No Bayesian Solution to the Transposition Fallacy: More Reason to Be Skeptical of Statistical Proof of Discrimination

Kingsley R. Browne

Follow this and additional works at: https://scholarlycommons.law.hofstra.edu/hlelj

Part of the Labor and Employment Law Commons

Recommended Citation
Available at: https://scholarlycommons.law.hofstra.edu/hlelj/vol35/iss2/3

This document is brought to you for free and open access by Scholarly Commons at Hofstra Law. It has been accepted for inclusion in Hofstra Labor & Employment Law Journal by an authorized administrator of Scholarly Commons at Hofstra Law. For more information, please contact lawlas@hofstra.edu.
NO BAYESIAN SOLUTION TO THE TRANSPOSITION FALLACY:
MORE REASON TO BE SKEPTICAL OF
STATISTICAL PROOF OF DISCRIMINATION

Kingsley R. Browne*

Courts evaluating statistical evidence of discrimination routinely engage in a reasoning process that is widely recognized by statisticians to be fallacious.¹ In a typical case, an expert would testify that the deviation between the employer’s workforce and the workforce that would be “expected” if the employer were selecting randomly with respect to the forbidden criterion of interest (race, sex, etc.) is “statistically significant” because the “p-value” associated with the disparity falls beneath some pre-established level, typically five percent.² Courts conclude from that evidence that there is less than a five-percent probability that the disparity occurred by chance.³ Using the process of “null hypothesis significance testing” (“NHST”), the expert would reject the null hypothesis, which is that any deviations from proportional representation are due to chance.⁴ This is enough for many courts to declare that the plaintiff has established a prima facie case of discrimination.⁵

Despite the ubiquity of the above reasoning, it is unambiguously wrong. It is not possible to derive the likelihood that a result is caused by chance from the p-value used in NHST. To equate the p-value with

---

² See Kingsley R. Browne, supra note 1.
³ See Kingsley R. Browne, supra note 1.
⁴ Id.; see also GEOFF CUMMING, supra note 1.
⁵ See, e.g., Adams v. Ameritech Servs., Inc., 231 F.3d 414, 424 (7th Cir. 2000) (noting that once the null hypothesis of random selection was rejected, it was the employer’s responsibility to offer alternative explanations).
the likelihood of random selection is an example of the "transposition fallacy."

Although legal academic commentators often commit the transposition fallacy, and some who may recognize it seem unwilling to challenge it when made by others, still others have acknowledged that it is not actually possible to quantify the likelihood of nonrandom selection (and, a fortiori, of discriminatory selection) from the kinds of statistical evidence that is available in most discrimination cases. These commentators recognize that the probability of discrimination cannot be derived simply from the p-value; it also requires knowledge of the "prior probability" of discrimination—in other words, the "base rate" of discrimination out in the real world. Thus, they have argued for the use of a different statistical tool—Bayesian analysis—which incorporates base rates of discrimination into the probability calculation.

While avoiding the transposition fallacy, use of Bayesian analysis in this context raises additional problems. First, and foremost, it is virtually impossible to estimate with any reasonable accuracy what the appropriate base rate of discrimination actually is. Moreover, even if it were possible, there is a substantial question whether it is appropriate to base an employer's liability explicitly on what other employers have done (or are thought to have done). Such a course seems inconsistent with the spirit, if not the letter, of evidentiary rules that prohibit reliance on character evidence and evidence of "other bad acts."

This article will proceed in three parts. First, I will briefly describe the error of the transposition fallacy, which provides the impetus behind application of Bayesian methods. Second, I will describe how Bayesian methods solve the transposition-fallacy problem. Third, I will discuss why the proposed solution to the transposition fallacy—incorporating estimates of base rates of discrimination—is impractical or

7. Id. at 133-35.
9. See infra Part II.
10. See infra Part III.
11. See infra Part I.
12. See infra Part II A.
impossible. 13 Finally, I will describe why, even if the technical problems could be overcome, use of base rates of discrimination is inappropriate under the rules of evidence. 14

I. THE TRANSPOSITION FALLACY

The transposition fallacy is a simple logical error, but it has had a tremendous impact on courts’ thinking about discrimination cases. 15 It is easiest to understand the fallacy in the concrete context of a hypothetical discrimination case. Assume that an employer of 500 employees is located in an area where 20% of the relevant local labor force is black. The statistically “expected” number of blacks would be 20% of 500, or 100. Thus, the hypothetical universe of nondiscriminating employers of this size would have an average of 100 black employees, with some having more and some having fewer, with deviations from 100 being simply the “luck of the draw.” That universe could be described by a normal (bell-shaped) curve with a mean of 100. The farther one moves away from the mean and out into the tail, the less frequent the deviations from expectation. Put another way, in a normal distribution, small deviations are more likely than large deviations.

The role of statistics in discrimination cases has generally been to test whether a particular deviation from expected is great enough to support an inference that it is not simply random variation but rather the product of a systematic cause, which the plaintiff would argue is a discriminatory one. 16 In our hypothetical, suppose that instead of having 100 black employees, the employer has just 82. Is that so many fewer black employees that we ought to be suspicious of the cause?

The statistical expert would use “null hypothesis significance testing” (“NHST”) to decide whether the disparity is statistically significant. 17 The null hypothesis would be that there is no racial difference in the probability of selection. 18 Under the null hypothesis, five percent of all employers (non-discriminators all) will, by definition, have statistically significant disparities if the standard five-percent significance level is used. 19 If the result is one that would occur in less

13. See infra Part II B.
14. See infra Part III.
15. See generally Browne, Pernicious P-Values, supra note 1.
16. See id. at 114.
17. For a critique of NHST generally, see id. at 155-163.
18. See id. at 114, 118.
19. See id. at 114.
than five percent of nondiscriminating employers chosen at random (that is, if the $p$-value is less than .05), courts typically find the disparity sufficient to establish a prima facie case of discrimination or even to justify a finding of discrimination, based on their conclusion that there is less than a five-percent probability that chance was the cause.20 They do so by making a fundamental mistake about the meaning of the $p$-value.

The $p$-value is calculated under the assumption that the null hypothesis is true.21 That is, by definition, less than five percent of the distribution of nondiscriminating employers will have disparities corresponding to a $p$-value of less than .05.22 That is a far cry, however, from telling you that when you do observe an employer with a disparity associated with a $p$-value of less than .05, there is less than a five-percent probability that the disparity occurred by chance.23 To equate the two probabilities is to commit the transposition fallacy. That is, the transposition fallacy consists of equating the probability of the evidence given the null hypothesis ($p(E|H)$) with the probability of the null hypothesis given the evidence ($p(H|E)$).24 Logic should tell you that because the $p$-value is calculated based on the assumption that the null hypothesis is true, it cannot simultaneously provide a probability that the null hypothesis is true.

A simple coin-flip example may make the point clearer. If one wanted to test whether a coin was fair—that is, whether heads and tails are equally likely to turn up when the coin is tossed—one could use NHST, testing the null hypothesis that there is no difference in the probability of heads or tails. If one were to flip the coin 100 times and obtain 40 heads, the question would be whether the deviation from 50-50 is sufficiently great that we would reject the null hypothesis and conclude that the coin is unbalanced. Using standard statistical techniques, we would see that one would expect a deviation that great or greater from equality slightly less than five-percent of the time. Employing the transposition fallacy, as courts routinely do in discrimination cases, one would conclude that there is less than a five-percent chance that the coin is fair.

But is it true that we can assess the probability that the coin is fair merely from the 60:40 split? A disparity this great is expected in one out of twenty series of coin tosses of fair coins, so by what logic do we now

20. See id. at 115-126.
21. Id. at 121.
22. Id. at 120.
23. Id. at 121.
24. Id. at 114-115.
say that there is only a one in twenty chance that the coin is fair? The fact is that we cannot. If, for example, the coin tosses had been performed with a coin that had already been verified to be fair, we would know that the probability that the coin was fair was 100%, irrespective of the p-value. If conversely, we had previously measured the balance of the coin and determined that it was asymmetrical in a way that would strongly bias coin-flip results, we would know that the probability that the coin was unfair was 100%, again irrespective of the p-value.

