," while the logic of statistical reasoning operates on quantitative evidence of probabilities, confidence coefficients, and statistical significance. When judges are confronted with statistical evidence, they are forced to define mathematically those terms of art so as to determine whether the quantitative evidence satisfies the legal standard.\textsuperscript{119} It is a nearly impossible task, as judges often misapprehend important statistical principles,\textsuperscript{120} disagree on the significance of legal terms,\textsuperscript{121} and sometimes operate on misguided assumptions.\textsuperscript{122} Indeed, studies in the quantification of standards of proof have shown that subjective interpretations of the standard vary widely among judges.\textsuperscript{123} The case also were difficult for the plaintiff to meet. Even under *Griggs*, the plaintiff's analysis usually had to satisfy a 95\% statistical significance cut-off rule. See *supra* note 56 and accompanying text.

\textsuperscript{117} See, e.g., E.E.O.C. v. Western Electric Co., 713 F.2d 1011, 1019 (4th Cir. 1983) (misconstruing standard deviation analysis); Davis v. Califano, 613 F.2d 957, 965 (D.C. Cir. 1979) (accepting plaintiff's statistical evidence as probative although it was inconclusive); see also Cassell v. Texas, 339 U.S. 282, 284-85 (1950) (undertaking an erroneous statistical analysis in a grand jury discrimination case); Commissioner v. Indiana Broadcasting Corp., 350 F.2d 580, 582-83 (7th Cir. 1965) (misunderstanding the "Poisson distribution"), cert. denied, 382 U.S. 1027 (1966).

\textsuperscript{118} See, e.g., Kaye, *supra* note 39, at 13 (arguing that "[i]n most endeavors concerned with the acquisition of knowledge, quantitative information is welcomed. . . . [while in] law, however, it appears sometimes that scientific or numerical evidence makes cases harder, not easier.").

\textsuperscript{119} This may be why "[t]he idea that formal hypothesis tests should or must be used to assist the judge or jury . . . is a recent phenomenon. Before 1970, almost no federal cases adverted to 'statistically significant' evidence." Kaye, *supra* note 30, at 1334.

\textsuperscript{120} See *supra* note 117.

\textsuperscript{121} See Goldberg, *On Legal and Mathematical Reasoning*, 22 JURIMETRICS J. 83, 85 (1981) ("[J]udges often disagree not only on specific axioms but on entire classes of axioms.").

\textsuperscript{122} See, e.g., id. ("The problem is not that judges sometimes make errors in logic; published mathematical proofs are not invariably free of that vice. The problem is rather the shabby conditions of the axioms used by the judge." (footnote omitted)).

\textsuperscript{123} See United States v. Fatico, 458 F. Supp. 388, 410 (E.D.N.Y. 1978) (reporting an informal survey of ten judges' quantitative interpretation of the burden of proof), aff'd, 603 F.2d 1053 (2d Cir. 1979), cert. denied, 444 U.S. 1073 (1980). *Fatico* indicates that although all of the judges translated the civil standard of proof into values in excess of 0.50, they gave values ranging from 0.76 to 0.90 for the criminal standard.
impact of these misinterpretations has, over time, been borne by the disparate impact plaintiff.

The purposes of legal and scientific reasoning also account for the imperfect nature of court decisions with respect to statistical evidence. Court decisions constitute an instrument of social policy that serves to draw lines between what is permissible and impermissible. For that reason, majority opinions must convey a sense of certainty and finality, to communicate the idea that they are writing the rule of law. The Supreme Court is particularly confronted with this problem, as its opinions will serve as a guide to lower courts. Statistical studies are, on the other hand, a form of free scientific debate designed to draw conclusions concerning a given area. Statisticians compile their data, select mathematical methods to analyze the data, and derive conclusions based on the analysis. Any of these steps may be and usually are criticized by experts who disagree on the mechanics of the statistical analysis. While such scientific debate is encouraged for its societal benefits of truth and understanding, Supreme Court opinions are rarely debated for this purpose.

Judges thus must have a solid understanding of statistical reasoning if they are to become experts in "jurimetrics." Yet because they are subject to case precedent and authority, they do not have the luxury to rule as statisticians. Theirs is a difficult role. Whether the disparate impact framework may effectively be modified to simplify this role is the subject of Part III, which proposes a more straightforward approach to the review of statistics in disparate impact litigation.

See id.; see also Simon & Mahan, Quantifying Burdens of Proof: A View From the Bench, the Jury, and the Classroom, 5 LAW & SOC'Y REV. 319, 324-25 (1971) (reporting similar values for the civil standard, but finding that about one third of judges believe the criminal standard to be equivalent to a probability of 1.0).


125 For the Supreme Court "[t]o decide a case too narrowly is to squander the Court's resources; to rule too broadly is to squander its legitimacy. The tension is inherent in the system of judicial review." Greenhouse, supra note 113, at A18, col. 5.

126 The debate over legal doctrine is not conducted on the premise—which holds true among scientists—that if the doctrine is shown to be illogical by its critics, its conclusions will no longer hold true and the doctrine will be abandoned. At best, the Court may subsequently overrule the doctrine, though rarely because it is unpopular. In any case, legal scholarly debate focuses instead on understanding and interpreting the Court's reasoning which is often left vague for the reasons outlined above. See id.
III. SIMPLIFYING DISPARATE IMPACT THEORY

A. The Objective of a Simpler Model

The main purpose of simplifying the disparate impact framework is to allow judges to make legal decisions in a consistent and straightforward fashion. Such an approach would partially relieve the plaintiff of the particularly heavy burden that she shoulders throughout the disparate impact case. In practice, this means devising a model in which the risks of error are borne equally by both the plaintiff and defendant throughout the litigation. The model must also offer fairly relaxed data and causation rules that are consistent with the preponderance of the evidence standard.

Since the touchstone of such a model is consistency with the traditional preponderance of the evidence standard, the model should dispel the fears of frivolous litigation raised by the Court in Wards Cove. Indeed, if the model is consistent with the traditional preponderance of the evidence standard, then frivolous litigation will be no more likely than it is in other types of civil litigation. The result of applying this standard is that, in short, the party that has displayed the more cogent analysis will prevail. Thus, the employer will also gain from any benefits realized by the plaintiff as she will never be drawn into court unless the plaintiff’s statistical case meets the preponderance of the evidence standard. The plaintiff’s ability to build such a case implies not an instance of frivolous litigation, but the existence of a wrong to be corrected.

