

COMMENTS

From *Price Waterhouse* to *Dukes* and Beyond: Bridging the Gap Between Law and Social Science by Improving the Admissibility Standard for Expert Testimony

By R. Matthew Wise†

I. INTRODUCTION	546
II. THE ADMISSIBILITY STANDARD FOR EXPERT TESTIMONY	551
A. <i>The Frye “General Acceptance” Standard</i>	551
B. <i>The Emergence of the Federal Rules of Evidence and Daubert</i>	552
C. <i>The Admissibility Standard Set Out by Daubert</i>	552
D. <i>Evaluating the Court’s Use of Scientific Terminology in Daubert</i>	554
E. <i>Kumho and the Post-Daubert Era</i>	555
F. <i>The 2000 Amendment to Federal Rule of Evidence 702</i>	557
G. <i>The Question Whether to Conduct a Full Daubert Analysis at the Class Certification Stage</i>	557
H. <i>Summarizing the Transformation in the Admissibility Standard and Noting Its Significance to Employment Discrimination Cases</i>	558
III. THE USE OF SOCIAL FRAMEWORK EXPERTS IN EMPLOYMENT DISCRIMINATION CASES	559
A. <i>Distinguishing Social Framework Experts from Employment Practices Experts</i>	559

† J.D., University of California, Berkeley School of Law (Boalt Hall), 2005. B.A., University of California, Los Angeles, 2002. Many thanks to Professor Linda Krieger and Aaron Kaufmann, Esq., Hinton, Alfert & Sumner, for their guidance and suggestions. Thanks also to Christine Wise and Kristine Shedd for their insights on non-legal aspects of the Comment, and to the BJELL editors, especially Sarah Laubach, Ann O’Leary, Brian Moskal, Bridget Smith, and Eli Naduris-Weissman, for their patience and hard work in revising the Comment.

B.	<i>The Body of Social Science Research on which Social Framework Evidence is Based</i>	562
C.	<i>The Advantages of Social Framework Evidence over Original Research in a Litigation Context</i>	563
IV.	THE DIVERGENCE OF LAW AND SOCIAL SCIENCE IN THE DAUBERT STANDARD	565
A.	<i>Examining Whether Judges are Equipped to Recognize Flawed Research</i>	566
B.	<i>Problematic Interpretations and Explanations of Social Framework Testimony</i>	569
1.	<i>Interpreting Social Framework Evidence Inconsistently With the Scope of the Findings of This Evidence</i>	569
2.	<i>Regarding as Common Knowledge Specialized Knowledge Helpful to the Jury</i>	570
3.	<i>Relying Solely On Case Precedent And The Expert's Reputation</i>	571
4.	<i>Treating Sound Methodology as an Issue of Weight and Not Admissibility</i>	572
5.	<i>Failing to Substantiate the Decision to Admit or Exclude Testimony</i>	572
V.	BRIDGING THE GAP BETWEEN LAW AND SOCIAL SCIENCE.....	573
A.	<i>Strengthening Scientific Training for Attorneys and Judges</i> ..	574
B.	<i>Modifying the Language Added in the 2000 Amendment to Federal Rule of Evidence 702</i>	575
1.	<i>The Current Role of the Four Daubert Factors and the 2000 Amendment</i>	575
2.	<i>Revising the 2000 Amendment</i>	577
C.	<i>A Recommended Approach to the Admissibility Standard for Expert Testimony</i>	578
VI.	CONCLUSION.....	580

I.

INTRODUCTION

On June 23, 2004, headlines in newspapers around the country announced the largest workplace-bias lawsuit in United States history.¹ The

1. See Steven Greenhouse & Constance L. Hays, *Wal-Mart Sex-Bias Suit Given Class-Action Status*, N.Y. TIMES, June 23, 2004, at A1; see also Lisa Girion & Abigail Goldman, *Wal-Mart Must Face Huge Sex-Bias Suit*, L.A. TIMES, June 23, 2004, at A1; Bob Egelko & Jenny Strasburg, *Wal-Mart Faces Huge Sex Discrimination Suit*, S.F. CHRON., June 23, 2004, at A1; Amy Joyce, *Wal-Mart Bias*

day before, Judge Martin Jenkins had certified *Dukes v. Wal-Mart Stores, Inc.*,² a class action with 1.6 million current and former female employees who claim that Wal-Mart paid them less and offered them fewer opportunities for promotion than male employees.³ The decision sparked much commentary, including predictions of the financial implications for Wal-Mart⁴ and of the effects on female employees in other companies throughout the country.⁵

Lost in the media scrutiny of the case was analysis of the vital role expert testimony played in establishing the plaintiffs' claims against Wal-Mart.⁶ To demonstrate Wal-Mart's company-wide discrimination against women, the plaintiff class relied both on Professor Richard Drogin's statistical findings, which identified gender disparities in pay and promotions at Wal-Mart, and on Dr. William T. Bielby's social framework analysis,⁷ which produced evidence that Wal-Mart's policies and practices included subjective decision making, a culture of corporate uniformity, and gender stereotyping.⁸ Despite Wal-Mart's motions to strike,⁹ Judge Jenkins

Case Moves Forward, WASH. POST, June 23, 2004, at A1; Greg Burns, *Class Action OKd against Wal-Mart*, CHI. TRIB., June 23, 2004, at C1.

2. 222 F.R.D. 137 (N.D. Cal. 2004). *Dukes* is the latest in a series of employment discrimination class action lawsuits certified in the Northern District of California over the last decade or so. See, e.g., *Stender v. Lucky Stores*, 803 F. Supp. 259 (N.D. Cal. 1992); *Butler v. Home Depot, Inc.*, 1996 WL 421436 (N.D. Cal. 1996).

3. Greenhouse, *supra* note 1.

4. Ann Zimmerman, *Judge Certifies Wal-Mart Suit as Class Action*, WALL ST. J., June 23, 2004, at A1.

5. Amy Joyce, *Wal-Mart Pay-Gap Suit a Caution to Employers*, S.F. CHRON., June 25, 2004, at C3.

6. With the exception of the coverage in *The New York Times*, major newspapers ignored this foundational aspect of the case. See Greenhouse, *supra* note 1 (describing how the plaintiff class relied on the findings of statistical expert, Dr. Richard Drogin, who demonstrated pay disparities between male and female employees at Wal-Mart, in combination with the findings of sociology expert, Dr. William T. Bielby, who demonstrated that women fill a disproportionately large number of nonmanagerial positions and a disproportionately small number of managerial positions).

7. Dr. Bielby provided a description of social framework analysis in his expert report:

Over the past twenty years, much of my research has focused on issues of workplace discrimination, and on organizational policies and practices more generally . . . I have reviewed the deposition testimony of Wal-Mart managers responsible for creating and implementing the company's personnel policies as well as the testimony of managers who made decisions about compensation, hiring, promotion, job assignment, and related personnel matters. I have also reviewed the documents used as exhibits in the depositions of these individuals . . . I have also relied upon a large body of social research on organizational policy and practice and on workplace bias . . . My method is to look at distinctive features of the firm's policies and practices and to evaluate them against what social science research shows to be factors that create and sustain bias and those that minimize bias. In litigation contexts, this method of analysis is known as "social framework analysis."