To make a judgment about the fairness of the coin, one needs to know not only the results of the coin tosses but also the prevalence of unfair coins. If you knew, for example, that because of quality-control techniques employed in coin manufacturing, virtually all coins are fair, then you should be quite confident when you get that 60:40 split that it was indeed a random event. On the other hand, if you knew that roughly half of all coins were unfair, then if you got such a disparity, you would judge it much more likely that the coin was one of the unfair ones rather than a fair coin that simply showed a chance disparity.

That reasoning shows the importance of base rates in assessing probabilities, and it applies to discrimination cases as well. If discrimination is very rare, then statistically significant disparities are very weak evidence of discrimination; if discrimination is very common, their evidentiary force is greater. Thus, if we knew that there was no discrimination in this world, then all statistically significant disparities would have been caused by chance. If one percent of employers engage in systematic discrimination, then one percent of employers would have statistically significant disparities caused by discrimination, and among the remaining 99 percent of employers, five-percent would have disparities caused by chance, meaning that only approximately one-sixth (16.9%) of all statistically significant disparities were caused by discrimination, and the remainder were caused by chance. Similar probabilities of discrimination can be calculated with base rates of discrimination of five percent (51.3% of disparities would be caused by discrimination), ten percent (69%), and fifty percent (95.2%). Thus, only when the prevalence of systematic discrimination reaches fifty percent is there less than a five-percent probability that a particular

---

25. *Id.* at 122-25. Their force still may not be very strong if there are other nonrandom factors that plausibly might contribute to the result. *Id.* at 125. For purposes of the current discussion, we are assuming that all statistically significant disparities are caused either by chance or by discrimination. *Id.* at 125. In the real world, of course, there will be many such disparities that are caused by neither. See discussion infra pp. 19-20.

employer with a disparity obtained it by chance. So, depending on the 
base rate of discrimination chosen in these hypotheticals, the probability 
that an employer having a deviation significant at the .05 level obtained 
it by chance ranges from 100 percent to 5 percent. Yet, the p-value 
calculated for purposes of testing the null hypothesis is .05 in each case, 
and the strength of the prima facie case in each scenario would be 
deemed equivalent.27 It is this disconnect between the meaning of the p-
value and the meaning that is typically ascribed to it that has led to calls 
for use of Bayesian analysis.28

II. BAYESIAN ANALYSIS: A WAY TO AVOID THE TRANSPOSITION 
FALLACY?

Recognizing the limitations of NHST and the error of equating a p-
value with the likelihood of discrimination, some commentators have 
argued for use of Bayesian analysis in attempting to compute the 
probability that a given employer engaged in discrimination.29 Recently, 
both Deborah Weiss30 and Jason Bent31 have argued that explicitly 
Bayesian methods would, at least in theory, allow a more meaningful 
calculation of the probability of discrimination.32

A. What Is Bayesian Analysis?

Bayesian analysis is a method by which prior beliefs or 
probabilities can be incorporated into the statistical analysis, such as, in 
an earlier example, the prior probability that a coin selected at random 
would be fair.33 If a result is far-fetched, then even a low p-value is not 
strong proof of its validity, while weaker statistical evidence of a highly 
plausible claim might be may be enough to carry the day.34 Consider 
two studies designed to test treatments for upper-respiratory infections, 
each comparing the treatment group to a placebo group. Study 1 tests

27. Id.
28. Id.
29. See generally Jonathan J. Koehler & Daniel N. Shaviro, Veridical Verdicts: Increasing 
Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods, 75 CORNELL L. 
30. See Weiss, supra note 8.
31. See Bent, supra note 8.
32. See Bent, supra note 8.
33. Kaye & Friedman, supra note 1, at 258-59.
34. Steven N. Goodman, Of P-Values and Bayes: A Modest Proposal, 12 EPIDEMIOLOGY 
the effect of a new antibiotic that is derived from, and chemically similar
to, an antibiotic already known to be effective. Study 2 tests the effect
of intercessory prayer. Both studies produce equivalent results that are
statistically significant at the five-percent level. Should we treat these
two studies as equally meaningful? What separates the two is that the
result in Study 1 is quite plausible, while the result in Study 2 is far less
so (even most religious people go to doctors). Thus, it is not the $p$-value
that gives us confidence in the result but the plausibility of the
alternative hypothesis, which turns on Bayesian “priors.”

The flaw in the transposition fallacy, it will be remembered, is that
it wrongly equates the probability of the evidence given the hypothesis
$(p(E|H))$ with the probability of the hypothesis given the evidence
$(p(H|E))$, probabilities that have no clear relationship to each other. In
contrast, Bayesian methods allow direct calculation of the probability of
the hypothesis given the evidence, which is the probability that one
would want. The formula for calculating $p(H|E)$ is:

$$p(H|E) = \frac{p(H)p(E|H)}{p(E)}$$

where $p(E) = p(H)p(E|H) + p(-H)p(E|-H)$.

Applying a Bayesian analysis to a hypothetical case in which the
base rate of discrimination is 10% and the $p$-value is .05, the relevant
terms of the formula are:

- $p(H)$ = the “prior probability”; that is, the probability of the
  hypothesis that the employer in question discriminated before we
  know whether the employer has a significant disparity. In
  this example, the prior probability is provided exclusively by the

35. See William S. Harris et al., A Randomized, Controlled Trial of the Effects of Remote,
Intercessory Prayer on Outcomes in Patients Admitted to the Coronary Care Unit, 159 ARCH.
INTERNAL MED. 2273 (1999) (reporting a statistically significant ($p = .04$) advantage in coronary-
care-unit course scores in patients who had been prayed for). But see Donald A. Sandweiss, P Value
Out of Control, 160 ARCH. INTERNAL MED. 1872 (2000) (arguing that the above results “suggest the
need to reassess our statistical methods for judging efficacy rather than the need to reassess the
fundamental theories of science”); See also Eric-Jan Wagenmakers et al., Why Psychologists Must
Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011), 100 J. PERS.
& SOC. PSYCHOL. 426 (2011) (commenting on an article purporting to find support for precognition,
stating that “[i]nstead of revising our beliefs regarding psi, Bem’s research should instead cause us
to revise our beliefs on methodology”).

36. See Koehler & Shaviro, supra note 29, at 255 n.27.
37. Id. at 256.
38. Id. at 255.
39. Id. at 255 n.27.
base rate of discrimination = .10

- $p(-H) =$ the prior probability that the employer did not discriminate = .90

- $p(E|H) =$ the probability of a significant disparity if the employer discriminates = .1

- $p(E|-H) =$ the probability of a significant disparity if the employer does not discriminate (the significance level) = .05

- $p(E) =$ the probability that any random employer will have a significant disparity = (1)(.10) + (.90)(.05) = .1 + .045 = .145

Therefore, $p(H|E)$, which is the probability that the employer discriminated given the observed disparity, equals

$$\frac{(1)(1)}{.145} = .69$$

As can be seen, this is the same probability calculated using the more intuitive approach described previously. Where the base rate is 0, of course, the prior probability, denoted by $p(H)$, is 0, so the numerator is 0; hence, the probability of discrimination, represented by $p(H|E)$, would likewise be 0.

So, if Bayesian analysis gives an appropriate probability untainted by the transposition fallacy, what's not to like? Well, the devil, as always, is in the details, and the details present three perhaps insuperable problems. The first is how to quantify the “prior probability,” that is, the base rate of discrimination, which is necessary to compute the “posterior probability,” which is the probability of interest. The second is the assumption that $p(E|-H)$ is equal to the significance level; put another

40. Id.

41. For purposes of explication, it is assumed that all discriminating employers would have statistically significant disparities, which would not necessarily be the case, as the study my simply not have sufficient power to detect it. Another way of saying this is that nonsignificant results do not prove the null hypothesis. See CUMMING, supra note 1, at 29 (“We must therefore be careful not to take statistical nonsignificance (the null is not rejected) as evidence of a zero effect (the null is true).”).