Ironically, such a model must make minimal use of statistics to avoid the aforementioned problems of trying a case that draws on vague legal terminology and complex statistical evidence. It should embody the values of the community, so as to prevent the need for accuracy in statistics from distorting the legal decision. Indeed, any statistical or mathematical system of proof requires that certain assumptions be made as to how lines will be drawn between “right and wrong” or “true and false.” In the above analysis, for example, it was demonstrated that courts have forced plaintiffs to bear a particularly large degree of risk, by requiring that their

128 See, e.g., People v. Collins, 68 Cal. 2d 319, 320, 438 P.2d 33, 33, 66 Cal. Rptr. 497, 497 (1968) (noting that “[m]athematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not [be allowed to] cast a spell over him”).
statistical analyses satisfy a 5% statistical significance test. This requirement deviates from the principle behind the traditional preponderance of the evidence standard, which is the product of the values that have shaped our legal system. Thus, it may be argued that this requirement must be changed, as any model of statistical proof must be consistent with such values. Such a change will ensure that all civil cases are adjudicated according to the same standard.

Finally, an improved statistical model must also be straightforward, so that the important variables behind the analysis are not hidden behind a misleading “black and white” conclusion, such as that generated by hypothesis testing. Ultimately, the model must release the mind of the trier of fact from the confines of complex quantitative analysis, as “[t]he main role of models is not so much to explain and to predict—though ultimately these are the main functions of science—as to polarize thinking and to pose sharp questions.”

B. The Legal Precedent for Simplification

In recent years, scholars have devised a number of legal decision-making models in their pursuit of a more equitable legal system. At the same time, courts have been faced with the need to adjudicate complex cases of sometimes exceedingly large proportions, the resolution of which has required much statistical evidence. These modern-day legal Goliaths have been, for the most part, in the form of antitrust and products liability suits. Responding to this challenge, courts and scholars have devised a number of theories to encourage the prompt and fair resolution of such civil actions. The purpose of these algebraic, statistical, and probabilistic models is to simplify legal reasoning and improve the framework in which the legal system can use quantitative evidence. They provide a basis for developing a straightforward disparate impact framework model.

See supra notes 60-64 and accompanying text.

1. The Hand Negligence Model

One of the oldest instances of the use of simplified formulas in the legal process is Learned Hand's negligence model described in United States v. Carroll Towing Co.:\textsuperscript{131}

[T]he owner's duty, as in other similar situations, to provide against resulting injuries is a function of three variables: (1) The probability that [a moored ship] will break away; (2) the gravity of the resulting injury, if [the ship] does; (3) the burden of adequate precautions. Possibly it serves to bring this notion into relief to state it in algebraic terms: if the probability be called P; the injury, L; and the burden, B; liability depends on whether B is less than L multiplied by P: i.e., whether B is less than PL.\textsuperscript{132}

The model was subsequently applied in a number of negligence cases,\textsuperscript{133} and in the less obvious context of a First Amendment action.\textsuperscript{134} It provided a simple framework with which judges could interpret the often hazy notions of duty and breach. The pitfall of the model was its simplified reliance on the difficult task of estimating probabilities and harm. Yet it provided a necessary analytical framework, since "[w]ithout such a generalization, it would be impossible to resolve many theoretical and practical questions without tortuous and, in most cases, nonreplicable analysis."\textsuperscript{135}

2. The Expected Value Rule

An offshoot of the Hand model has emerged in the form of the "expected value rule."\textsuperscript{136} Expected value theorists question the

\textsuperscript{131} 159 F.2d 169 (2d Cir. 1947).
\textsuperscript{132} Id. at 173.
\textsuperscript{133} See, e.g., Burgess v. M/V Tamano, 564 F.2d 964, 981 (1st Cir. 1977) (applying model to supertanker oil spill), cert. denied, 435 U.S. 941 (1978); Andros Shipping Co. v. Panama Canal Co., 298 F.2d 720, 725-26 (5th Cir. 1962) (involving damages sustained when a ship ran aground); Rosenquist v. Isthmian S.S. Co., 205 F.2d 486, 489 (2d Cir. 1953) (applying model to seaman's personal injury suit); see also Comment, supra note 130, at 54-55 (reviewing the use of the Hand rule by the courts).
\textsuperscript{134} See United States v. Dennis, 183 F.2d 201, 212 (1950) (questioning "whether the gravity of the 'evil', discounted by its improbability, justifies [an] invasion of free speech as is necessary to avoid the danger"), aff'd, 341 U.S. 494 (1951); see also Comment, supra note 130, at 65-66 (discussing the apparent influence of the Hand model in the Dennis decision).
\textsuperscript{135} Comment, supra note 130, at 40.
A COHERENT DISPARATE IMPACT MODEL

traditional legal framework which provides full recovery when the probability of events having taken place is greater than 50%, and denies recovery when the probability is less than 50%. They advocate the expected value rule which calculates damages "by multiplying the probability that the defendant is culpable by the magnitude of the damages." If, for example, the damages are in the amount of $100 and the plaintiff has proven the harm to a 49% certainty she will recover $49, while under the traditional preponderance of the evidence standard she will not recover at all. If the plaintiff proves her case to 51% certainty she will recover $51, while under the traditional "all-or-nothing" preponderance of the evidence standard she would recover the full $100. Thus, the model strives to minimize Type I error, the risk of incorrectly finding for the defendant, and Type II error, the risk of incorrectly finding for the plaintiff. Such a goal is worthwhile, for the above example shows that under the preponderance of the evidence standard, a 2% difference in the degree of certainty that the defendant is culpable may be the difference between no recovery and full recovery for the plaintiff.

Analyses of the expected value rule have drawn mixed conclusions. Two scholars have conducted a series of tests to determine whether this model is preferable to traditional preponderance of the evidence theory and have concluded that neither model is preferable in all circumstances. Another has argued that the preponderance of the evidence standard should be applied except in situations in which the probability that the defendant caused the harm never changes from case to case. In that situation, the expected value theory is the error-minimizing alternative.