William T. Bielby, Expert Report of William T. Bielby, Ph.D. (Feb. 3, 2003), <http://www.walmartclass.com/staticdata/reports/r3.html>. Dr. Bielby's social framework analysis led him to conclude that employment decisions made based on Wal-Mart's personnel policy and practice are vulnerable to gender bias. *Id.*

8. *Dukes v. Wal-Mart Stores, Inc.*, 222 F.R.D. 137, 187-88 (N.D. Cal. 2004).

admitted Dr. Drogin's testimony because his statistical methods created an inference of company-wide discrimination,¹⁰ and citing *Price Waterhouse v. Hopkins*,¹¹ he admitted Dr. Bielby's testimony "because it raise[d] an inference of corporate uniformity and gender stereotyping . . . common to all class members."¹²

Over a decade has passed since *Price Waterhouse*, where the Supreme Court gave an approving nod to Dr. Susan Fiske's explanation of sex stereotyping,¹³ and *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, where the Court transformed the judiciary's role in evaluating expert testimony.¹⁴ In this period of time, plaintiffs who have alleged that they are victims of employment discrimination increasingly have relied on experts like Dr. Bielby to provide the courts with a "social framework" from which to view issues crucial to their cases. Social framework evidence is the product of social science research that an expert compiles and uses to construct a frame of reference for specific issues central to the resolution of a case.¹⁵ This evidence is especially useful in employment discrimination cases, where the expert constructs such a framework to examine whether an organization's policies and practices are vulnerable to stereotyping and bias.¹⁶ Pursuant to *Daubert*, judges must decide whether to admit social framework testimony like that in Dr. Bielby's report. As in *Dukes*, these decisions can determine the outcome of key phases in employment discrimination cases, because plaintiffs generally rely on this testimony to explain the mechanisms that create statistical disparities in the workforce.¹⁷

This Comment examines the treatment of social framework testimony in federal employment discrimination cases with the primary purpose of highlighting potential flaws in these decisions that stem from weaknesses in the admissibility standard for expert testimony. In doing so, this Comment also should provide employment law practitioners with a sense of how judges have dealt with social framework testimony since *Daubert*. Despite plaintiffs' increasing dependence on social framework testimony to show employment discrimination, few articles move past *Price Waterhouse* in

9. *Id.* at 189.

10. *Id.* at 164.

11. 490 U.S. 228 (1989).

12. *Dukes*, 222 F.R.D. at 154.

13. See *Price Waterhouse v. Hopkins*, 490 U.S. 228 (1989).

14. See 509 U.S. 579 (1993).

15. Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559, 570 (1987).

16. William T. Bielby, *Can I Get a Witness? Challenges of Using Expert Testimony on Cognitive Bias in Employment Discrimination Litigation*, 7 EMP. RTS. & EMP. POL'Y J. 377, 390 (2003) [hereinafter Bielby, *Can I Get a Witness?*]; see Jane Goodman & Robert T. Croyle, *Social Framework Testimony in Employment Discrimination Cases*, 7 BEHAV. SCI. & L. 227, 238-39 (1989).

17. See Bielby, *Can I Get a Witness?*, *supra* note 16, at 385-86.

discussing this evidence. This Comment attempts to fill this void by touching on a broader set of cases involving social framework evidence.

The cases selected for analysis are all federal employment discrimination cases; excluding state employment discrimination cases ensures that the admissibility standard set out by *Daubert* governed all cases in the sample. However, the sample includes both individual cases and class actions, raising the question whether the admissibility standard in a class action should be the same at the class certification stage as at trial.¹⁸ While the sample is thorough, it is not exhaustive, in large part because employment discrimination cases that involve social framework evidence are difficult to pinpoint, as this evidence is known by many different names and is used in a variety of practice areas.

The Comment's primary objective is to examine whether the federal admissibility standard for expert testimony might be revised to help the legal community analyze this and other expert testimony in a manner more consistent with the purpose of *Daubert*. *Daubert* was groundbreaking in that it compelled federal judges to evaluate expert testimony in terms of the soundness of its methodology, not its "general acceptance" in the field.¹⁹ For a long time, the methodology of social science evidence fell into one of two categories; either general research was used to resolve general questions of law and policy (in a process intended to prove "legislative facts"), or specific research was used to resolve issues specific to the case at hand (in a process intended to prove "adjudicative facts").²⁰ In contrast, the methodology of social framework testimony is unusual in that it applies *general* research principles to draw conclusions about *specific* issues in dispute.²¹

Thus, in determining whether this distinctive methodology is sound, a central factor is whether the social framework expert's research generalizes

18. See *infra* Part II.G.

19. See *infra* Part II.C.

20. Walker & Monahan, *supra* note 15, at 559. Kenneth Culp Davis first made this distinction between "legislative facts" and "adjudicative facts" in 1942. *Id.* at 561 (citing Kenneth Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364, 402-03 (1942)). Walker and Monahan provide examples of how social science research is used for these purposes. To determine the scope of the exclusionary rule—a legislative fact—the Supreme Court took into account the findings of social science research concerning the effects of the exclusionary rule on the disposition of felony arrests. *Id.* at 562 (discussing *United States v. Leon*, 468 U.S. 897 (1984)). To determine whether the defendant had created consumer confusion by selling a toy replica of a copyrighted car—an adjudicative fact—the plaintiff introduced a social science survey intended to discern whether people identified the item with the copyrighted car. *Id.* at 562-63 (discussing *Processed Plastic v. Warner Commc'ns*, 675 F.2d 852 (7th Cir. 1982)).

21. *Id.* at 559. In describing cases that use social framework evidence, Walker and Monahan observe how this evidence is different from evidence intended to prove legislative or adjudicative facts: "The expert witnesses relied heavily—and in some cases, exclusively—on . . . research studies published before the events that gave rise to the litigation took place, studies performed by researchers and using subjects with no knowledge of the case at bar." *Id.* at 568.

to the specific issue in dispute.²² In scientific terms, this factor is known as “external validity.”²³ The admissibility decisions involving social framework evidence offer a unique vantage point from which to observe the transition to *Daubert* because they reveal that judges often underutilize factors such as external validity and instead fall back on proxy factors for “trustworthiness” such as the old admissibility standard of general acceptance in the field.²⁴ More generally, research on the legal community’s understanding of social science evidence demonstrates that judges have difficulty applying the admissibility standard set out in *Daubert*.²⁵ These observations suggest that the admissibility standard for expert scientific testimony is an area where reform might facilitate the legal community’s use of social science.²⁶

Part II introduces the admissibility standard for expert testimony, outlined in Federal Rule of Evidence 702 and in the Court’s decisions in *Daubert* and *Kumho Tire Co. v. Carmichael*. Part III begins by distinguishing between social framework experts, who offer a framework based on social science research from which to observe conduct in the workplace, and “employment practices” experts, who generally offer unscientific expertise about the workplace as human resources experts, consultants, employment lawyers, or the like. Part III goes on to provide an overview of the research on which social framework testimony is based and a few examples of how this testimony is used in employment discrimination cases. Part IV highlights research about whether judges know how to apply the *Daubert* standard and whether they recognize weaknesses in social science evidence. Part IV also questions whether judges interpret and explain this evidence accurately in their decisions. Finally, with the hope of providing the legal community a more solid foundation on which to examine social science evidence, Part V offers three suggestions: (1) upgrading scientific training for law students, attorneys, and judges, (2)

22. Walker and Monahan treat this factor as part of the relevance inquiry of FED. R. EVID. 401 by asking “whether the framework makes any fact in the case ‘more probable or less probable.’” *Id.* at 572. However, *Daubert* and FED. R. EVID. 702 pose the additional hurdle of evaluating the soundness of the methodology of the evidence.

23. “External validity” considers the question, “How far are we justified in taking the data, assuming them to be accurate in their own terms, and in drawing implications or conclusions in other context either very closely related . . . or much further away?” Mark P. Denbeaux & D. Michael Risinger, *Kumho Tire and Expert Reliability: How the Question You Ask Gives the Answer You Get*, 34 SETON HALL L. REV. 15, 39 (2003). External validity can be contrasted from “internal validity,” which considers the question, “Are the data any good in their own apparent terms, given the way they were generated?” *Id.*

24. *See infra* Part II.D.

25. *See infra* Part IV.A.

26. In recent years, the legal and social science communities have recognized the need to search for ways to promote interdisciplinary approaches. *See, e.g.*, Jeremy A. Blumenthal, *Law and Social Science in the Twenty-First Century*, 12 S. CAL. INTERDISC. L.J. 1, 4 (2002).

aligning the legal and scientific definitions of “reliability” and “validity” in the 2000 amendment to Federal Rule of Evidence 702, and (3) providing judges with guidance in evaluating social science evidence by adding to the Rule 702 advisory note a recommended approach to making admissibility decisions. The first suggestion piggybacks on proposals made by several academics, while the latter two arise from the observations in this Comment.

II.