42. For purposes of explication, it is assumed that all significant disparities are caused by chance or discrimination, which is also not necessarily the case, as there are numerous nondiscriminatory, but nonrandom, causes that could cause such results. See PART II(C), infra.

43. See text accompanying note 7, supra.
way, the assumption that chance is the only nondiscriminatory explanation for statistically significant disparities. The third is that use of background rates of discrimination seems inconsistent with the rules of evidence, which typically require evidence specific to a party in litigation and not evidence about the tendency of other similarly situated actors to engage in the conduct that the party is accused of engaging in.

B. Where Do the Prior Probabilities Come From?

The inability to specify the prior probability with any accuracy at all has long been thought to be a perhaps insurmountable barrier to use of Bayesian analysis in many legal contexts. Both Professors Weiss and Bent disagree. They argue, not without force, that prior probabilities ("priors") necessarily play a substantial role in any discrimination case. That is, the trier-of-fact, in deciding whether an observed statistical disparity (or even an isolated event such as an individual discharge) is a consequence of discrimination, must either use priors presented in the courtroom or, instead, rely on its own pre-existing priors, or, as Professor Bent calls them, "hidden priors." That is what Professor Weiss means by her claim that discrimination law cannot be "agnostic." One way or another, the argument goes, someone's priors

44. See Bent, P-Values, Priors, and Procedures, supra note 8, at 96. See also David H. Kaye, The Numbers Game: Statistical Inference in Discrimination Cases, 80 Mich. L. Rev. 833, 854 n.67, 854-55 (1982) (describing the "difficulty of institutionalizing Bayesian inference in the law," even in relatively straightforward cases like Hazelwood and concluding that "it would not provide that much additional guidance to the fact-finder"); Richard Lempert, Statistics in the Courtroom: Building on Rubinfeld, 85 Colum. L. Rev. 1098, 1101 (1985) (stating that "the Bayesian perspective is unlikely to transform the way statistical data is evaluated for litigation purposes at any time in the foreseeable future").

45. See Bent, Hidden Priors, supra note 8, at 840-849. Professor Bent seems to use the term "hidden priors" in two different ways. One way is the suggestion that jurors rely on their "hidden priors"—that is, their view of the prevalence of discrimination—in reaching their verdicts after evaluating all of the evidence. The other sense in which he uses the term is more specific. He says that "Decision rules turning on statistically significant results from traditional hypothesis testing necessarily involve a built-in, unstated, and unexamined assumption about the prevalence, or base rate, of discrimination— a hidden prior." Here he seems to be referring to the "prior" as being built in to traditional hypothesis testing using p-values. But there really is no assumption about prevalence built in. As explained earlier, the p-value is computed assuming the null hypothesis is true, so there is no base-rate information (other than the "zero" base rate assumed by the null hypothesis). It is true, as noted above, that the probability of discrimination is roughly equal to the five-percent significance level if the base rate of discrimination happens to be .5, but that's just where the lines happen to cross; it really does not rest on anyone's "assumptions"—either explicit or implicit—about the base rate.

46. Weiss, supra note 8, at 1678 (stating that "whenever the factual occurrence of differential treatment is at issue, triers of fact must make background assumptions about the societal pattern of discrimination").
are going to factor into a jury’s verdict. Thus, rather than relying on the intuitive priors of jurors, they argue, the priors should be supplied by social science.47

The need for the trier-of-fact to rely on a set of priors is not unique to discrimination cases.48 Indeed, as Judge Frank Easterbrook has noted, “[a]ll inferences are statistical—whether implicitly or explicitly does not matter.”49 Whenever facts are contested, the trier-of-fact is called upon to resolve the conflict and, in doing so, will make judgments of the plausibility of the differing accounts.50 Those judgments are likely to be strongly influenced by the trier-of-fact’s intuition about how the world works and how people in various settings are likely to behave. Those are priors every bit as much as assumptions about the statistical frequency of discrimination in society, and evidence relating to the accuracy of those assumptions might well be relevant to their decision-making process. Usually, however, we leave it to the common sense of the trier-of-fact to sort out. Take, for example, a lawsuit involving allegations of excessive force by police officers. Different jurors are likely to have different levels of receptivity to such claims depending upon whether they think there is a lot of police brutality or very little. Those in the latter camp are probably going to require more evidence than those in the former before concluding that excessive force was used in a particular instance. Logically, therefore, the rate of police brutality is relevant to the jurors’ decisions.

Attempting to specify priors in discrimination cases (or, for that matter, police brutality cases) raises a host of practical problems. What is the base rate of discrimination, after all? Specifically, what is the base rate of systematic discrimination?51 Professor Bent notes that “[t]he background rate of discrimination on the basis of race by all employers in the United States may be practically unknowable, given our current constraints on data sources and polling abilities, but theoretically it could

47. Id. at 1679; Bent, supra note 8 at 849.
48. Bent, supra note 8, at 848 (noting that “the use of Bayesian statistics is not unheard-of in litigation, especially in paternity cases and DNA match criminal cases”).
49. Baylie v. Federal Reserve Bank of Chicago, 476 F.3d 522, 523 (7th Cir. 2007).
50. Bent, supra note 8, at 835.
51. It is likely that any employer of substantial size has either engaged in some discriminatory act in its history or, at least, could be presented with prima facie cases of discrimination that it would be unable to rebut. See McDonnell Douglas v. Green, 411 U.S. 792 (1973). However, in a pattern or practice case, the question is not whether an employer has discriminated at all, but whether “discrimination was the company’s standard operating procedure the regular rather than the unusual practice.” International Brotherhood of Teamsters v. United States, 431 U.S. 324, 336 (1977). Therefore, presumably it is the rate at which employers engage in discrimination as their standard operating procedure that would be relevant.
be an empirically measured fact."52 But one reason that social science does not tell us the background rate of discrimination is very obvious: there is not a single base rate of discrimination. There are no doubt different base rates for racial discrimination depending upon what race is at issue, what industry is at issue, what region of the country, or even what city, is at issue.53 Moreover, the rates of discrimination are likely to be very different in a single locale for a particular racial group depending upon the nature of the job from which the people have allegedly been discriminatorily excluded.54 All of these rates are likely to be different for sex discrimination than they are for race discrimination—and different for discrimination against men and discrimination against women—different for religious discrimination (and different by religion) than they are for age discrimination, and so forth.55 Yet the number actually chosen from the array of possible—but practically indeterminate—discrimination rates is likely to be outcome-determinative in the litigation.56 As Professor Bent states, "estimates of priors can make the difference between deciding that a defendant more likely than not engaged in systemic disparate treatment and deciding exactly the opposite, even where the parties offer the exact same statistical evidence."57 As a result, he acknowledges that the priors "will necessarily be at least somewhat subjective and possibly deeply uncertain . . ."58 All the more reason, one might have thought, to avoid assessing liability on the basis of hypothetical assumptions.