---

137 See id. at 1159-60.
138 Id. at 1160.
139 See supra notes 42-45 and accompanying text.
140 Professors Orloff and Stedinger conclude that [if the goal is to minimize the number of erroneously decided cases or the sum of the wrongful payments that are made, the preponderance-of-the-evidence rule emerges as the superior choice. However, if the goal is to minimize making large errors or to avoid bias in the distribution of the errors among plaintiffs and defendants, the expected value rule emerges as the superior choice. Orloff & Stedinger, supra note 136, at 1172.
142 See id. at 514 n.76 (stating that the expected value theory best applies in
The expected value rule suffers from some of the same problems as the Hand formula, namely its reliance on quantification of the harm and the probability that the harm was caused by the defendant’s negligence. Furthermore, it is unlikely to gain favor among the legal community because the preponderance of the evidence standard is ingrained in our legal tradition, and the injection of an additional issue would further complicate an already overburdened legal system.\footnote{The untraditional nature of the expected value rule makes its application in court unlikely, so that it does not provide a workable theoretical basis for a new disparate impact doctrine.} The untraditional nature of the expected value rule makes its application in court unlikely, so that it does not provide a workable theoretical basis for a new disparate impact doctrine.

3. Market Share Liability

In the early 1980s, an inferential model of causation gained favor in a number of courts. In Sindell v. Abbott Laboratories,\footnote{See Orloff & Stedinger, supra note 136, at 1174.} the plaintiff brought a class action against eleven drug companies that had manufactured and marketed diethylstilbestrol, commonly known as DES.\footnote{26 Cal. 3d 588, 607 P.2d 924, 163 Cal. Rptr. 132, cert. denied, 449 U.S. 912 (1980).} The plaintiff alleged having been harmed by the drug, but could not identify which company had manufactured the DES that had caused her injuries.\footnote{See id. at 593, 607 P.2d at 925, 163 Cal. Rptr. at 133.} The trial court dismissed the action, and the California Supreme Court reversed, holding that if the plaintiff could prove her allegations of negligence and damages, she would be entitled to recover from each of the companies in proportion to their percentage share of the statewide DES market.\footnote{See id. at 611-13, 607 P.2d at 937, 163 Cal. Rptr. at 145.}

The DES causation framework apportions fault on the basis of the actual market shares of the various defendant drug companies, thus relieving the plaintiff of the burden of identifying the company which actually manufactured and marketed the particular sample that was responsible for the plaintiff’s harm. Identifying the responsible defendant can be a painstaking process, as it requires situations in which there is a very “real possibility of repeated recovery from the same defendant,” such as in DES cases since in those cases, the single largest marketer of DES in California “could be held liable for every DES-caused injury in California”). For a discussion of market share liability, which may be used to solve the same problem, see infra notes 144-52 and accompanying text.
the plaintiff to determine the origin of the DES that harmed her. Responding to this practical difficulty, the court allowed the plaintiff to prevail based on a showing of negligence and harm, inferring the causation element of her tort claim.

One of the problems of the market share liability model is that it does not focus on the probability that a particular defendant caused the harm. As the dissent in Sindell indicated, "a particular defendant may be held proportionately liable even though mathematically it is much more likely than not that it played no role whatever in causing plaintiffs' injuries."148 Three scholars have advocated adopting a probabilistic approach to multiple causation harms.149 One proposed model would apportion fault among multiple tortfeasors on the basis of their "probabilistic marginal product,"150 that is, "their relative marginal contributions to the risk of harm suffered by the plaintiff."151

The pitfall of the DES probability model is that both parties are still required to calculate probabilities which may not be very accurate. Furthermore, probability analysis does nothing to alleviate the war of statistics between the plaintiff and the defendant, as both parties will introduce statistical evidence in court to buttress their probabilistic assessments. Nevertheless, the probability model is a more intuitively understandable framework because it avoids the black hole of hypothesis testing.

The difficulties of proof faced by a DES plaintiff are not unlike those of a disparate impact plaintiff. Determining the exact composition of the statewide market for a widely used pharmaceutical product is no more difficult than isolating the precise origins of a rule that causes a disparate impact on protected groups and showing specifically how it causes such an impact. The wide

---

148 Id. at 616, 607 P.2d at 939, 163 Cal. Rptr. at 147 (Richardson, J., dissenting).
150 Rizzo & Arnold, supra note 149, at 1406 n.39.
151 Robinson, supra note 149, at 759. Such a model may be applied to a disparate impact hypothetical. Assume a plaintiff claims that the employer has turned down her job application for a position as a security guard on the basis of a hiring procedure that has a disparate impact with regard to gender. If the court finds that the relative apportionment of probable causes of the disparity in workforce hiring is 65% due to a height requirement that is not a valid business justification, 10% due to the structure of the local union, and 25% due to the justified need to hire individuals without a criminal record, then under a multiple causation probabilistic model, the plaintiff will recover 65% of her damages.
discrepancy in the standards required of DES claimants and those required of disparate impact plaintiffs supports the argument that the legal system is less willing to insulate the employer/manufacturer from claims by consumers than from those by job applicants. Consistency in the law requires that the relaxed causation standards must also be applied to disparate impact plaintiffs. Adoption of the DES causation framework is an important precedent for formulating a simpler and fairer disparate impact framework.

C. The Value of Probabilistic Analysis

The analysis in Part I of this Comment regarding the use of hypothesis testing in the courtroom indicates that its complexities and statistical nuances are ill-suited to judicial decision-making. Even if the system were improved to guard against the statistical fallacies perpetuated over the years by courts, it is likely that such an overhaul would merely complicate the already confused evidentiary process of disparate impact cases. The solution is to search for a new framework that will both conform to the conventional standard for the burden of persuasion applied in civil trials and allow the judge to interpret inferential statistical evidence more easily. Thus, such a framework should prevent the judge from being overwhelmed by complex statistical analyses whose effect may be to convey the often false impression that they are always accurate.

The above review of alternate statistical theories indicates that the statistical evidence should take the form of likelihoods, thus lessening the current focus of courts on significance levels and hypothesis testing. Probabilistic analysis is intuitively more understandable than hypothesis testing, although the latter theory has been more widely used by courts. The relative disuse of probability theory is likely due to the legal system’s fixation on “black and white” decisions of culpability, similar to the bottom-line

---

152 Neither the traditional Griggs nor the current Wards Cove disparate impact framework reaches even the minimum level of fairness applied in standard civil trials. See supra notes 60-64 and accompanying text. To quote Justice Stevens, “[o]rdinary principles of fairness require that Title VII actions be tried like ‘any lawsuit.’” Wards Cove Packing Co. v. Antonio, 109 S. Ct. 2115, 2133 (1989) (Stevens, J., dissenting).