THE ADMISSIBILITY STANDARD FOR EXPERT TESTIMONY

Over the past decade, the admissibility standard for expert testimony has evolved significantly. The source most responsible for this development is *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, which instituted a new doctrine for the admissibility of scientific evidence in federal cases.²⁷ *Daubert* resolved the debate about whether Federal Rule of Evidence 702 had asserted itself over the long-standing *Frye* “general acceptance” standard, recognizing the definitive role of Rule 702 in assessing scientific evidence.²⁸ Part II reviews the progression from *Frye* to *Daubert* and post-*Daubert* developments.

A. The *Frye* “General Acceptance” Standard

In *Frye v. United States*, the Court of Appeals for the District of Columbia held that for scientific evidence to be admissible, it “must be sufficiently established to have gained general acceptance in the particular field in which it belongs.”²⁹ Decided over eighty years ago, *Frye* dealt with whether a systolic blood pressure deception test—an early version of the modern polygraph test—should be admitted into evidence.³⁰ The Court of Appeals determined that this technique had not gained a general level of recognition and approval, and accordingly, it held that in order for scientific evidence to be admitted, it must meet a standard of “general acceptance.”³¹

This “general acceptance” standard eventually became a measure used by federal and many state courts to judge the admissibility of scientific evidence.³² The purpose of *Frye* was to prevent the admission of evidence based on untested scientific principles that might guide the court to an

27. 509 U.S. 579 (1993).

28. *Id.* at 588-89.

29. 293 F. 1013, 1014 (D.C. Cir. 1923).

30. *Id.* at 1013.

31. *Id.* at 1014.

32. 1 DAVID L. FAIGMAN, MODERN SCIENTIFIC EVIDENCE § 1-2.3 (2d ed. 2002).

unfounded conclusion.³³ Over the years, critics questioned many aspects of *Frye*, principally its vague and easily manipulated criterion³⁴ and its tendency to exclude probative evidence that had not yet been “generally accepted.”³⁵

B. *The Emergence of the Federal Rules of Evidence and Daubert*

While the *Frye* “general acceptance” standard was not cited for a decade after it surfaced, it became increasingly popular by the 1970s.³⁶ However, the Supreme Court approved the Federal Rules of Evidence in 1972, and Congress adopted them in 1975.³⁷ Federal Rule of Evidence 702, which governs all expert testimony in federal court cases, provided: “If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.”³⁸ For almost two decades, judges and scholars debated whether *Frye* and Federal Rule of Evidence 702 could coexist, or whether they were mutually exclusive.³⁹

In 1993, the Supreme Court settled this dispute, instituting a new doctrine in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* that relied solely on the “plain meaning” of Federal Rule of Evidence 702 in the decision whether to admit expert scientific testimony.⁴⁰ The Court rejected the exclusive use of the *Frye* “general acceptance” standard, observing that Rule 702 should be interpreted like any statute⁴¹ and that its language does not mandate that scientific evidence be “generally accepted” to be admissible.⁴²

C. *The Admissibility Standard Set Out by Daubert*

Perhaps the most noteworthy aspect of *Daubert* is the responsibility it places on judges in the analysis of expert testimony.⁴³ Under *Frye*, judges

33. Henry F. Fradella et al., *The Impact of Daubert on the Admissibility of Behavioral Science Testimony*, 30 PEPP. L. REV. 403, 406 (2003).

34. 1 FAIGMAN, *supra* note 32, § 1-2.4.

35. Fradella, *supra* note 33, at 406.

36. 1 FAIGMAN, *supra* note 32, § 1-2.2.

37. John V. Jansonius & Andrew M. Gould, *Expert Witnesses in Employment Litigation: The Role of Reliability in Assessing Admissibility*, 50 BAYLOR L. REV. 267, 274 (1998).

38. FED. R. EVID. 702 was amended in 2000 to include additional language. *See infra* Part II.F.

39. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 587 n.5 (1993).

40. 1 FAIGMAN, *supra* note 32, § 1-3.0.

41. *Daubert*, 509 U.S. at 587.

42. *Id.* at 589.

43. 1 FAIGMAN, *supra* note 32, § 1-3.0.

simply deferred to the opinions of scientists in the “particular field” when deciding whether expert testimony was “generally accepted,” and thus admissible.⁴⁴ *Daubert* was revolutionary in the sense that it purported to pass the duty of evaluating expert testimony to judges, compelling them to become familiar with scientific methodology.⁴⁵ The new doctrine, according to the Court, requires judges to act as “gatekeepers” in determining whether expert testimony displays both reliability and relevance.⁴⁶

In its discussion of the evidentiary reliability prong, the Court focused on the language of Federal Rule of Evidence 702, emphasizing that the words “scientific” and “knowledge” imply that testimony must have grounding in the methods and procedures of an objective scientific inquiry.⁴⁷ The Court also set out some general guidelines to help trial court judges, in their function as gatekeepers, assess the evidentiary reliability of the reasoning and methodology of expert testimony.⁴⁸ These factors include the following: (1) whether the theory or technique has been or can be reliably tested, (2) whether the theory or technique has been subjected to peer review and publication, (3) the technique’s known or potential rate of error, and (4) whether the theory or technique has been “generally accepted.”⁴⁹ Unlike *Frye*, where the “general acceptance” standard was dispositive, *Daubert* emphasized that the “inquiry envisioned by Rule 702 is . . . a flexible one.”⁵⁰

The relevance prong of the *Daubert* admissibility standard is codified in Federal Rules of Evidence 402 and 702. Pursuant to Rule 402, expert scientific testimony, like all evidence, must be relevant to be admissible.⁵¹ Rule 702 focuses on relevance more explicitly in terms of the “fit” of the testimony to a disputed issue in the case, requiring testimony to “assist the trier of fact to understand the evidence or to determine a fact in issue.”⁵² “Fit” traditionally refers to the helpfulness of the expert’s knowledge to resolve the specific issue in dispute. Some judges assessing social framework testimony have used “fit” to assess external validity—whether the expert’s research generalizes to the issue in dispute.⁵³

44. *Id.* § 1-2.3.

45. *Id.* § 1-3.0.

46. *Daubert*, 509 U.S. at 594-95, 597.

47. *See id.* at 589-90.

48. *Id.* at 592-93.

49. *Id.* at 593-94.

50. *Id.* at 594.

51. *Id.* at 587.

52. *See* FED. R. EVID. 702; *see also Daubert*, 509 U.S. at 591.

53. *See* 1 FAIGMAN, *supra* note 32, § 1-3.2. By asking whether the expert’s research “generalizes” to the issue in dispute, judges examine whether the expert’s research provides a sound basis for deciding the issue.

Finally, although *Daubert* discussed the evaluation of the expert's qualifications only briefly, Rule 702 also requires that the expert be qualified to present testimony. Rule 702's definition of expertise is broad, including those qualified "by knowledge, skill, experience, training, or education."⁵⁴ The analysis of the expert's qualifications should be conducted within the context of the proposed testimony's evidentiary reliability and relevance.⁵⁵

D. Evaluating the Court's Use of Scientific Terminology in *Daubert*

Unfortunately, in the bulk of the *Daubert* opinion, the Court used "reliability" and "validity" interchangeably, despite their precise scientific meanings.⁵⁶ In a mere footnote,⁵⁷ the Court acknowledged that scientists differentiate between validity, which assesses whether a principle supports what it purports to show, and reliability, which assesses whether a principle produces consistent results.⁵⁸ However, the Court followed by asserting that when it used the term "reliability," it meant *evidentiary* reliability, a measure of "trustworthiness" it likened to "scientific validity."⁵⁹

By sometimes referring to reliability by its scientific meaning (reproducibility) and other times equating reliability with evidentiary

54. See FED. R. EVID. 702; see also 1 FAIGMAN, *supra* note 32, § 1-3.3.1.

55. See 1 FAIGMAN, *supra* note 32, § 1-3.3.2.

56. See Joelle Anne Moreno, *Beyond the Polemic Against Junk Science: Navigating the Oceans that Divide Science and Law with Justice Breyer at the Helm*, 81 B.U. L. REV. 1033, 1065 ("Daubert introduced the concepts of scientific 'validity' and 'reliability' to a federal bench and bar that likely knew little of the precise scientific meanings of these terms."). As Moreno explains, "[t]o a scientist, 'reliability' and 'validity' are terms of art that have different and precise definitions. The difference between scientific reliability and validity is belied by the Supreme Court's assertion in *Daubert* that 'the difference between . . . validity and reliability may be such that each is different from the other by no more than a hen's kick.'" *Id.* at 1066.