So, where would these theoretically knowable priors come from? Professor Bent suggests that there are a number of sources, including empirical studies of the outcomes of employment discrimination litigation,59 "social framework" evidence to "inform or adjust a prior

52. Bent, P-Values, Priors, and Procedure, supra note 8, at 117.
54. See Weiss, The Impossibility of Agnostic Discrimination Law, supra note 8, at 1712 (recognizing that the likelihood of discrimination occurring can vary based on the nature of the occupation and the specific workplace location).
55. Id.
56. Id.
57. Bent, Hidden Priors, supra note 8, at 837.
58. Id.
59. Bent, P-Values, Priors, and Procedure, supra note 8, at 131-132. Bent acknowledges, however, that "empirical studies are rare, deeply flawed, or non-existent." Even if they were not, however, they would be of limited utility. Litigated cases do not represent a random sample of employment decisions. Of the millions of employment decisions that are made each year, only an infinitesimal fraction of them lead to litigation, and those that do probably tend to involve circumstances in which at least a somewhat plausible argument of discrimination could be asserted.
probability,\textsuperscript{60} and academic theory,\textsuperscript{61} although how “academic theory” would establish an empirical fact is not obvious—indeed, the prospect is somewhat frightening.\textsuperscript{62} The alternative, he says, is to allow triers-of-fact to rely on prior probabilities that are just “unsupported guesses, estimates, or assertions about the state of the world, based on idiosyncratic beliefs held by judges or fact-finders, and not based on any evidence in the record.”\textsuperscript{63} But in the absence of a recognized, authoritative source for the priors to which the expert would testify, it is unclear that the expert’s testimony would be an improvement. Having two competing experts argue about what they think the base rate of discrimination is (based in large part on speculation, albeit informed speculation) and having the jurors base their decisions on the relative persuasiveness, showmanship, or attractiveness of the experts does not seem to be an improvement over basing their decisions on their own common sense and understanding of the way the world works.

Even if social scientists could generate estimates of the prevalence of discrimination using some sociological definition of discrimination, it is unlikely that they could come up with a reasonable estimate of the prevalence of illegal discrimination. That is, any “study” of the prevalence of discrimination nationwide would almost certainly be based on the same kinds of statistical patterns of disproportionate representation that the plaintiffs would be challenging in the case in which the evidence is to be admitted. But lack of proportional representation is not illegal.\textsuperscript{64} For an employer to be liable in a “pattern or practice” case, the lack of proportional representation must be a consequence of intentional invidious discrimination.\textsuperscript{65} Again, as the Teamsters Court said, the plaintiff’s burden is to show that invidious

\textsuperscript{60} Id. at 118, 132-134.
\textsuperscript{61} Id. at 134-136.
\textsuperscript{62} Rule 702 of the Federal Rules of Evidence requires that the scientific knowledge made admissible by the rule be testable and have been tested. Daubert v. Merrell Dow Pharm, Inc., 509 U.S. 579, 580, 593 (1993). It is unclear how the academic theory that might support an expert’s opinion about the prevalence of discrimination could be tested.
\textsuperscript{63} Bent, \textit{P-Values, Priors and Procedure}, supra note 8, at 136; Professor Weiss seems to view the Bayesian priors as more subjective and qualitative, allowing social-framework evidence to be introduced to substitute for (or at least augment) the jurors’ own subjective priors. See Weiss, supra note 8 at 1696 n.65 (stating that although her numerical examples might seem open to the objection that “none of us have quantified views on the probabilities in question,” Bayesian analysis “also holds for qualitative judgments about relative likelihood although the formal representation of these relationships is more complex”). See also id. at 1719 (describing ways that either courts or experts could convey information concerning background assumptions to the jury). It is far from clear how this would work—or even whether it could—in the real world.
\textsuperscript{64} See Weiss, supra note 8 at 1689.
discrimination is the company’s “standard operating procedure—the regular rather than the unusual practice.” 66

Given these circumstances, it is not clear that expert testimony, which would be the source of the base-rate data, should even be admissible. The rules of evidence require that expert testimony be helpful to the jury to determine the facts at issue, based on sufficient facts or data, and be the product of reliable principles and methods. 67

For the sociological studies to be helpful to the jury in assessing the likelihood that the employer engaged in unlawful employment discrimination, they would at a minimum have to speak to the prevalence of the specific kind of unlawful discrimination at issue in the case. 68

Even if experts purported to opine on the prevalence of unlawful discrimination, one might question whether agenda-driven expert testimony is likely to give a jury a more accurate picture of the base rate of discrimination than the jurors are likely to possess merely from being members of the community. That is to say, there seems little reason to believe that jurors’ priors are likely less accurate than those of social scientists—many of whom are heavily invested in careers that reward “finding” high levels of discrimination 69—and there is little chance that academic sociologists would bring the diversity of perspectives on the question that a randomly selected jury pool would.

There is an additional problem with the Bayesian approach advocated by Professor Bent. In discussing application of the Bayesian approach, he explicitly assumes that statistical disparities are caused either by chance or by discrimination. 70 For illustration of the operation

66. Id.
67. Rule 702 of the Federal Rules of Evidence provides:
   A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
   (a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
   (b) the testimony is based on sufficient facts or data;
   (c) the testimony is the product of reliable principles and methods; and
   (d) the expert has reliably applied the principles and methods to the facts of the case.
Fed. R. Evid. 702.
68. It also may be that the relevant statistic is not what percentage of employers discriminate but rather what percentage of individual employment decisions are tainted by discrimination, which even for an employer found to engage in systematic discrimination may be relatively small.
69. Weiss, supra note 8, at 1684, discusses the work of William Bielby, who is, in her words, “both a leading scholar on discrimination and a leading expert for plaintiffs.” Bielby has made a comfortable living testifying for plaintiffs in discrimination cases over the last three decades, and he relies to a large extent on his “scholarly” work to inform his testimony.
70. See Bent, P-Values, Priors, and Procedure, supra note 8, at 108.
of a Bayesian analysis, that is a reasonable thing to do, and I did it when I was discussing the meaning of the $p$-value.\footnote{See supra Part I.} However, it assumes that the only prior probability of interest is the prior probability of discrimination.\footnote{See Bent, \textit{P-Values, Priors, and Procedure}, supra note 8, at 108-09; See also Bent, \textit{Hidden Priors}, supra note 8, at 821-23.} Another critical probability in the real world is the prior probability that nondiscriminatory \textit{nonrandom} factors are also operating to create a deviation from strict proportional representation. As discussed below, that is a "prior" just as important as the base rate of discrimination.

\textit{A. The Bayesian "Solution," Like NHST, Tends to Focus Excessively on the False Dichotomy of Chance Versus Discrimination}

Expert testimony itself is not free of its own "hidden priors."\footnote{Weiss, \textit{supra} note 8, 1710, accurately observes:
Experts are no more exempt than the rest of us from the logic of Bayes' Law. They choose their experimental designs and evaluate the experimental results of others in the light of their own background assumptions, and empirical research on discrimination is no doubt influenced by prior belief with political overtones.}

When sociologists examine statistical patterns and derive from them estimates of the prevalence of discrimination, they are likely to be based on one of the greatest "hidden priors" of all, which is an assumption about the extent to which nonrandom statistical effects are necessarily due to discrimination or rather to relevant group differences—whether quantified or quantifiable, observed or unobserved—in factors such as occupational interest, qualifications, and the like.\footnote{Application of Bayesian inference in this context thus violates the requirement that the set of hypotheses must be "mutually exclusive and collectively exhaustive." David Trafimow, \textit{Hypothesis Testing and Theory Evaluation at the Boundaries: Surprising Insights from Bayes's Theorem}, 110 PSYCHOL. REV. 526, 526, 534-35 (2003). Failure to account adequately for systematic nondiscriminatory causes of statistical disparities is a problem for all statistical analyses used in discrimination cases. See Kingsley R. Browne, \textit{Statistical Proof of Discrimination: Beyond "Damned Lies"}, 68 WASH. L. REV. 477, 503-513 (1993) (criticizing the "central assumption" of statistical proof set forth in \textit{Teamsters}, 431 U.S., at 339-40 n.20, that "absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired").} That is a major source of contention in the literature.\footnote{See generally Kingsley R. Browne, \textit{The Quixotic Quest for "Gender Equality"}, U. TOL. L. REV. (forthcoming) KINGSLEY R. BROWNE, \textit{BIOLOGY AT WORK: RETHINKING SEXUAL EQUALITY} (2002); Kingsley R. Browne, \textit{Evolved Sex Differences and Occupational Segregation}, 27 J. ORG. BEHAV. 143 (2006).} Thus, if the plaintiff is permitted to introduce evidence attempting to demonstrate the background rate of discrimination, it would be equally open to the defendant to introduce...
general evidence of group differences in interest and qualifications. This would include not just employer-specific evidence, which has been successfully used by employers to explain statistical disparities in their own workforces but also labor-force-wide evidence that would shed light on the meaning of the plaintiffs’ expert testimony concerning estimates of labor-force-wide discrimination, leaving jurors to decide whose social science is right.