153 See, e.g., Aickin, supra note 28, at 385 (“In trying to cope with the infusion of data analytic technicalities into the courts, lawyers and judges have tried to become statisticians. They would do much better by trying to become probabilists.”).
conclusions of hypothesis testing. Use of a probabilistic framework will allow the judge both to gauge correctly the somewhat tenuous nature of the statistical evidence and to avoid stretching the meaning of such evidence beyond the interpretation given by statisticians. Such a result will neither deter plaintiffs from satisfying the prima facie case nor discourage defendants from rebutting it, since the framework will impose on both parties the probabilistic interpretation of the preponderance of the evidence standard. Thus, the model will accomplish its ultimate goal, namely peaceful cohabitation of legal constructs and statistical evidence in the disparate impact trial process.

1. Quantifying and Applying the Burden of Proof

Advocates of probabilistic decisionmaking generally agree that the preponderance of the evidence standard applied in civil trials translates in probabilistic terms to a test in which the court rules for the plaintiff if the probability of the existence of facts indicating liability exceeds 0.5. Thus, the plaintiff satisfies her burden by introducing evidence that satisfies this test. Although such a test would appear to be easily applied in the courtroom, some scholars indicate that quantification of the burden of proof does not fit into the concept of legal decisionmaking.

154 See supra text accompanying note 32; cf. M. ZIMMER, C. SULLIVAN & R. RICHARDS, supra note 28, at 117 (stating that “[p]robability is the basis of the science of statistics,” and therefore of hypothesis testing). In other words, the probabilities generated by statistical analysis should not be analyzed in the traditional hypothesis testing framework. Hypothesis testing generates misleadingly black and white conclusions that an original null hypothesis of no discrimination must either be accepted or rejected. An improved framework will examine the statistical analysis from a purely probabilistic standpoint, thus allowing the trier of fact to rank the analysis on a spectrum of confidence.

155 The utility of a probability function stems from the fact that it “contains all the information that the statistical findings can convey. It forces the fact-finder to confront all the hypotheses concerning the parameters of a probability distribution, and it avoids the arbitrary character of hypothesis testing at a uniform significance level.” Kaye, supra note 35, at 852.

156 See, e.g., Kaye, Naked Statistical Evidence (Book Review), 89 YALE L.J. 601, 609 (1980) (“I wish to argue for the orthodox view that a party shoulders its burden of proof by a preponderance of the evidence by persuading the finder of fact that the probability in question exceeds one-half.”); see also Cohen, supra note 45, at 394; Kaye, Naked Statistical Evidence, 1982 AM. B. FOUND. RES. J. 487, 493; Martin & Schum, Quantifying Burdens of Proof: A Likelihood Ratio Approach, 27 JURIMETRICS J. 383, 388 (1987) (quantifying the criminal standard of the burden of proof).
This problem may be illustrated with an example. Suppose 499 of 1,000 people in attendance at a rodeo have paid the admission price, while the remaining 501 are gatecrashers. Assume further that there is no way of distinguishing the paying customers from the gatecrashers. If the proprietor of the rodeo brings suit against any spectator to recover the lost admission charge, and if the only pieces of evidence are the above statistics of paying and non-paying audience, then according to the probability model, the proprietor will presumably recover against all those in attendance. This example, known as the "paradox of the gatecrasher," demonstrates how probabilistic analysis may run counter to accepted judicial principles which would prohibit recovery in this instance.

This argument highlights the difficulties of applying statistical and mathematical models in the legal system. Such models, however, need only be used as a guide rather than a tool that will be dispositive of the outcome of a case. Indeed, one advantage of probability theory is that it highlights possible fallacious court rulings that might remain hidden if the court were solely dealing with evidence in the form of hypothesis testing. Thus, although the paradox of the gatecrasher does present a situation in which a guilty verdict will cause injustice to the paying audience, it does not weaken the case for using probability analysis in court. Probability analysis is not as desirable for its accuracy as it is for its clarity. The strength of probability analysis is that the trier of fact will be better able to identify borderline situations such as that of the gatecrasher hypothetical. Once the probability is identified, it will be an additional piece of evidence that the trier of fact will consider in her decision-making process.

The temptation to employ probabilistic models as black boxes rather than guideposts stems from the perception that mathematical analyses generate pinpointed solutions to problems. This, however, stems from "confusion between the precision that is characteristic

---

159 See supra note 28. Such an approach may be problematic when dealing with so-called "pure impact" employment discrimination cases, in which only statistical evidence is presented at trial. Since no other evidence is available in such a case, probabilistic evidence becomes determinative of the outcome of the suit. The problem of analysis of "naked statistical evidence" is discussed infra notes 188-96 and accompanying text.
of mathematical analysis of uncertainty, and precise quantitative or numerical values." Indeed, it is possible to give a mathematical statement assessing a probability between 30% and 40%, as mathematical analysis does not necessarily require exact precision. From this perspective, probabilistic analysis is most useful in employment discrimination litigation. The probabilities may guide, not enslave the trier of fact, if they are used as general indicators of the probability that employment discrimination has taken place.

2. The Subjectivist/Frequentist Controversy

Use of probabilities in law is also debated on philosophical grounds by the two major schools of probabilistic theory. The belief that one must be cautious in interpreting statistics is shared by the "subjectivist" school of statisticians. Subjectivists argue that a statistically measured probability is actually a measure of the statistician’s "degree of belief that the event . . . is or will turn out to be true." Thus, subjectivist probabilities may be viewed as "the odds an individual would be willing to accept in betting on the outcome of an event . . . ."

Frequentists, on the other hand, believe that the probability indicates the relative frequency with which the predicted outcome would be observed if the operation were repeated a large number of times. Thus, while the frequentist interpretation of a conclusion that an event has a certain probability of occurrence is in the form of an objective prediction about future repetitions, the subjectivist would claim that the probability is a "property of the observer's perception about that particular opportunity for the event to occur, and has no reference to hypothetical repetitions of the opportunity." Thus, in the employment discrimination context, the subjectivist would report a probability of discrimination with the caveat that the probability represents a measure of

---

161 See id. at 385 n.11.
162 A scale could, for example, be devised through which probabilistic evidence would be rated by the court on a scale of five levels of discrimination (0-19.99%, 20-39.99%, 40-59.99%, 60-79.99%, 80-100%).
163 See id. at 350.
164 Kaye, supra note 158, at 43.
165 See id. at 42.
166 Aickin, supra note 28, at 350.
subjective belief. By formulating somewhat modest conclusions, the subjectivist approach recognizes that statistical analysis will always be inherently biased. Such an approach might seem to advocate a more rigorous review of statistical analyses presented in court, since it recognizes that no finding of discrimination based on statistical data is truly objective. In fact, subjectivists simply recognize that different people may be presented with the same evidence and still come to opposite conclusions.