57. Footnote 9 of *Daubert* is far from clear in its attempt to simplify the concept of reliability:

We note that scientists typically distinguish between "validity" (does the principle support what it purports to show?) and "reliability" (does application of the principle produce consistent results?). Although "the difference between accuracy, validity, and reliability may be such that each is distinct from the other by no more than a hen's kick," our reference here is to *evidentiary* reliability—that is trustworthiness. In a case involving scientific evidence, *evidentiary* reliability will be based upon *scientific* validity.

509 U.S. at 590 n.9 (citations omitted).

58. *Id.* Moreno further distinguishes and defines reliability and validity:

As a term of art in science and statistics, reliability refers to the reproducibility of data. A reliable test can be repeated under identical circumstances and yield the same results. The results may be consistently wrong, but that is an issue of validity, not reliability. . . . "Validity" is also a term of art in science . . . [that] relate[s] to the connection between a theory or results of a particular study and the empirical world. . . . [Thus,] validity depends instead on the concordance between the theory by which one interprets data and how the inferences drawn from the data are going to be used."

Moreno, *supra* note 56, at 1067-69 (quoting KENNETH R. FOSTER & PETER W. HUBER, *JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS* 143-46 (1999)).

59. *Daubert*, 509 U.S. at 590 n.9.

reliability or “trustworthiness,” the Court not only exacerbated judges’ generally unsophisticated understanding of scientific terminology, but also muddled the admissibility standard for expert testimony.⁶⁰ The distinction between reliability and validity is vital to the admissibility standard, because as one legal scholar observes, “[t]o understand and use science, judges and lawyers need to know more than just that another experimenter can replicate a particular test. [They] need to know whether a particular test assists in understanding natural phenomena relevant to the resolution of a specific legal dispute.”⁶¹ By equating evidentiary reliability with trustworthiness, the Court gave life to the *Frye* “general acceptance” standard, tempting judges to place undue weight on this standard and on other proxy factors for trustworthiness in their assessment of reliability.⁶²

The Court also reverted back to *Frye* in laying out the four *Daubert* factors. Significantly, only the first and third *Daubert* factors (whether the theory has been or can be reliably tested and the technique’s rate of error) focus *directly* on the scientific rigor of the evidence—and these factors only assess reliability (the consistency of the technique), not validity (whether the technique measures what it is intended to measure). The second and fourth factors (whether the theory has been subjected to peer review and publication and has been generally accepted), like the *Frye* standard, are proxies for trustworthiness, as they defer to the opinions of those in the particular field.⁶³ Therefore, depending on how judges interpret *Daubert*, it may not represent as sharp a break from *Frye* as the Court implied.⁶⁴

E. *Kumho and the Post-Daubert Era*

While *Daubert* clarified that Federal Rule of Evidence 702 supersedes the *Frye* standard, still at issue was whether the scope of *Daubert* is limited to expert *scientific* testimony, or whether *Daubert* applies to all testimony based on specialized knowledge. In 1999, the Supreme Court provided the answer in *Kumho Tire Co. v. Carmichael*.⁶⁵ Recognizing that the language of Rule 702 makes no distinction between “scientific” knowledge and

60. See Moreno, *supra* note 56, at 1068; see also Michael H. Graham, *The Expert Witness Predicament: Determining “Reliable” Under the Gatekeeping Test of Daubert, Kumho, and Proposed Amended Rule 702 of the Federal Rules of Evidence*, 54 U. MIAMI L. REV. 317, 336-337 (2000) (analyzing the Court’s use of reliability as both an indicator of sound scientific methodology and as the equivalent of “trustworthiness”).

61. Moreno, *supra* note 56, at 1069.

62. Cf. Graham, *supra* note 60, at 340-41.

63. 1 FAIGMAN, *supra* note 32, § 1-3.4.

64. See *id.*

65. 526 U.S. 137 (1999).

“technical” or “other specialized” knowledge,⁶⁶ the Court held that *Daubert* applies not only to expert scientific testimony, but to *all* expert testimony.⁶⁷

The Court also confirmed that trial judges, in their role as “gatekeepers,” may use a flexible approach in considering the four *Daubert* factors:

[W]e can neither rule out, nor rule in, for all cases and for all time the applicability of the factors mentioned in *Daubert*, nor can we do so now for subsets of cases categorized by category of expert or by kind of evidence. Too much depends upon the particular circumstances of the particular case at issue.⁶⁸

The Court concluded that trial judges have discretion to choose which procedure to use when assessing the evidentiary reliability of expert testimony and may even choose to ignore the *Daubert* factors if they are inapplicable to the evidence at issue.⁶⁹

Kumho emphasized that the task for judges is not to make an evaluation based on whether the evidence is scientific or non-scientific, but instead to assess the rigor of the *methods* relied upon by the expert.⁷⁰ This approach is especially instructive for the handling of social science evidence, which may rely on principles that are difficult to test and that are based more on advocacy than on objective reasoning.⁷¹

One of the more subtle but important developments in *Kumho* was partly responsible for inspiring this Comment’s suggestion to add to the Rule 702 advisory note a recommended approach to making admissibility decisions. In contrast to *Daubert*, where the Court conducted separate inquiries into evidentiary reliability and relevance, *Kumho* represented a subtle shift toward examining evidentiary reliability in the context of the relevance or “fit” of the testimony to the legal dispute.⁷² By making the relevance inquiry the first and primary focus, *Kumho* aligned the admissibility standard for expert testimony with the admissibility standard for all other evidence.

66. *Id.* at 147.

67. *Id.* at 141.

68. *Id.* at 150.

69. *Id.* at 152; see 1 FAIGMAN, *supra* note 32, § 1-3.5.

70. See 1 FAIGMAN, *supra* note 32, § 1-3.5.

71. See *id.*

72. See Moreno, *supra* note 56, at 1054. Summarizing its decision, the Court stated, “Rule 702 grants the district judge the discretionary authority, reviewable for its abuse, to determine reliability *in light of the particular facts and circumstances of the particular case.*” *Kumho*, 526 U.S. at 158 (emphasis added).

F. *The 2000 Amendment to Federal Rule of Evidence 702*

The Court's list of factors in *Daubert* and the expansion of the admissibility standard to incorporate all expert testimony in *Kumho* raised questions about how these decisions related to Federal Rule of Evidence 702. In 2000, Rule 702 was amended with the intention of clarifying the function of the rule in light of *Daubert* and *Kumho*.⁷³ The amendment includes three tests that expert testimony must meet to ensure evidentiary reliability: (1) testimony must be based upon sufficient facts or data, (2) testimony must be the product of reliable principles and methods, and (3) the witness must apply the principles and methods reliably to the facts of the case.⁷⁴

The first subpart of the amendment codifies *Kumho*, confirming that Rule 702 applies to all expert testimony, whether scientific or non-scientific, research- or experience-based, as long as the testimony is supported by facts or data. Notably, the amendment does not codify the *Daubert* factors, although the advisory committee's note asserts that they "remain relevant to the determination of the reliability of expert testimony under Rule 702 as amended."⁷⁵ Accordingly, the second subpart of the amendment implicitly incorporates the first and third *Daubert* factors, as testability and rate of error may be used to assess whether the principles can be reproduced consistently. The amendment also ensures consistency in the third subpart, requiring that the witness not stray from these reliable principles and methods when applying them to the facts of the case.