In suggesting the means by which Bayesian analysis can solve the problems with statistical evidence, the focus on chance-versus-discrimination obscures the importance of demonstrating that the employer has actually discriminated. In order to narrow the focus down to these two potential causes, one has to assume that there are no unaccounted-for differences between the relevant groups in interest in the jobs or in ability or qualifications, such as education, training, and experience. This assumption would be far-fetched. One of the most robust findings of the organizational psychology literature, for example, is that men and women exhibit large differences in occupational interest, and despite dramatic changes over the last half-century in women's participation in the workforce, these differences have persisted. Similarly, the disparate-impact literature reveals that there are large differences between the races in the predictors of productivity. Multiple regression analysis can help in this respect, but the data are invariably incomplete and often use crude proxies for the traits of interest, which, almost by definition, means that the null hypothesis—that members of the groups at issue have the same probability of being included in the employer's workforce if the employer is not

76. See EEOC v. Sears, Roebuck & Co., 839 F.2d 302, 309 (7th Cir. 1988).
77. Id. at 308.
78. Id.
82. See Amy Wax, Disparate Impact Realism, 53 WM. & MARY L. REV. 621, 623 (2011) (arguing that the quest for valid predictors of job performance that do not cause a disparate impact by race is chimerical because of the "validity-diversity tradeoff": "the most effective job selection criteria consistently generate the smallest number of minority hires").
83. See, e.g., EEOC v. Morgan Stanley, 324 F. Supp. 2d 451, 458 (S.D.N.Y. 2004) (rejecting challenge to regression analysis that used experience and education as proxies for performance and productivity, stating that "[d]isputes regarding the proper variables to employ in statistical studies are more properly left for juries to consider and decide").
discriminating—is not true, and perhaps not even remotely so.84 Thus, there is no good reason to assume that once sampling error has been rejected as the cause of a statistical disparity, discrimination is the likely explanation. This problem, which has already been shown to be a fundamental problem with NHST itself,85 is not any closer to being solved through Bayesian analysis.

II. USE OF BAYESIAN PRIORS IS INCONSISTENT WITH THE RULES OF EVIDENCE

Even apart from its unreliability, the kind of evidence that Bayesian enthusiasts advocate is, in a fundamental way, inconsistent with principles of evidence law. It is unlikely that a social scientist ever would have adequate data to back up a statement such as "[b]etween thirty and forty percent of employers in the defendant's industry engage in systematic illegal sex discrimination" or "of the top one hundred firms in the defendant's industry, thirteen of them engage in systematic illegal race discrimination against blacks." But suppose the plaintiff could in fact find an expert willing to do so—and the availability of experts willing to provide fanciful opinions should not be underestimated.86 Suppose further that the testimony is found to satisfy the reliability dictates of Rule 702. What would be the appropriate response to such evidence introduced to establish Bayesian priors?

The appropriate response under the evidence rules would be to exclude the evidence, even if it were deemed a reliable indicator of the base rate of discrimination. Fundamentally, such evidence runs afoul of Rule 404 of the Federal Rules of Evidence, which limits the kind of evidence that can be used to support an inference that a person acted in a particular way on a particular occasion. Rule 404(a) of the Federal Rules of Evidence bars admission of "[e]vidence of a person's character or character trait . . . to prove that on a particular occasion the person acted in accordance with the character or trait."

Suppose the plaintiff could produce evidence of "[b]etween thirty and forty percent of employers in the defendant's industry engage in systematic illegal sex discrimination" or "of the top one hundred firms in the defendant's industry, thirteen of them engage in systematic illegal race discrimination against blacks." But suppose the plaintiff could in fact find an expert willing to do so—and the availability of experts willing to provide fanciful opinions should not be underestimated. Suppose further that the testimony is found to satisfy the reliability dictates of Rule 702. What would be the appropriate response to such evidence introduced to establish Bayesian priors?

The appropriate response under the evidence rules would be to exclude the evidence, even if it were deemed a reliable indicator of the base rate of discrimination. Fundamentally, such evidence runs afoul of Rule 404 of the Federal Rules of Evidence, which limits the kind of evidence that can be used to support an inference that a person acted in a particular way on a particular occasion. Rule 404(a) of the Federal Rules of Evidence bars admission of "[e]vidence of a person's character or character trait . . . to prove that on a particular occasion the person acted in accordance with the character or trait."

84. See Browne, Pernicious P-Values, supra note 1, at 156-58.
85. Id. at 161-63.
86. See Jack B. Weinstein, Improving Expert Testimony, 20 U. RICH. L. REV. 473, 482 (1986) (noting that "[a]n expert can be found to testify to the truth of almost any factual theory, no matter how frivolous . . .").
87. FED. R. EVID. 404(a); Edward J. Imwinkelreid, Reshaping the "Grotesque" Doctrine of Character Evidence: The Reform Implications of the Most Recent Psychological Research, 36 SW. U. L. REV. 741, 743 (2007-2008) (noting that forty-one states have adopted rules of evidence patterned after the Federal Rules and that the remaining states also generally ban introduction of character evidence). FED. R. EVID. 413-415 provide exceptions to the general rule for cases involving sexual assault and child molestation, but those are not relevant here.
some narrow exceptions generally favoring defendants in criminal cases, the rule absolutely forbids the use of character evidence to prove conduct in a civil action.\footnote{88} Moreover, Rule 404(b)(1) provides that prior acts are "not admissible to prove a person's character in order to show that on a particular occasion the person acted in accordance with the character."\footnote{89} Rule 404(b)(2) provides that prior acts may be admissible for other purposes, however, including "motive, opportunity, intent, preparation, plan, knowledge, identity, absence of mistake, or lack of accident," but for reasons discussed below, this subsection would be of no help to the expert in a discrimination case.

The central function of Rule 404 is to bar evidence that is intended to demonstrate that an actor has a "propensity" to act in a particular way in order to prove that he acted that way at the time in question.\footnote{90} A number of rationales for the rules barring such evidence have been offered. First, in many cases, it may be of only slight probative value yet very prejudicial, because it may cause the trier of fact to find against a defendant not because of "what he did" in the particular case but of "who he is."\footnote{91} Although the prejudicial "you're guilty because you're a bad person" inference may be less likely in a case involving misbehavior of third persons,\footnote{92} that is simply because the evidence of what others do

\footnote{88} FED. R. EVID. 404, advisory committee notes to 2006 amendment ("The Rule has been amended to clarify that in a civil case evidence of a person's character is never admissible to prove that the person acted in conformity with the character trait.").

\footnote{89} But see FED. R. EVID. 413 and FED. R. EVID. 414, (allowing in criminal cases, evidence of prior acts of sexual assault and child molestation by the defendant, respectively, to "be considered on any matter to which it is relevant.") See also FED. R. EVID. 415(a) (applicable in civil cases).


\footnote{91} See United States v. Caldwell, 760 F.3d 267, 276 (3rd Cir. 2014) (emphasis in original). See also Michelson v. United States, 335 U.S. 469, 475-76 (1948): The inquiry is not rejected because character is irrelevant; on the contrary, it is said to weigh too much with the jury and to so overpersuade them as to prejudice one with a bad general record and deny him a fair opportunity to defend against a particular charge. The overriding policy of excluding such evidence, despite its admitted probative value, is the practical experience that its disallowance tends to prevent confusion of issues, unfair surprise and undue prejudice.

\footnote{92} Cf. United States v. McKinney, 156 F.3d 1233 (6th Cir. 1998) (rejecting defendant's
is only marginally relevant to what the defendant did. But if the jury does not link the base rate to the defendant’s inclination to discriminate, it is not clear what the evidence is there for at all. As the Supreme Court noted in *Huddleston v. United States*, under Rule 404(b), “similar act evidence is relevant only if the jury can reasonably conclude that the act occurred and that the defendant was the actor.”