D. A Bayesian Disparate Impact Framework

1. Applying Bayes' Theorem in Disparate Impact Cases

One statistical model advocated by subjectivists for its ability to embody the subjective nature of statistical evidence is Bayes' theorem. Bayes' theorem, a theory of statistical inference, combines prior and posterior probabilities to determine the likelihood of an event's occurrence. It has been thought of as a "straightforward device for assessing the probative value of certain evidence that might otherwise be difficult to assess," mainly because of its manipulation of "subjective belief," that is, the prior personal probabilistic belief of the decision-maker. Because Bayesian analysis is governed by certain advanced mathematical principles, the following explanation of Bayes' theorem is accompanied by an illustrative example that has been constructed to facilitate understanding of the model.

The hallmark of Bayes' theorem is that it allows one to process statistical data by incorporating previous experience into the statistical analysis. This is undertaken by first determining the prior distribution, that is, the set of probabilities representing the judge's prior subjective belief with respect to the problem being analyzed. The likelihood function, which is the set of probabilities that are

---


168 Faigman & Baglioni, supra note 167, at 2 n.1.

A COHERENT DISPARATE IMPACT MODEL

derived from the available data, is then determined. The product of the prior distribution and the likelihood function yields the posterior distribution, that is, the set of probabilities that incorporates the judge's subjective knowledge into the probabilities directly derived from the data.

As an example, assume that a corporation currently employs 200 welders, 60 of whom are black. An analysis undertaken by the plaintiff indicates, however, that 40% of the welders in the region are black. The plaintiff's suit against the corporation, based on both disparate treatment and disparate impact, alleges that its hiring practices for welders discriminate against blacks. The plaintiff's figures indicate that the employer's welder hires should be approximately 40% black to conform to the racial breakdown in the area, assumed to be that of the qualified applicant pool of welders for the purposes of this example.

The first step in the Bayesian analysis is to determine the prior distribution in an employment discrimination context. This is the prior subjective belief of the judge that there has been hiring discrimination. This prior distribution is determined before the likelihood function is known—that is, before the statistical analysis is presented. In a mixed disparate impact/disparate treatment claim, the prior distribution is a function of qualitative evidence of discrimination presented at the trial. Because such evidence is "soft," the main difficulty in a Bayesian analysis consists of accurately quantifying such evidence in order to construct the prior distribution. The problem of converting subjective belief into probabilities is compounded by the need for the judge to gauge her subjective belief of the degree of discrimination for every possible degree of discrimination. In our example, the judge must assess the probability that the actual degree of discrimination equals each of the possible degrees of discrimination between 0% (complete

---

170 This evidence could, for instance, take the form of testimony of members of the employer's workforce. Such evidence is often presented in mixed disparate impact/disparate treatment cases. See, e.g., supra note 76. The remainder of the discussion will assume that the court is presented with a mixed impact/treatment claim. The problem of modeling a prior distribution in a pure impact claim (with no qualitative evidence available) is treated infra notes 188-96 and accompanying text.

171 See Tribe, supra note 127, at 1361-62 (defining soft variables as those that are not readily quantifiable).

172 The difficulties of modelling the prior distribution have been cited as one problem associated with applying Bayes' theorem in court. See D. BALDUS & J. COLE, supra note 12, at 504-05; Kaye, supra note 35, at 855; see also G. IVERSEN, supra note 169, at 70-71.
discrimination—no black hires) and 40% (no discrimination—hiring that is representative of the qualified applicant pool). Once this is done, the judge will have determined the prior distribution which may then be graphed.\(^\text{173}\)

A simplified approach to quantifying the prior distribution is outlined by Professor Iversen in his paper on Bayesian statistical inference.\(^\text{174}\) The starting point is the standard formula for the prior distribution,

\[
f(\Theta) = C \Theta^{a-1} (1-\Theta)^{b-1}
\]

where \(f(\Theta)\) is the prior distribution, \(\Theta\) is the unknown degree of discrimination, \(C\) is a constant, and \(a\) and \(b\) are non-negative constant integers.\(^\text{175}\) To determine the prior distribution, \(C\), \(a\) and \(b\) must be quantified. These may be derived from the mean and variance of the distribution.\(^\text{176}\) The mean of the prior distribution represents the expected value of \(\Theta\). In the above example, it is assumed to be 20%. That is, the judge believes that the most likely degree of discrimination is 20% (i.e., the corporation hires 20% black welders). Furthermore, if the judge believes that most of

\(^{173}\) The degree of the employer’s discrimination is unknown. To find the “most likely” degree of discrimination, the likelihood of each of the possible degrees given the observed data must be determined. Although this procedure may appear to be complex and time consuming, it may be simplified by constructing “probability intervals.” If the judge is quite certain that the degree of discrimination lies between 10% and 30%, she may construct a bell shaped curve representing the prior distribution. Such a procedure may appear to be exceedingly imprecise—it may, however, be no less so than the actual subjective belief of the judge which will often be a simple “hunch” as to the degree of discrimination. 

\(^{174}\) See G. IVERSEN, supra note 169, at 22-29.

\(^{175}\) This formula is the general expression of the probability density known as the beta distribution.