G. *The Question Whether to Conduct a Full Daubert Analysis at the Class Certification Stage*

In a class action, the class must be certified under Federal Rule of Civil Procedure 23 before the lawsuit can reach the merits.⁷⁶ Because plaintiffs often rely on expert testimony to satisfy Rule 23, the question arises whether the court should conduct a full *Daubert* analysis at the class certification stage, or whether the court should lower the standard at this stage to avoid considering the merits. In *Dukes v. Wal-Mart Stores, Inc.*,

73. See FED. R. EVID. 702 advisory committee's note. Rule 702 now reads:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

FED. R. EVID. 702 (emphasis added to reflect 2000 amendment).

74. FED. R. EVID. 702.

75. FED. R. EVID. 702 advisory committee's note.

76. See FED. R. CIV. P. 23.

for example, this lower standard measured only “whether Dr. Bielby’s opinion [was] so flawed that it lack[ed] sufficient probative value to be considered in assessing commonality.”⁷⁷ Although the Supreme Court has not ruled on this issue, several lower courts have indicated that a lower *Daubert* standard is appropriate at the class certification stage.⁷⁸

H. *Summarizing the Transformation in the Admissibility Standard and Noting Its Significance to Employment Discrimination Cases*

In summary, the past decade has witnessed a transformation of the admissibility standard for expert testimony. A crucial point to recognize, however, is that while *Daubert*, *Kumho*, and the 2000 amendment laid out tools to help judges assess the relevance, fit, and reliability of expert testimony, they failed to distinguish reliability from validity.⁷⁹ While some judges have interpreted the *Daubert* standard flexibly to incorporate an analysis of validity, this Comment suggests a modest revision of the 2000 amendment to integrate measures of validity explicitly. Such a revision, if supplemented with an advisory committee note distinguishing reliability from validity, would provide a clearer approach to the *Daubert* admissibility standard, encouraging judges to apply this standard with confidence and ameliorating judges’ tendency to revert back to the comfort of *Frye* and other proxy factors for trustworthiness in their admissibility decisions.

The recent shift in the admissibility standard is of special interest in employment discrimination cases, where reliance on social framework testimony is common.⁸⁰ Just a few years before *Daubert*, the Court decided *Price Waterhouse v. Hopkins*, raising awareness of the potential for this testimony in employment discrimination cases.⁸¹ Along with *Price*

77. 222 F.R.D. 189, 192 (2004).

78. See *Thomas & Thomas Rodmakers, Inc. v. Newport Adhesives and Composites, Inc.*, 209 F.R.D. 159, 162-63 (C.D. Cal. 2002) (“It is clear to the Court that a lower *Daubert* standard should be employed at this stage of the proceedings. Courts have declined to engage in a *Daubert* analysis at the class certification stage of an action on the ground that an inquiry into the admissibility of the proposed expert testimony under *Daubert* would be an inappropriate consideration of the merits of the plaintiff’s claims.”); *O’Connor v. Boeing N. Am., Inc.*, 184 F.R.D. 311, 321 n.7 (C.D. Cal. 1998) (“At [the class certification] stage of the litigation . . . an inquiry into the admissibility of Plaintiffs’ proposed expert testimony as set forth in *Daubert* would be inappropriate.”); *In re Visa Check/Mastermoney Antitrust Litig.*, 192 F.R.D. 68, 76-77 (E.D.N.Y. 2000), *aff’d*, 280 F.3d 124 (2d Cir. 2001) (“Although there is a role for a *Daubert* test here, it is a limited one, tailored to the purpose for which the expert opinion is offered. The question is not, therefore, whether a jury at trial should be permitted to rely on [the expert] report to find facts as to liability, but rather whether I may utilize it in deciding whether the requisites of Rule 23 have been met.”). Perhaps the Ninth Circuit will address this issue in the appeal of *Dukes*.

79. The Court’s decisions and the 2000 amendment have led judges to conflate reliability with internal validity and “fit” with external validity. See *infra* Part V.A.2.

80. See Bielby, *Can I Get a Witness?*, *supra* note 16, at 399-400.

81. See 490 U.S. 228 (1989).

Waterhouse, the changing patterns of discrimination in recent decades have further encouraged plaintiffs to rely on social framework experts to identify and explain the operation of subtle and unconscious forms of discrimination in the workplace.⁸²

III.

THE USE OF SOCIAL FRAMEWORK EXPERTS IN EMPLOYMENT DISCRIMINATION CASES

Social framework experts have provided plaintiffs with valuable testimony about gender, race, and age stereotyping and bias, discriminatory management practices, and harassment. Part III distinguishes social framework experts from employment practices experts and demonstrates how social framework testimony may prove helpful in showing employment discrimination.

A. Distinguishing Social Framework Experts from Employment Practices Experts

Social scientists offer specialized knowledge in a number of areas relevant to employment discrimination cases. For example, plaintiffs often retain research professors to present testimony linking statistical disparities within the workforce to discrimination⁸³ or mental health professionals to illustrate the relationship between negative events at the workplace and emotional distress.⁸⁴

This Comment primarily focuses on testimony provided by social framework experts on behalf of plaintiffs in employment discrimination cases. In a recent article, Dr. Bielby describes what a social framework analysis entails:

[T]he expert reviews testimony, documents, and other quantitative and qualitative information about a case in order to draw conclusions about how extant social science theory and research applies to the specific

82. See Tristin K. Green, *Discrimination in Workplace Dynamics: Toward a Structural Account of Disparate Treatment Theory*, 38 HARV. C.R.-C.L. L. REV. 91, 91 (2003). Green states:

As traditional social norms permitting overt racism and segregation give way to a modern norm of egalitarianism, and as well-defined, hierarchical, bureaucratic structures delineating clear paths for advancement within institutions give way to a globalized workplace of flexible governance and movement between institutions, discrimination often operates in the workplace today less as a blanket policy or discrete, identifiable decision to exclude than as a perpetual tug on opportunity and advancement. It often takes form in a fluid process of social interaction, perception, evaluation, and disbursement of opportunity. It creeps into everyday impressions of worth and assignment of merit on the job, lurking constantly behind even the most honest belief in equality, perpetuating the very injustice that we decry.

Id.

83. Jansonius & Gould, *supra* note 37, at 287.

84. *Id.*

circumstances of the organizational setting where discrimination is alleged to have occurred. She or he analyzes the specific features of the organization's policies and practices and evaluates them against what that social science scholarship has shown to be factors that create and sustain bias and those that minimize bias.⁸⁵

While social framework evidence commonly is used in employment discrimination cases, it first appeared outside the employment context, in cases concerning eyewitness identification, assessments of dangerousness for purposes of sentencing, battered woman syndrome, and sexual victimization.⁸⁶

Because social framework evidence often gets lumped together with other expert evidence in employment discrimination cases, this Comment draws a distinction between social framework evidence and employment practices evidence. Employment practices experts⁸⁷ examine the internal personnel policies and practices that operate generally in the workplace or that govern a particular industry or field.⁸⁸ Unlike social framework experts, employment practices experts generally do not rely on a body of social science research to reach their conclusions; instead, most employment practices experts look to their practical experience with the workplace, often with a specific industry: "The employment practices expert may be a lawyer who has worked for years in the employment field, an experienced personnel manager, or a general manager who has extensive experience in personnel matters. Some experts are also consultants in the personnel field, whereas others come from academia."⁸⁹ In part because

85. Bielby, *Can I Get a Witness?*, *supra* note 16, at 390.

86. See Walker & Monahan, *supra* note 15, at 563.

87. These experts may also be referred to as "human resources" or "management practices" experts.

88. Edward T. Ellis & Mona Ross, *The Admissibility of Expert Opinion Testimony about Human Resources Practices, Sexual Stereotyping and Sexual Behavior in the Workplace*, in CURRENT DEVELOPMENTS IN EMPLOYMENT LAW 703, 712 (Am. Law Inst. ed. 1999), available at SE05 ALI-ABA 703 (Westlaw).