Evidence of prior bad acts also tends to distract the fact-finder from the question of what happened on the occasion in question, compared to what has happened on other occasions (in this case with others not party to the litigation). Thus, much of the rationale for Rule 404 rests on the same concerns animating Rule 403, which allows the trial judge to exclude relevant evidence “if its probative value is substantially outweighed by a danger of... unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.” In the context of a discrimination claim against a single employer, evidence about base rates of discrimination changes the focus of the trial from what Acme Employer did during the legally relevant time period to what employers across the country have done over a broader period of time and to the myriad innocent contributors to disparities in workplace outcomes by race, sex, age, religious, and other categories, in workforces generally.

Evidence of Bayesian priors does not satisfy the requirements of Rule 404(b) for introduction of evidence of other acts for non-propensity-related purposes. The Supreme Court in *Huddleston v. United States* established a four-pronged inquiry for other-acts evidence

---

404(b) objection to introduction of bad acts evidence relating to third person, where there was no risk that the jury would attribute those bad acts to the defendant).


94. Id. at 688.

95. See also United States v. Cardall, 885 F.2d 656, 671 (10th Cir. 1989) (“We do not believe that the relevance requirement under Rules 404(b), 402 and 104(b) can be met with respect to one defendant by introducing evidence of the bad acts of the defendant’s other associates.”).

96. See United States v. Alayeto, 628 F.3d 917, 922 (7th Cir. 2010) (upholding exclusion of “other acts” evidence on the ground that it would distract jurors from the central issue in case).

97. See Edward W. Cleary, McCormick on Evidence § 188 (3d ed. 1984) (“Character evidence used for this purpose, while typically being of relatively slight value, usually is laden with the dangerous baggage of prejudice, distraction, time consumption [and surprise.]”). Accordingly, for example, courts sometimes view the admissibility of gang membership in criminal cases as a matter to be decided just on relevance and prejudice grounds under Rules 401 and 403, see United States v. Ford, 761 F.3d 641, 648 (6th Cir. 2014); United States v. Irvin, 87 F.3d 860, 863 (7th Cir. 1996); United States v. Ozuna, 674 F.3d 677, 682 (7th Cir. 2012); United States v. Skillman, 922 F.2d 1370, 1373 (9th Cir. 1990), cert. dismissed, 502 U.S. 922 (1991); and sometimes as a Rule 404 issue, see United States v. Jobson, 102 F.3d 214, 220 (6th Cir. 1996); United States v. Ellison, 616 F.3d 829, 833 (8th Cir. 2010); People v. Trujillo, 338 P.3d 1039, 1045 (Colo. App. 2014).

98. See Fed. R. Evid. 404.
under that rule. First, the evidence must be offered for a purpose that is proper under Rule 404(b)—that is, for a purpose other than propensity. Second, under Rules 401 and 402, the evidence must be relevant to proving that purpose. Third, the trial court must determine under Rule 403 whether the probative value of the evidence is substantially outweighed by its potential for unfair prejudice. Finally, under Rule 105, the trial court must, if requested, instruct the jury that the evidence may be considered only for the proper purpose for which it is admitted and not for the forbidden propensity inference.

Before considering the admissibility of evidence of background rates of discrimination, it should be noted that even evidence that the defendant employer itself previously engaged in discrimination is not necessarily admissible against it in a discrimination action. In *Sprint/United Management Co. v. Mendelsohn*, the District Court had excluded the age-discrimination plaintiff’s proffered evidence of age discrimination against other employees on the basis of objections raising relevance and prejudice issues under Rules 401 and 403. The Tenth Circuit reversed, interpreting the District Court’s action as applying a *per se* rule excluding such evidence. The Tenth Circuit then reweighed the relevance and prejudice and ruled the challenged testimony admissible. The Supreme Court unanimously reversed, expressing doubt that the District Court had actually applied a *per se* rule and remanding with instructions to clarify the basis for its ruling. Rules 401 and 403, according to the court “do not make such evidence *per se* admissible or *per se* inadmissible.”

100. Id. at 691.
101. Id.
102. Id.
103. Id. at 691-92.
105. Id. at 383.
107. Id. at 381. The Court did not mention Rule 404(b), but logically the evidence was 404(b) evidence, as recognized in a number of briefs on both sides in the case. See Brief for Employers Group as Amicus Curiae in Support of Petitioner, Sprint/United Management Co. v. Mendelsohn, No. 06-1221; id., Brief Amicus Curiae of AARP in Support of Respondent; id., Brief of Amici Curiae Lawyer’s Committee for Civil Rights Under Law et al.; Reply Brief of Appellant, 10th Cir. No. 05-3150. The decisions of the District Court and the Tenth Circuit also did not address Rule 404, but rather, like the Supreme Court, analyzed the case under Rules 401 and 403. The Third Circuit has explicitly relied on Rule 404(b) to exclude evidence that the employer had engaged in discrimination against a different employee. Becker v. ARCO Chemical Co., 207 F.3d 176 (3d Cir. 2000) (overturning District Court’s refusal to order new trial based upon admission of evidence suggesting that corporate defendant had discriminated against another employee). The *Becker* court quoted an earlier Third Circuit case for the “self-evident proposition” that “a proponent’s
Evidence of Bayesian "priors" presents just the kind of risks that Rule 404 was designed to mitigate. The purpose of the evidence is to provide the trier of fact with information about rates of discrimination by employers generally to allow it to assess the probability that the defendant employer discriminated. Jurors with such evidence would be told, in essence: "In finding against this defendant, you may rely on the testimony that you have heard about unlawful acts committed by other employers as an indication of what this employer did." In that sense, of course, the evidence goes beyond ordinary character evidence, which involves evidence of a particular person's character to support an inference that he acted in a particular way.

incantation of the proper uses of [Rule 404(b) evidence]... does not magically transform inadmissible evidence into admissible evidence." 207 F.3d at 191 (quoting United States v. Morley, 199 F.3d 129, 133 (3rd Cir. 1999). Instead, "the proponent must clearly articulate how that evidence fits into a chain of logical inferences, no link of which may be the inference that the defendant has the propensity to commit the crime charged." Id. The court ruled that the plaintiff's evidence of another discriminatory act by the defendant "fails this test because the logical connection between ARCO's alleged 'fabrication' of performance problems in relation to [the other employee's] dismissal and its purported conduct in terminating [plaintiff] is the inference that ARCO was likely to have fabricated customer complaints and other performance problem's in [plaintiff's] case merely because ARCO previously engaged in a similar impropriety in facilitating [the other employee's] dismissal." 207 F.3d at 191. The court stated: "this is the very evil that Rule 404(b) seeks to prevent." Id. (quoting Morley, supra, 199 F.3d at 134). It should also be noted that under Huddleston even if the evidence is supported by such a "proper" purpose, it still may be excluded under Rule 403. See generally Lisa Marshall, The Character of Discrimination Law: The Incompatibility of Rule 404 and Employment Discrimination Suits, 114 Yale L.J. 1063 (2005) (noting that much evidence admitted in employment discrimination cases violates 404(b) but arguing that the rule should be modified).

108. See Laurens Walker & John Monahan, Social Frameworks: A Now Use of Social Science in Law, 73 Va. L. Rev. 559, 581 (1987) (noting that "[t]he policy concern that gave rise to a rule barring the admissibility of evidence of an individual's 'characteristic' behavior applies with equal force to the use of information on behavior characteristic of the groups to which he or she belongs"); Allan G. King & Syeeda S. Amin, Social Framework Analysis as Inadmissible "Character" Evidence, 32 Law & Psychol. Rev. 1 (2008). See also Allan G. King & Syeeda S. Amin, The Propensity to Stereotype as Inadmissible "Character" Evidence, 27 ABA J. Lab. & Emp. Law 23 (suggesting that implicit-bias evidence should be barred as impermissible character evidence).