\(^{176}\) A property of the beta distribution is that \(a\) and \(b\) may be derived from the mean and variance, as

\[
a = \mu \left[ \frac{\mu(1-\mu)}{\delta^2} - 1 \right] \quad \text{and} \quad b = [1-\mu] \left[ \frac{\mu(1-\mu)}{\delta^2} - 1 \right]
\]

where \(\mu\) is the mean and \(\delta^2\) is the variance—which is the square of the standard deviation—of the distribution. When \(a\) and \(b\) are integers, the constant \(C\) may also be derived, as

\[
C = \frac{(a + b - 1)!}{(a - 1)! (b - 1)!}.
\]

G. IVERSEN, supra note 169, at 22-23.
the probability of the prior distribution lies between 15% and 25%, then it may be assumed that there is a distance of two standard deviations from the mean to 15% and 25% so that the standard deviation is 2.5%. Using these figures and the above formulas to determine a and b yields a = 51 and b = 204. C may then be derived from a and b. This constant is, however, a much larger figure which is of little use except to graph the curve for f(θ). These figures may then be used to derive the prior distribution, which is

\[ f(\theta) = C\theta^{50}(1-\theta)^{203}. \]

The second step of this Bayesian analysis is to derive the likelihood function from the data provided in the example. If there is a fixed probability θ that the employer will have a black welder every time, and if independence between observations is assumed, the likelihood function will take the form of a binomial distribution, so that

\[ g(\theta/x) = nC^x_x(1-\theta)^{n-x} \tag{2} \]

where \( g(\theta/x) \) is the likelihood function, \(_nC^x_x\) is the binomial coefficient, and \( x \) is the number of black welders hired in \( n \) hires. In our example, \( x = 60 \) and \( n = 200 \), so that

---

177 When the distribution is a bell-shaped "normal" curve, 95% of the probability lies within two standard deviations of the mean. Thus, the "2 or 3 standard deviation" rule... permits one to assess statistical significance in terms of the ratio of the estimated disparity to its standard error or standard deviation." D. BALDUS & J. COLE, supra note 12, at 294-95.

178 See supra note 176.

179 See id.

180 See G. IVERSEN, supra note 169, at 25.

181 The binomial model is a mathematical model that gives the probability that a postelection pool will contain \( x \) minority candidates if (a) the size of the combined postelection pool is fixed, (b) the minority representation rate in the preselection pool is known or hypothesized, (c) members of the postelection pool are chosen through a completely random process, and (d) the preselection pool is much larger than the postelection pool, exceeding the latter by a factor of ten or more.

D. BALDUS & J. COLE, supra note 12, at 350.
The product of this likelihood function and the prior distribution yields the posterior distribution, that is

\[ f(\theta|x) = f(\theta)g(\theta|x), \]

where \( f(\theta|x) \) is the posterior distribution. In our example, the posterior distribution may be derived as

\[ f(\theta|x) = C''\theta^{110}(1-\theta)^{348} \]

where \( C'' \) is a constant.\(^{182}\) The new mean and standard deviation for \( f(\theta/\mu) \) may then be derived from the function, so that \( \mu = 24\% \) and \( \delta = 2.0\% \).\(^{183}\) These figures represent the results of the Bayesian analysis, indicating that the mean degree of discrimination after having taken into consideration prior subjective knowledge and the likelihood function derived from statistical evidence is 24%. If the derived posterior distribution is a normal distribution, that is, it is bell-shaped,\(^{184}\) then a probability estimate may be calculated from basic statistical analysis,\(^{185}\) resulting in a finding that there is a .95 probability that the unknown degree of discrimination \( \theta \) lies between 20% and 28%, thus clearly pointing to the guilt of the defendant.

The effect of the prior distribution—that is, the subjective belief of the judge—may be seen by comparing the posterior distribution to the likelihood function derived from the statistical evidence. In

\[^{182}\text{The exponent for } \theta \text{ is derived by adding the exponent for } \theta \text{ in } f(\theta) \text{ and } g(\theta|x). \text{ The same method is used to derive the exponent for } (1-\theta). \text{ The constant } C'' \text{ is equal to the product of the binomial coefficient in } g(\theta|x) \text{ and constant in } f(\theta). \text{ Because the posterior distribution is a beta distribution, see supra note 176 and accompanying text, } C'' \text{ may be derived from } a \text{ and } b \text{ to graph the posterior distribution.}\]

\[^{183}\text{When } f(\theta|x) \text{ is a beta distribution,}\]

\[ \mu = \frac{a}{a + b} \text{ and } \delta^2 = \frac{\mu(1 - \mu)}{a + b + 1}. \]

Solving for \( a = 111 \) and \( b = 344 \), one derives \( \mu = 24\% \) and \( \delta^2 = .0004 \) (adjusted for rounding). The standard deviation is the square root of the variance so that \( \delta = .0201. \)

\[^{184}\text{See supra note 177 (describing the characteristics of a normal curve).}\]

\[^{185}\text{Because we are dealing with a normal distribution, there is a .95 probability that } \theta \text{ lies } \pm 1.96 \text{ standard deviations from the mean, so that there is a .95 probability that } 20\% < \theta < 28\%.}\]
the likelihood function, the estimated degree of discrimination was 30% and the standard deviation was 3.2%. Incorporating the qualitative evidence has in effect lowered the mean, as such qualitative evidence seemed to point towards a stronger degree of discrimination than did the quantitative evidence embodied in the likelihood function. In short, the Bayesian analysis has allowed the judge to incorporate all forms of evidence presented at the trial into a probability distribution illustrating the likelihood of each possible degree of discrimination. This benefit of Bayes' theorem is one of the reasons why some scholars advocate using Bayesian models in the courtroom.\footnote{See, e.g., Faigman & Baglioni, supra note 167, at 16 (concluding that "an expert's Bayesian formulation will not overwhelm the average trier of fact"); Fienberg & Schervish, supra note 167, at 795 (suggesting a system of training that would facilitate the application of Bayes' theorem). But see D. BALDUS & J. COLE, supra note 12, at 304-05 (stating that both the need to construct a prior distribution in a Bayesian analysis and lack of familiarity of courts with Bayesian theory are such that it should not be applied in the courtroom); Kaye, supra note 35, at 855 (concluding that Bayes theorem would not provide much additional guidance to the factfinder); Kaye, supra note 54, at 779 (asserting that the costs of Bayesian analysis outweigh its benefits).}

These results must not, however, overshadow some of the structural difficulties involved in applying Bayes' theorem to employment discrimination cases. Indeed, a common criticism of applying Bayes' theorem in the courtroom is that quantifying the prior distribution is a difficult task.\footnote{See, e.g., D. BALDUS & J. COLE, supra note 12, at 304 n.32 (stating that the problem with applying Bayes' theorem in the courtroom is that "there is no judicially acceptable way we know of to quantify an essential variable in the procedure (the 'prior distribution')").} Thus one could argue that the above quantification of subjective belief was undertaken in an oversimplified fashion so that the resulting prior distribution is inaccurate. One response to this criticism is that the subjective belief of a judge will often remain rather vague, so that merely approximating the prior distribution—for example, by constructing a bell shaped curve around the judge's best estimate of the mean degree of discrimination—is sufficient for the purposes of the Bayesian analysis. If the judge's prior belief is too vague or uncertain to engage in such an exercise, then the problem becomes essentially one of representing ignorance.
The problem of modeling ignorance arises in pure disparate impact cases, in which "soft" non-statistical evidence of employment discrimination is unavailable. In such a case, there is no data from which to specify the prior distribution required in a Bayesian analysis. The problem is compounded by the fact that there is no accepted statistical method through which to represent complete ignorance in a prior distribution. One proposed solution is to refrain from modelling subjective belief with a prior distribution and instead to construct a series of prior distributions and to examine their impact on the data through the posterior distributions. Such a procedure would allow the judge to gauge the probative value of the data by examining the effect of the various prior distributions on the likelihood function derived from the statistical evidence.