89. Richard G. Moon & Jennifer Humphreys, *Litigating Sexual Harassment Cases, The Same Old Questions but Some of the Rules are Different: The Use of the Employment Practices Expert*, 606 PRAC. L. INST. LITIG. 413, 423 (1999). Employment practices experts often appear in sexual harassment cases to "testify as to the employer's training, policies and procedures for handling sexual harassment complaints in the workplace." *Id.* at 424. For example, former police officer turned sociology professor Dr. Stephen Leinen testified in both *Valentin v. New York City* and *Katt v. City of New York* about police culture and sexual harassment. *Valentin v. New York City*, No. 94 CV 3911(CLP), 1997 WL 33323099 (E.D.N.Y. Sept. 9, 1997); *Katt v. City of New York*, 151 F. Supp. 2d 313 (S.D.N.Y. 2001). The judges in both cases approved Dr. Leinen's qualifications because of his extensive professional and academic experience with police culture and harassment in police departments. See 1997 WL 33323099, at *17; 151 F. Supp. at 354-55. In *Katt*, Dr. Leinen had greater flexibility to offer his opinions because the judge was familiar with the field of criminology and recognized the soundness of Dr. Leinen's methodological approach of relying in part on "anecdotal" research, such as interviews. See 151 F. Supp. at 357. However, in *Valentin*, the judge excluded Dr. Leinen's testimony about anecdotal evidence of previous unrelated incidents because the evidence was not helpful to the jury. See 1997 WL

employment practices experts commonly draw conclusions that are not grounded in scientific research, their testimony is especially susceptible to exclusion for lacking a foundation in specialized knowledge.⁹⁰

In contrast, social framework experts draw from the fields of social cognitive psychology and industrial-organizational psychology to develop their evidence. Social cognitive psychology operates at the micro-level, with a focus on cognitive processes that result in bias and stereotyping.⁹¹ Industrial-organizational psychology operates at the macro-level, with a focus on hiring standards, methods of evaluating employee qualifications and performance, and pay systems.⁹² Dr. Bielby incorporated findings from both of these strands of psychology to draw the conclusion in his expert report for *Dukes v. Wal-Mart Stores, Inc.* that Wal-Mart discriminates against its female employees.⁹³ The remainder of Part III provides a brief overview of the social science research on which a social framework analysis is constructed, offers a demonstration of how Dr. Bielby uses this research to develop a social framework analysis for litigation, and concludes with an introduction to some cases in which a social framework analysis proved helpful in showing employment discrimination.

33323099, at *19-20. Exclusion on this basis—because the testimony does not “fit” the issue in question—is another way of stating that the testimony lacked “external validity” to resolve the issue in question. See I FAIGMAN, *supra* note 32, § 1-3.2.

90. In variations of this trend, judges have excluded this testimony because it was not helpful to the jury, because the expert lacked the appropriate qualifications to offer such testimony, and because the testimony consisted of opinions rather than objective observations. See *Wilson v. Muckala*, 303 F.3d 1207, 1218-19 (10th Cir. 2002) (excluding human resources expert’s testimony about the defendant’s response plan to incidents of sexual harassment because this evidence was within the jury’s common knowledge, and therefore, the testimony would not have been helpful to the jury); *Brink v. Union Carbide Corp.*, 41 F. Supp. 2d 402, 405 (S.D.N.Y. 1997) (excluding expert’s testimony because his limited background in corporate human resources did not bestow upon him the qualifications to evaluate age discrimination claim with specialized knowledge helpful to the jury); *Voisine v. Danzig*, No. 98-340-P-DMC, 1999 WL 33117132 (D. Me. Oct. 26, 1999) (excluding expert testimony on sexual harassment because plaintiff failed to offer any evidence of the expert’s qualifications and specialized knowledge); *EEOC v. Ind. Bell Tel. Co.*, No. IP 95-0217 C-M/S, 1997 WL 1098409 (S.D. Ind. Dec. 22, 1998) (excluding employment law educator and practitioner’s testimony about sexual harassment prevention policies because her testimony appeared to consist of conjectures and legal conclusions); *Oubre v. Entergy Operations, Inc.*, No. Civ.A. 95-3168, 1998 WL 684154, at *3 (E.D. La. Sept. 24, 1998) (excluding psychologist’s testimony in age discrimination case that the plaintiff was constructively discharged because the testimony constituted a legal conclusion that invaded the province of the jury).

91. See Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1187 (1995).

92. See Michael M. Harris, *The Industrial Psychologist as Expert Witness*, EXPERT WITNESS J., Oct. 1999, at 3, 4.

93. See Bielby, *supra* note 7.

*B. The Body of Social Science Research on which
Social Framework Evidence is Based*

Social cognitive psychology is based on the principle that intergroup bias that leads to discrimination results from “normal” cognitive processes related to categorization, rather than from a “special” motivation to stereotype.⁹⁴ As Professor Linda Krieger details, the social cognitive approach to discrimination involves three assumptions: (1) stereotyping is a form of categorization employed by all people, (2) stereotypes function as implicit theories that bias intergroup judgment and decision making, and (3) stereotypes operate on a level unconscious to the decisionmaker.⁹⁵ These assumptions support the fundamental premise of social cognition theory— “that cognitive structures and processes involved in categorization and information processing can in and of themselves result in stereotyping and other forms of biased intergroup judgment.”⁹⁶

In short, the tendency to invoke stereotypes in making judgments is automatic. When these stereotypes affect organizational policies and practices, such as personnel decisions, employment discrimination can result.⁹⁷ Stereotyping especially shapes employment decision making in the absence of measures to hold the decision-maker accountable, such as specified decision-making criteria.⁹⁸ Accordingly, stereotyping is also particularly influential when employment decision making is based on informal, arbitrary, and subjective factors.⁹⁹

Psychologists use a number of methods of analysis when developing and substantiating theories that explain discrimination. Traditionally, they have conducted surveys to determine the content and extent of stereotypes and have observed the effects of stereotypes on decision making in “judgment task” exercises¹⁰⁰ and in “simulation paradigms,” such as job interviews.¹⁰¹ More recently, they have relied on response latency measures, which prevent a participant from hiding biases by linking the length of time a participant takes to respond to certain questions to the

94. Krieger, *supra* note 91, at 1187.

95. *Id.* at 1187-88.

96. *Id.* at 1187 (footnote omitted).

97. See Bielby, *Can I Get a Witness?*, *supra* note 16, at 379.

98. *Id.* at 380.

99. *Id.* at 381.

100. A typical “judgment task” exercise is to present participants with information about an individual, tell the participants that the individual is either a man or woman, and observe how stereotyping affects the participants’ judgments about the individual. See 2 FAIGMAN, *supra* note 32, § 18-2.3.1.

101. *Id.*

accessibility of stereotypes in that participant's mind.¹⁰² Perhaps the most important method psychologists use as a basis for creating a social framework is the "meta-analysis," the procedure by which existing studies in the field are compiled and integrated in order to summarize their results.¹⁰³ This body of research has been conducted over many decades, generating considerable information about how stereotyping and bias perpetuate workplace inequalities.¹⁰⁴

C. *The Advantages of Social Framework Evidence over Original Research in a Litigation Context*

Social science experts rarely use materials from a case in litigation to conduct original research for the purpose of testing for stereotyping and cognitive bias.¹⁰⁵ Rather, they conduct a social framework analysis, which has distinct advantages over original research:

[A] social framework analysis is not designed to test theories, nor are the methods of evaluation identical to those that an expert on cognitive bias or organizational inequality would employ in substantively motivated research [A]n expert on stereotyping, cognitive bias, and organizational inequality with a thorough command of the relevant scientific literature and access to [materials produced in discovery] will often be able to reach conclusions that meet the standard of a "reasonable degree of scientific certainty" with a much more concrete empirical grounding than would normally be possible with a case study method of social research.¹⁰⁶

Social framework analysis has been recognized by the courts as a sound methodology in a variety of litigation contexts, including employment discrimination cases.¹⁰⁷

102. *Id.* § 18-2.3.2. For example, social cognitive psychologists can infer that a participant has stereotypes if he or she responds more quickly to situations consistent with stereotypes and less quickly to situations inconsistent with stereotypes. *Id.*

103. *Id.* § 18-2.3.3.