109. See Marshall, supra note 107 at 1802.

110. Id. at 1076.

In short, then, when plaintiffs purport to offer evidence of an employer's "motive," they overwhelmingly do so based on the following logic: The employer's prior acts reveal that the employer has some discriminatory mindset; ipso facto, the employer was motivated to discriminate. Nothing more than semantics differentiates this "motive" from character propensity, while the underlying theory of admissibility in no manner complies with Rule 404(b)'s prohibition of prior act evidence "to prove the character of a person in order to show action in conformity therewith."

111. Id. at 1073-74.
The Bayesian priors are evidence of the propensity of other organizations—that is, employers in general or perhaps some crudely specified subset of employers—introduced to prove that the defendant-employer acted in conformity with its character and indulged the same propensity to discriminate.\textsuperscript{112}

If this were not statistical evidence—that is, evidence that is reduced to a number for a statistical study—it seems unlikely that many would suppose for an instant that it was admissible. It is hard to imagine, for example, an expert's being permitted to testify: “I have no specific evidence that the defendant engaged in discrimination, but let me tell you about a lot of other employers that did.” The fact that a social scientist might aggregate instances of discrimination and characterize the frequency of employment discrimination as a scientific-sounding “Bayesian prior probability” to insert into a statistical formula does not change its essential nature.\textsuperscript{113} Although evidence of “other acts” under Rule 404(b) usually relates to discrete actions, such as having previously committed a particular crime, the Bayesian argument is simply an aggregation of such acts by some estimated percentage of all potential actors.\textsuperscript{114} Thus, the trier of fact is being asked to draw an inference that the defendant-employer discriminated by virtue of being an employer or, if the base-rate evidence is more narrowly focused, by virtue of being a particular kind of employer—which is really a kind of “profiling” evidence.\textsuperscript{115}

The similarity of profiling evidence and character evidence (really

\textsuperscript{112} See supra Sections II.A., II.B.

\textsuperscript{113} See Michael D. Claus, Profiles, Syndromes, and the Rule 405 Problem: Addressing a Form of Disguised Character under the Federal Rules of Evidence, 88 NO\_TRE DAME L. REV. 973, 977 (2012) (“Disguised character evidence should not be admissible simply because it is presented by an expert.”). There is a large literature, beyond the scope of this article, on the question whether and in what circumstances it is appropriate to treat base rates as relevant. See, e.g., Jonathan J. Koehler, When Do Courts Think Base Rate Statistics Are Relevant?, 42 JURIMETRICS J. 373 passim (2002); Koehler & Shaviro, supra note 26, at 257-61.


\textsuperscript{115} As stated by Brilmayer & Kornhauser:

If a plaintiff tries to prove discrimination in a property owner’s refusal to rent him a house, he should not be allowed to introduce data from the market as a whole that shows discriminatory patterns in the neighborhood. The individual owner simply cannot be held accountable for the activities of others, just as he should likewise be unable to refute through general market statistics a showing of discrimination. Even if a utilitarian overview of the situation shows that it would result in greater overall accuracy to grant the would-be tenant a remedy, we would consider the evidence irrelevant to the particular case.

Id. at 150 n.129.
“group character evidence”) has been recognized by both commentators and courts. For example, in State v. McMillan, the Ohio Court of Appeals overturned a conviction for child sexual abuse based on the prosecution’s use of profile evidence. The defendant had acknowledged on cross-examination that he had been sexually abused as a child; thereafter, the prosecution called a police detective to testify, over a defense objection, that in his experience 85 to 90 percent of child sexual abusers had themselves been victims of abuse. The appellate court observed that this testimony was improper character evidence. It stated:

‘Group’ character evidence attempts to prove that because other persons have acted in certain ways in the past, a defendant who shares common characteristics with those persons is likely to have acted the same way with respect to the crime charged. A moment’s reflection on these categories of evidence reveals that ‘group’ character evidence is objectionable for the same reason as is traditional character evidence: probative value depends upon the jury drawing the forbidden inference that the defendant has a propensity to commit the crime with which he is charged.

116. See Andrew E. Taslitz, Myself Alone: Individualizing Justice Through Psychological Character Evidence, 52 Md. L. Rev. 3, 26 (1993) (noting that “profile” evidence is a form of group character evidence that is used to show that a person does or does not fit the profile of a rapist or child abuser in order to support the inference that it is more or less likely that the defendant committed the charged crime); Robert P. Mosteller, Legal Doctrines Governing the Admissibility of Expert Testimony Concerning Social Framework Evidence, 52 Law & Contemp. Probs. 85, 104-05 (1989) (recognizing the similarity between social framework evidence and character evidence); Mark S. Brodin, Behavioral Science Evidence in the Age of Daubert: Reflections of a Skeptic, 73 U. Cin. L. Rev. 867, 882 n.66 (2005) (“It must be noted that profile evidence offered by the prosecution regarding the defendant (unlike syndrome evidence regarding the victim) is generally not admissible, as it is viewed as unreliable and runs afoul of the character evidence prohibition against painting the defendant as a particular ‘criminal-type.’”).


118. Id. at 33.

119. Id. at 31. The detective first testified that in seminars he had attended he was told that in 85 to 90% of cases, the offender had himself been a victim. Coincidentally, his own experience matched those lessons exactly. Id. Even apart from the character-evidence issue, this evidence should have been excluded as misleading anyway. The detective’s testimony invites (though it does not compel) the jury to engage in the transposition fallacy, equating the probability that a person was subjected to abuse as a child with the probability that he is a child molester if he was subjected to abuse as a child. Id. at 31-32.

120. Id. at 32. See also State v. Claffin, 690 P.2d 1186, 1190 (Wash. App. 1984) (finding that testimony that 43% of child molestations were committed by “father-figures,” in a case where defendant was a father-figure, was “extremely prejudicial and should not have been admitted”). Cf.
Similarly, in *Bushell v. Dean*, the court overturned a judgment in favor of a sexual-harassment plaintiff on the ground that the trial court had admitted expert testimony that the accused harasser fit the profile of a sexual harasser.  

The court held that the profile testimony was inadmissible character evidence under Texas’s version of Rule 404 (and also that it was irrelevant and unfairly prejudicial). Other courts have relied instead on Rules 401 and 403 to exclude profiling evidence, but the rationales are similar.

In order for evidence of other acts to be admissible under Rule 404(b), the proponent “must explain how [the evidence] fits into a chain of inferences—a chain that connects the evidence to a proper purpose, no link of which is a forbidden propensity inference.” How does the fact that X percent of employers engage in discrimination bear upon the question of whether the defendant employer did so, if not to rely on a propensity inference?

As previously mentioned, Rule 404(b) allows admission of evidence of other acts to prove something other than propensity. In discrimination cases, the most likely candidate would be intent, since intent is critical in both individual and systemic disparate-treatment cases. The “me too” evidence at issue in *Mendelsohn* is sometimes

---

United States v. Doe, 149 F.3d 634, 638 (7th Cir. 1998) (rejecting the “group character evidence” argument where testimony about Nigerian drug trafficking patterns was not aimed at demonstrating that defendant had a propensity to import drugs but “served only to illuminate the modus operandi of Nigerian importers of Southeast Asian heroin”). The *Doe* court did, however, express sympathy with the “core concern” underlying the “group character evidence” argument, “which is that juries should not conclude that a particular defendant is guilty simply because the defendant shares some characteristics with a particular group,” which is the main argument in the case of Bayesian priors. *Id.*

121. Among other things, the expert testified that a typical harasser is male, married, the victim’s supervisor, and has known the victim for at least six months. *Id.* at 655. Sounds dangerous. *See also* Miller v. Commonwealth, 77 S.W.3d 566, 572 (Ky. 2002) (“[A] party cannot introduce evidence of the habit of a class of individuals either to prove that another member of the class acted the same way under similar circumstances or to prove that the person was a member of that class because he/she acted the same way under similar circumstances.”). Note again the implied application of the transposition fallacy. Even if most harassers are male, married, the victim’s supervisor, and have known the victim for at least six months, that tells us very little about the probability that a man with those characteristics is a harasser.