It may be that there will rarely be a situation in which the disparate impact judge will be faced with complete ignorance. Indeed, although cases are generally tried in a vacuum, so that, for example, the only admissible evidence in a pure disparate impact case will be statistical in nature, judges are often influenced by other factors. Statutory construction and stare decisis may both provide information for the judge to model the prior distribution. If, for example, the case at hand resembles, by way of statistical evidence, other cases presented to the judge, she will inevitably be influenced by her prior decisions.

Finally, one may question the value of applying Bayesian theory in instances in which the posterior distribution is heavily determined by the statistical evidence presented in the case.

---

188 The notion of ignorance may be examined from a Bayesian perspective as distinguishing between "informative and noninformative prior distributions." G. Iversen, supra note 169, at 62.
189 See supra note 171.
190 See, e.g., Kaye, supra note 54, at 779 (asserting that "the use of the uniform distribution to represent complete ignorance is subject to attack on philosophical grounds, on mathematical grounds, and on legal grounds" (citations omitted)).
191 See id. Kaye feels, however, that this procedure would not add much to a presentation of statistical evidence in the form of likelihood functions. See id.; see also Kaye, The Probability of an Ultimate Issue: The Strange Case of Paternity Testing, 75 IOWA L. REV. 75, 94 n.94 (1989) (concluding that "[t]he classic question of how to represent complete ignorance within the established theory of probability admits of no unique answer, at least on mathematical grounds").
192 This is somewhat similar to "replicating or building on an old study where the analysis was done using Bayesian inference. In such a case we would simply take the posterior distribution from the previous study and use it as the prior distribution in the new study." G. Iversen, supra note 169, at 62-63.
193 Generally, if the judge has very little prior information about the degree of
Applying Bayes’ theorem in such pure impact cases poses a problem, however, only if the posterior distribution indicates by a slim margin that the defendant is guilty—that is, if the distribution in the above example indicates that the probability that \( \Theta \) lies between 20% and 28% is only slightly greater than 50%. Critics of the theory thus argue that to rely largely on statistical evidence that narrowly points to the guilt of the defendant thwarts the purposes of the legal process.\(^{194}\) One may respond, however, that there is no reason to believe that naked statistical evidence is inferior to soft qualitative evidence.\(^{195}\) To put it differently, there may be instances in disparate treatment cases in which the defendant is found guilty, although the evidence only slightly points to a greater than 50% probability of guilt. The problem is that the tenuousness of statistical evidence is more easily discernible than that of qualitative evidence. According to the rules of the preponderance of the evidence test, the plaintiff should prevail as long as the evidence—regardless of its nature—points to a greater than 50% probability of guilt. Thus, it may be the assumptions and values underlying such a test that make certain critics uncomfortable. Practical solutions may be devised whereby the court would require the plaintiff to show a probability of guilt of 60% or more to prevail. Legal purists may claim that such a test would taint the legal process. The problems of the current system indicate, however, that it is preferable to recognize the limits of statistical and legal theory by devising a pragmatic framework that attempts to minimize errors.

Thus, the philosophy behind Bayes’ theorem is useful in pure impact cases, as it focuses on clarifying the relationship between probability and proof. In a more pedagogical sense, it also emphasizes the values and assumptions behind the preponderance of the evidence standard of civil trials by highlighting some of the perhaps troubling implications of the 50% test. It may be disconcerting to realize that a defendant has been found guilty because the probability of her guilt was 50.01%. Yet this concern should not lead us to discriminate against statistical evidence as opposed to other forms of evidence, since qualitative soft evidence may be

\(^{194}\) See supra notes 157-58 and accompanying text (discussing the paradox of the gatecrasher).

\(^{195}\) See Brilmayer, Second-Order Evidence and Bayesian Logic, in Probability and Inference in the Law of Evidence 151 (P. Tillers & E. Green eds. 1988).
equally tenuous, and equally inculpatory. It should prompt us to question whether we have applied the preponderance of the evidence test consistently, or whether instead we have increased the standard of proof in those cases for which we are too uncomfortable to apply the traditional test.¹⁹⁶

2. Bayes' Theorem as a Normative Model

The Bayesian legal model is not without its critics. From a practical standpoint, some have argued that the trier of fact is too unsophisticated in the complexity of mathematical techniques to understand the formula and its significance,¹⁹⁷ or that any statistical evidence may be simply ignored.¹⁹⁸ Indeed, the analysis in Parts I and II of this Comment indicates the extent to which courts have misunderstood basic statistical analysis. Thus, skeptics have argued that since most individuals have little or no experience juggling probabilities, using Bayesian probabilities in the legal setting is really counter-intuitive; unaided intuitive probabilistic judgments are sufficiently accurate for legal purposes.¹⁹⁹ Others have contested the theoretical foundations of Bayesian probability analysis, claiming that Bayes' theorem only functions if conditional independence of evidence is assumed.²⁰⁰ These may be some of the reasons that relatively few courts are known to have applied Bayes' theorem to determine relative guilt or innocence.²⁰¹