104. Bielby, *Can I Get a Witness?*, *supra* note 16, at 383.

105. Bielby, *Can I Get a Witness?*, *supra* note 16, at 390. In the cases where social science experts did try to conduct original research for trial, the expert testimony was often excluded. *EEOC v. Dial Corp.*, No. CIV.A. 99-C-3356, 2002 WL 31061088 (N.D. Ill. Sept. 17, 2002) (excluding testimony based on plaintiff expert's sexual harassment survey because of its small sample of employees heavily represented by class members, its failure to focus on the time frame relevant for litigation, and its biased questions, which primed respondents to express negative views of the defendant); *Fisher v. Vassar Coll.*, 70 F.3d 1420, 1447-48 (2d. Cir. 1995) (excluding expert testimony based on survey with flawed methodology); *Indianapolis Minority Contractors Ass'n, Inc. v. Wiley*, No. IP 94-1175-C-T/G, 1998 WL 1988826 (S.D. Ind. May 13, 1998) (excluding expert testimony on affirmative action because the testimony was based in part on flawed methodology, including the lack of defined criteria and the reliance on unsubstantiated assumptions).

106. Bielby, *Can I Get a Witness?*, *supra* note 16, at 391.

107. *Id.* at 390 & n.29.

Dr. Bielby explained in his expert report for *Dukes v. Wal-Mart Stores, Inc.* how he conducted a social framework analysis based on a body of social science research.¹⁰⁸ Dr. Bielby first introduced the research on which his conclusions were based, which included experiments in controlled laboratory settings, ethnographies and case studies in “real world” organizations, surveys with representative samples of workers and employees, and historical studies based on archival materials.¹⁰⁹ Using the research he compiled as a backdrop, he reviewed the deposition testimony of Wal-Mart managers who created and implemented personnel policies and who made decisions about compensation, hiring, promotion, job assignment and other personnel matters.¹¹⁰ He also examined documents used as exhibits in these depositions, such as organizational charts, correspondence, memos, reports, and presentations relating to the culture and history of the company, personnel policy and practice, diversity, and equal employment opportunity issues.¹¹¹ Finally, he evaluated the statistical analysis of gender patterns in the Wal-Mart workforce, which was prepared by the plaintiffs’ statistical experts.¹¹²

Price Waterhouse v. Hopkins marked the Supreme Court’s first consideration of expert psychological testimony on stereotyping in a gender discrimination case.¹¹³ Social psychologist Dr. Susan Fiske testified that the partners of Price Waterhouse relied on gender stereotypes in their partnership selection process, specifically in their professional evaluations of the plaintiff.¹¹⁴ While Justice Brennan’s plurality opinion indicated that this testimony may have been only “icing on Hopkins’ cake,”¹¹⁵ it also implicitly acknowledged the importance of Dr. Fiske’s testimony by

108. See Bielby, *supra* note 7. The importance of Dr. Bielby’s explanation of his methodology is especially salient after observing *Collier v. Bradley University*, where the judge excluded the plaintiff expert’s testimony because she could not explain her methodology. See 113 F. Supp. 2d 1235, 1244-45 (C.D. Ill. 2000).

109. See Bielby, *supra* note 7.

110. See *id.*

111. *Id.*

112. *Id.*

113. See Susan Bisom-Rapp, *Of Motives and Maleness: A Critical View of Mixed Motive Doctrine in Title VII Sex Discrimination Cases*, 1995 UTAH L. REV. 1029, 1040; see *Price Waterhouse v. Hopkins*, 490 U.S. 228, 255 (1989).

114. *Price Waterhouse*, 490 U.S. at 255.

115. *Id.* at 256. Justice Brennan stated:

It takes no special training to discern sex stereotyping in a description of an aggressive female employee as requiring “a course at charm school.” Nor . . . does it require expertise in psychology to know that, if an employee’s flawed “interpersonal skills” can be corrected by a soft-hued suit or a new shade of lipstick, perhaps it is the employee’s sex and not her interpersonal skills that has drawn the criticism.

Id.

concluding that gender stereotyping played a part in the partners' ultimate employment decision.¹¹⁶

Butler v. Home Depot, Inc. provides a classic example of how plaintiffs can rely on social science experts to create a social framework to explain employment discrimination that occurs on a broad scale.¹¹⁷ Among plaintiffs' experts in *Butler* were: (1) Dr. Mary Gentile, seeking to testify that Home Depot had failed to include several features that would improve the design of its gender diversity program and that Home Depot had failed to accomplish its diversity goals on its stated timeframe;¹¹⁸ (2) Dr. Fiske, seeking to testify about gender stereotyping, the organizational circumstances which facilitate such stereotypes, and Home Depot's failure to implement specific corrective steps to address the disparate impact of its subjective decision-making processes on women;¹¹⁹ and (3) Dr. Bielby, seeking to demonstrate how Home Depot's arbitrary, subjective, and ambiguous decision-making criteria led to gender discrimination and hindered the professional advancement of women at Home Depot.¹²⁰ The court admitted the testimony of all three experts, finding that each contributed a unique perspective to explain the organizational structure and culture of the employer and the failure of the employer to prevent discrimination through its practices and policies.¹²¹

IV.

THE DIVERGENCE OF LAW AND SOCIAL SCIENCE IN THE *DAUBERT* STANDARD

Having introduced the *Daubert* admissibility standard and its general application to social framework evidence, this Comment moves on to inquire whether federal judges are equipped for the tasks of recognizing weaknesses in social science testimony and of interpreting and explaining this testimony accurately. Part IV examines whether judges have the tools to handle the enormous responsibility given them by the *Daubert* admissibility standard, and it provides examples of when judges may have fallen short in their admissibility decisions on social framework testimony.

116. *See id.*

117. *See* 984 F. Supp. 1257 (1997).

118. *Id.* at 1261-62.

119. *Id.* at 1262.

120. *Id.* at 1265.

121. *Id.* at 1262, 1265.

A. *Examining Whether Judges are Equipped
to Recognize Flawed Research*

The decision in *Daubert* to expand the analysis of expert testimony beyond what is “generally accepted” empowers judges to embark on a more flexible investigation of the soundness of the methodology of proposed evidence.¹²² However, this more active role carries with it risks, as judges purport not to defer to the judgments of relevant scientific communities, as was allowed in *Frye*.¹²³ This subpart demonstrates that the assumption of the Court in *Daubert* that trial judges would be effective gatekeepers rested on the false notions that (1) judges would understand the four *Daubert* factors provided by the Court, (2) judges could use the four *Daubert* factors to assess the soundness of the methodology of expert evidence, (3) judges would be capable of recognizing flawed research, and (4) if judges did admit untrustworthy testimony, procedural safeguards would prevent this testimony from receiving much weight.

In a survey of trial judges, ninety-one percent stated that the four factors set out by the Court in *Daubert* were useful in assessing scientific testimony.¹²⁴ Yet although over ninety percent of these judges understood the peer review/publication and general acceptance factors, less than ten percent could clearly define the testability and error rate factors.¹²⁵

To compound this problem, while the Court asserted that the *Daubert* factors would be helpful in assessing the “scientific validity” of expert testimony, the factors only serve this function in part.¹²⁶ First, testability and error rate, the first and third factors, focus on *reliability*, but not on *validity*, an equal or more important measure of the soundness of the methodology.¹²⁷ Second, peer review/publication and general acceptance, the second and fourth factors, are merely proxies for trustworthiness, not direct measures of the soundness of the methodology.¹²⁸ Thus, the *Daubert*

122. Joseph Sanders et al., *Legal Perceptions of Science and Expert Knowledge*, 8 PSYCHOL. PUB. POL'Y & L. 139, 142 (2002); David L. Faigman, *The Evidentiary Status of Social Science under Daubert: Is it “Scientific,” “Technical,” or “Other” Knowledge?*, 1 PSYCHOL. PUB. POL'Y & L. 960, 979 (1995).

123. Sanders, *supra* note 122, at 142; cf. Faigman, *supra* note 122, at 979.

124. Sanders, *supra* note 122, at 152.

125. *Id.*

126. See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 592 (1993).

127. Faigman, *supra* note 122, at 965. Testability, or falsifiability, as defined by Karl Popper, provides that “a statement or theory is . . . falsifiable if and only if there exists at least one potential falsifier—at least one possible basic statement that conflicts with it logically.” *Id.* at 966 (quoting KARL POPPER, REALISM, AND THE AIM OF SCIENCE, POSTSCRIPT TO THE LOGIC OF SCIENTIFIC DISCOVERY, at xx (W. Bartley, III ed., 1983)). Even if judges did understand error rate, this factor likely would not be prominent in evaluating testimony presented by social framework experts. See Faigman, *supra* note 122, at 968.