122. *See Bushell*, 781 S.W.2d at 655-56.

123. *See* Claus, *supra* note 113, at 992 (“A majority of jurisdictions considering profile and syndrome evidence analyze it exclusively under Rules 401, 403, and 702. That is, they do not even acknowledge that the evidence may raise a character inference.”). Claus also notes that Rule 405 seems to contemplate that character evidence would come in as lay testimony rather than expert testimony. *Id.* at 976.


125. *See* FED. R. EVID. 404(b)(2); *See supra* Part II.B.
ruled admissible—supposedly not to prove a propensity to discriminate but to prove intent, a permissible use under Rule 404(b)—"because of the inherent difficulty of proving state of mind." As Lisa Marshall has argued, however, when courts admit plaintiffs' evidence of other discriminatory acts by the employer in a discrimination case, they "routinely fail to comply with Rule 404." Courts either do not acknowledge that the other acts raise a Rule 404 issue at all, which seems to be the usual case, or they misapply the rule by holding that the other acts are admissible to show motive or intent. Yet, as Marshall argues, whether admitted to show motive or intent, the evidence "tends to show nothing more than that the defendant had some tendency—some propensity—to act discriminatorily, and it is from this fact alone that the fact-finder can infer an enduring discriminatory motive.

The use of other acts of discrimination to prove intent are on far shakier ground than what might be thought of as a classic 404(b) intent situation. In the latter kind of case, for example, the prosecution in a murder trial might introduce evidence that the defendant husband had previously beaten his wife to rebut his claim that when he shot her it was accidental. The earlier assault would tend to show that he had a harmful intent toward his wife. However, as emphasized above, base-rate evidence in discrimination cases is not of this sort, because it does not consist of evidence of prior acts of the employer itself that would shed light on its state of mind. One supposes that at the murder trial, the prosecutor could not introduce testimony about base rates of spousal murder, either in general or by people who have previously assaulted them. We would no doubt say that the question is whether this defendant had the intent to kill, not how many other husbands have the

---

126. See supra text accompanying notes 105-10.
127. Heyne v. Caruso, 69 F.3d 1475, 1480 (9th Cir. 1995).
128. Marshall, supra note 107, at 1065. She also notes a broader point, which is that "propensity-based inferences exist whenever employees attempt to prove intent through statistics," not just when they rely on discrete "me-too" evidence. Id. at 1082.
129. Id. at 1073-74.
130. Id. at 1075 (referring to motive). See also id. at 1082 ("Prior conduct goes to propensity, which goes to present intent. Nothing complicates this straightforward logical progression—other than a judge's insistence upon a faithful adherence to Rule 404."). Because "faithful adherence to the propensity ban would, in practice, drastically limit the relief available to plaintiffs," id. at 1087, Marshall advocates replacing Rule 404 with a rule that would presume inadmissibility of prior acts evidence and subject it to more rigorous scrutiny for prejudice, id. at 1096.
131. See id. at 1097.
133. Id.
134. See supra Sections I, II, II.A, II.B.
intent to kill their wives.\textsuperscript{136}

To reinforce the point that evidence of Bayesian priors is really propensity evidence, recall that the fourth prong of the \textit{Huddleston} analysis is to apply Rule 105, which requires the judge to provide a limiting instruction to the jury, upon request, cautioning the jury that it may consider the evidence only for its proper purpose and not for the forbidden propensity inference.\textsuperscript{137} It is instructive to consider what the limiting instruction would look like in a case in which the plaintiff attempted to introduce evidence of Bayesian priors.\textsuperscript{138} The jury should be told some version of the following: "You may consider the evidence about the rate of employer discrimination on the question of \underline{___________}, but you may not consider it as bearing on whether and to what extent the defendant possessed a propensity to engage in discrimination."\textsuperscript{139} What proper purpose unrelated to the employer's propensity for (or likelihood of) discrimination could the plaintiff use to fill in the blank? There simply does not seem to be one.\textsuperscript{140}

Even if introduction of Bayesian priors did not conflict with the Rules, instructing juries in Bayesian analysis would be a misguided enterprise. A central contention of Bayesians is that human decision-makers are non-Bayesian in their approach, to the chagrin of Bayesians, who believe this to reflect a flaw in the human capacity for reason.\textsuperscript{141} But it is precisely because fact-finders tend to find a Bayesian approach

\begin{itemize}
  \item \textsuperscript{136} \textit{Id.} at 342, 344-47.
  \item \textsuperscript{137} \textit{Id.} at 343-45 ("The proponents of Rule 404(b) evidence must do more than conjure up a proper purpose—they must also establish a chain of inferences no link of which is based on a propensity inference."); \textit{Ansell v. Green Acres Contracting Co., Inc.}, 347 F.3d 515, 520 (3d Cir. 2003) ("To be admissible under Rule 404(b), other acts evidence must be offered for a proper purpose, i.e., a purpose other than showing that an individual has a propensity or disposition for certain activity."); \textit{United States v. Patterson}, 20 F.3d 809, 813 n.3 (10th Cir. 1994) (disapproving the "laundry list" limiting instruction that "quoted almost all of the proper purposes listed in Rule 404(b) and went beyond the actual purposes for which the evidence was offered."); \textit{United States v. Cardall}, 885 F.2d 656, 671 (10th Cir. 1989) ( ["T]he trial court must 'specify the purpose for which such evidence is offered and a broad statement merely invoking or restating Rule 404(b) will not suffice."]).
  \item \textsuperscript{138} \textit{See Ansell}, 347 F.3d at 520-21.
  \item \textsuperscript{139} \textit{Patterson}, 20 F.3d 809 at 812-13.
  \item \textsuperscript{140} \textit{See United States v. Miller}, 673 F.3d 688, 701-02 (7th Cir. 2012). The \textit{Miller} court disapproved of a jury instruction stating, "You may consider this evidence only on the questions of knowledge, intent and in deciding whether the defendant's testimony is truthful in whole, in part, or not at all." \textit{Id.} at 701. Further, the \textit{Miller} court noted that "when the government cannot explain how the prior conviction [for drug dealing] relates to the question of intent without resorting to a propensity inference, it would be unfair to expect the jury to do so based only on this instruction." \textit{Id.} at 702.
  \item \textsuperscript{141} \textit{See Daniel Kahneman \& Amos Tversky, Subjective Probability: A Judgment of Representativeness}, 3 COGNITIVE PSYCHOL. 430, 449-50 (1972) ("In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all").
\end{itemize}
non-intuitive that it would be a mistake to attempt to incorporate it in the decisions of judges and juries.\textsuperscript{142} Even Judge Posner, a dedicated Bayesian, believes that the value of Bayes’ theorem in the law is heuristic rather than a prescription for decision.\textsuperscript{143} As he notes, efforts aimed at “formalizing the tacit, intuitive inferential procedures now employed to resolve factual disputes in trials would cause endless confusion.”\textsuperscript{144} Jurors should resolve cases based upon their conviction that an employer did or did not discriminate, not on the basis of an algorithm that does not track the way people actually think, thus divorcing the jurors’ job from their sense of justice.

CONCLUSION

The push for Bayesian analysis in discrimination cases is animated, at least in part, by a recognition that the typical mode of statistical inference in discrimination cases is faulty. Yet substituting one flawed methodology for another is hardly an improvement. The Bayesian methodology necessarily relies on arbitrary and speculative assumptions about base rates of discrimination and also allows an employer to be held liable based in large part on conclusions about the frequency of wrongdoing by other employers.\textsuperscript{145} Those advocating the advantage of one statistical technique over another should consider that perhaps in many cases statistical evidence simply lacks probative value and may not be very useful in determining whether a particular employer has engaged in discrimination.

\textsuperscript{143} \textit{Id.}
\textsuperscript{144} \textit{Id.}
\textsuperscript{145} \textit{See supra} Sections II.B, II.C.