¹⁹⁶ See supra notes 127-29 and accompanying text (discussing the impact of societal values on legal standards).
¹⁹⁷ See, e.g., Tribe, supra note 127, at 1361-65 (arguing that the factfinder may not be sophisticated enough to give appropriate weight to non-quantitative evidence when simultaneously faced with the seemingly more exact quantitative analysis provided by techniques such as Bayes' theorem).
¹⁹⁸ See, e.g., Saks & Kidd, Human Information Processing and Adjudication: Trial by Heuristics, 15 Law & Soc'y Rev. 123, 149 (1981) (suggesting that "[t]he more realistic problem is presenting statistical evidence so that people will incorporate it into their decisions at all").
¹⁹⁹ See Fienberg & Schervish, supra note 167, at 782.
²⁰⁰ If the evidence is not conditionally independent, it is impossible to update the odds on guilt versus innocence after each incorporation of statistical evidence, since introduction of the new evidence would affect the weight of the prior evidence that was supposed to have already been established, and on which the current odds have been calculated. See id.
²⁰¹ See, e.g., D. BALDUS & J. COLE, supra note 12, at 305 n.32 (claiming that they "know of no court that has applied Bayesian methods to an evidentiary problem"). But see Kaye, supra note 35, at 854 n.69 (1982) (citing law review articles that refer to cases that have applied Bayesian methods to a legal problem, and also citing Everett v. Everett, No. D-850-370 (Super. Ct., Los Angeles County, Cal. 1981); Arizona v. Wagner, No. DR122023 (Super. Ct., Maricopa County, Ariz. 1980)); Kaye, supra note
Most of these criticisms presuppose that the Bayesian model will be used in court solely to generate a probability estimate that will determine whether the burden of persuasion has been met. This application of the theory is misguided, as it replicates the fallacy of using a statistical method for the sole purpose of deriving a legally significant figure. To avoid the problems of relying on statistics that must always be based on assumptions regarding the source of the data or the manner in which they have been manipulated, judges must understand that Bayes' theorem is not a "black box." It must be used instead as a theoretical framework that will compel the judge to isolate her subjective belief from any statistical evidence, thereby preventing the overall picture of the case from being transformed by an excessive reliance on statistics. Thus, it is more important that judges accept the philosophical underpinnings of Bayes' theorem rather than learn how to generate probabilities. If they understand the basis of Bayesian theory, they will commit fewer errors of basic misinterpretation when handling statistical evidence.

In this sense, the theory of Bayesian inference may be viewed as a normative model. As such, it is one that provides a framework for analyzing the process of proof in employment discrimination trials. This function may be decried as making the legal process of proof a prisoner of a statistical theory, and therefore distorting the meaning of established judicial principles. At the very least, however, Bayesian theory may also be used in a more conservative manner as a "heuristic device" to instruct judges on how to model normative legal rules. In either case, Bayes' theorem fleshes out some of the assumptions behind the use of statistics in the legal process.

3. The Justification for Isolating Subjective Belief

To some, the courtroom may appear to be a strikingly inappropriate forum in which to apply Bayesian theory. Indeed, the legal system is ingrained with the notion that judges must not inject their personal values into the jurisprudential exercise. That function belongs to legislators, who enact the laws of the nation based on their personal beliefs and value systems. This principle is particular-

191, at 77 (citing paternity cases as examples of cases applying Bayesian analysis).


203 See id. at 446-48.
ly difficult to follow in Title VII cases, where the judge takes on the role of the jury, as she is the trier of fact. Still, any misgivings would be unfounded, as the subjective belief element of Bayes' theorem does not refer to personal values per se, but to the trier of fact's instincts with respect to the guilt or innocence of the defendant. Bayes' theorem simply recognizes that the trier of fact will ultimately make the final decision with respect to the employer's liability.

The issue is nonetheless an important one, as the deconstructive formulation of Bayes' theorem uncovers the difficulty of abiding by the cardinal principle that judges must be objective arbiters of the law. The psychology of decisionmaking is such that different judges apply different standards when adjudicating cases. The weight given to the qualitative and quantitative evidence in a Title VII action will vary with the judge who is trying the case. Thus, "[t]he legal realist may suggest that a jurist simply expresses his personal vision," so that, for instance, the Court's choice of strict statistical proof in *Wards Cove* may have "mirror[ed] its individual predilection and ethical regimen." Arguably, then, isolating subjective belief merely stands as a statement regarding the relative frailty of statistics-based judicial decisions.

It may be that the subjective belief of judges adjudicating disparate impact cases will reflect an inherent reticence to finding an employer guilty of discrimination, so that the end result for all disparate impact plaintiffs is the same. One could argue, for example, that the *Wards Cove* majority's subjective belief of guilt on the part of the employer was low even before any statistical evidence was introduced. The subjectivist would respond that with the deconstructive Bayesian framework, the Court might not have ignored "evidence of racial stratification [that] puts the specific employment practices challenged by [plaintiffs] into perspective." A second response is that isolating subjective belief may induce courts to attempt to gain greater understanding of the legislative mandate of Title VII. Sorting through statistical evidence and weighing the flaws of mathematical studies is not a simple task.

---

204 *See supra* note 24.
205 *See supra* note 123 and accompanying text.
207 *Id.* at 10-11.
208 *Wards Cove*, 109 S. Ct. at 2135 (Stevens, J., dissenting).
It forces the judge to stray from the intuitive subject matter of the action at hand. Applying Bayes' theorem in the disparate impact courtroom would guide the judge through a morass of numbers, by inducing her to respect her prior beliefs, which are inevitably influenced by the values behind Title VII. The Bayesian principle of subjective belief reminds the judge that the final ruling must be premised on the disparate impact directive of Title VII, namely, that it is unlawful "to limit, segregate, or classify ... employees or applicants for employment in any way which would deprive ... any individual of employment opportunities ... because of such individual's race, color, religion, sex, or national origin."209

CONCLUSION

The indisputable need for statistical analysis in employment discrimination litigation poses a dilemma for lawyers, judges, and statisticians. The decision to allow the plaintiff to use statistics in proving a disparate impact employment discrimination case is in itself a value-laden one that stems from the equal opportunity mandate of Title VII. But the adjudication of disparate impact cases implies a further choice of rules, as it requires developing a set of statistical guidelines for plaintiffs. That choice of statistical guidelines "may itself evidence and indeed constitute a change in the mix of basic values of the society that has made the choice in question."210 The choice of statistical requirements for plaintiffs is, then, not so much an objective one, as it is a value-laden one. Over time, that choice has been made by the courts, whose most recent set of rules has drastically altered the value system that had originally encouraged plaintiffs to use statistics in disparate impact cases. As this Comment has shown, many statistical rules are applied to the benefit of the employer, and their egregious effect on the plaintiff often remains cloaked under the misleading guise of objective reasoning. Applying a Bayesian analysis will ensure both that judges better understand the basic statistical principles with which they must work in disparate impact cases, and that "objective" statistical reasoning will neither cloud nor distort the actual assumptions and values behind the judge's decision.

210 Tribe, supra note 127, at 1392.