128. *Id.* at 969. The peer review/publication and general acceptance factors have the same weaknesses as the *Frye* standard. For example, if reviewers or journals employ lenient standards when

factors may actually lead judges away from evaluating the soundness of the methodology of expert evidence and back to the deceptive comfort of the *Frye* standard.

Also worrisome is a study conducted by psychologists Margaret Bull Kovera, Melissa Russano, and Bradley McAuliff that indicates that judges cannot identify the methodological shortcomings in an expert's research.¹²⁹ Kovera et al. provided one group of judges a prize-winning study examining the influence of sexually suggestive materials on subsequent male-female interactions and another group the same study modified to contain a methodological flaw.¹³⁰ Kovera et al. found that the judges' ratings of the study's quality did not differ based on the study's methodological quality.¹³¹ Furthermore, while judges with some scientific training rated the legitimate study more positively than judges without any scientific training, the latter group rated one of the flawed studies more positively, actually preferring that study *because* of the flaw.¹³² The flaw in this study was that it included a confound; instead of a single female research assistant interacting with all of the male participants, one female interacted with the males exposed to the sexual material while another female interacted with the males not exposed to the sexual material.¹³³ The judges with no scientific training apparently believed that a study using two research assistants was more sound than a study using one assistant.¹³⁴

Lastly, if judges admit flawed research—and the Kovera study suggests that they often do—then the role of procedural safeguards in reducing the weight of this testimony takes on greater significance. The Court in *Daubert* expressed a high level of confidence in these procedural protections: “Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.”¹³⁵

evaluating an expert's work, the peer review/publication factor may not lead to the prohibition of junk science. Likewise, the general acceptance standard is only as dependable as the field surveyed. *Id.* While these concerns may not be as significant for social framework testimony, which rests on an established body of research, they loom larger for employment practices testimony, which sometimes rests on a shakier foundation.

129. Margaret Bull Kovera et al., *Assessment of the Commonsense Psychology Underlying Daubert*, 8 PSYCHOL. PUB. POL'Y & L. 180, 186 (2002).

130. *Id.*

131. *Id.*

132. *Id.* The confound was that instead of a single research assistant interacting with all of the men, one research assistant always interacted with the men exposed to the sexual material while another research assistant interacted with the men not exposed to the sexual material. *Id.* The judges with no scientific training apparently believed that a study using two research assistants was more reliable than a study using one assistant. *Id.*

133. *Id.*

134. *Id.*

135. 509 U.S. 579, 596 (1993).

However, attorneys may not highlight methodological weaknesses in cross-examination if they do not recognize these weaknesses.¹³⁶ In a survey of employment lawyers, few mentioned challenging the validity of the evidence as one of their cross-examination strategies.¹³⁷ Those attorneys who observed a study with validity devoted as much attention in their cross-examinations to whether the study measured what it purported to measure as did attorneys who observed a study without validity.¹³⁸ Consistent with previous findings in this subpart, the attorneys tended to hone in on the expert's qualifications or potential sources of bias (e.g., payment for testimony, disproportionate appearances on behalf of plaintiffs), rather than on the methodological quality of the expert's research.¹³⁹

Moreover, a preliminary study indicates that even a strong cross-examination does not appear to expose methodological weaknesses in expert evidence, at least not to jurors.¹⁴⁰ In this study of juror evaluations of expert evidence in a hostile work environment case, Kovera et al. investigated whether jurors who listened to a cross-examination designed to reveal the methodological flaws in expert evidence could better evaluate the soundness of the evidence.¹⁴¹ In the "trial," an expert presented testimony about a study that she had conducted.¹⁴² In some versions of the "trial," the expert's study had good construct validity (i.e., it contained multiple measures of the construct under investigation), while in other versions, the expert's study had poor construct validity (i.e., it contained a single, self-report measure of the construct under investigation).¹⁴³ Kovera et al. also varied the scientific sophistication of the cross-examination (i.e., whether it attacked the construct validity of the expert's study).¹⁴⁴ Kovera et al. concluded that the nature of the cross-examination of the expert did not affect jurors' evaluations of the expert evidence.¹⁴⁵

This subpart casts doubt on the Court's assumption in *Daubert* that judges can apply the *Daubert* admissibility standard effectively to examine social science evidence. Further, it calls into question the Court's stopgap measure, for if judges erroneously allow methodologically flawed evidence, the procedural safeguard of cross-examination may not have the intended impact of reducing the weight of this evidence.

136. Kovera et al., *supra* note 129, at 192.

137. *Id.*

138. *Id.*

139. *Id.*

140. *Id.* at 193.

141. *Id.*

142. *Id.*

143. *Id.*

144. *Id.*

145. *Id.*

*B. Problematic Interpretations and Explanations of
Social Framework Testimony*

The few published decisions that examine social framework testimony provide only limited insight into how federal judges perform their gatekeeping role with regard to this evidence. However, these decisions bring to light a few shortcomings in judges' admissibility decisions on social science testimony. These shortcomings appear to result from a lack of scientific training and/or a lack of confidence in applying the current admissibility standard. They include (1) interpreting social framework evidence inconsistently with the scope of the findings of this evidence, (2) regarding as common knowledge specialized knowledge that may offer insights helpful to the jury, (3) relying solely on case precedent and the reputation of the expert to evaluate expert testimony, (4) treating the soundness of the methodology as an issue of weight instead of admissibility, and (5) failing to substantiate the decision to admit or exclude testimony. The Comment will discuss these five shortcomings in turn.

1. Interpreting Social Framework Evidence Inconsistently With the Scope of the Findings of This Evidence

Judges' interpretations of social framework evidence may be inconsistent with the scope of the findings of this evidence.¹⁴⁶ In *Price Waterhouse*,¹⁴⁷ Justice Brennan's analysis of Dr. Fiske's testimony reflected his limited comprehension of the way stereotyping and bias operate.¹⁴⁸ Consistent with social cognition theory, Justice Brennan recognized that unconscious stereotyping of women may have triggered some partners' strong reactions to the plaintiff, especially when partners who barely knew the plaintiff indicated that she was "universally disliked."¹⁴⁹ On the other hand, Justice Brennan reached flawed conclusions based on mixed motive

146. Similarly, social framework testimony may be excluded improperly when a judge takes an overly restrictive approach to antidiscrimination law. In *Naeem v. McKesson Drug Co.*, involving gender discrimination and retaliation claims, the judge allowed the plaintiff's human resources expert, Dr. William Anthony, to testify about whether the defendant corporation followed its own policies, while barring him from testifying about deficiencies in the defendant's employment policies that might lead to discrimination. No. 95C5425, 2001 WL 1141803, at *2 (N.D. Ill. Aug. 14, 2001). The court reasoned that the former testimony would be helpful to the trier of fact, but that the latter testimony did not fit: "The law is clear that an employer's conduct . . . need only be non-discriminatory. Therefore, the fact that [the defendant] acted in a manner which facilitated discrimination, retaliation or emotional distress is not an appropriate subject of expert testimony." See *id.* The court's approach is problematic because whether the defendant's employment practices "facilitated" discrimination would appear to be quite relevant and helpful to the jury to determine whether the defendant discriminated.

147. 490 U.S. 228 (1989).

148. See Bisom-Rapp, *supra* note 113, at 1042-43.

149. *Id.* at 1041-42; see *Price Waterhouse*, 490 U.S. at 235.

theory,¹⁵⁰ drawing a sharp distinction between permissible categorizations (of the plaintiff's interpersonal skills) and discriminatory categorizations (of the plaintiff's gender).¹⁵¹ Justice Brennan failed to realize that the former categorizations may *also* have been tainted by gender stereotypes, even if this form of discrimination was less overt and did not involve discriminatory intent.¹⁵² Therefore, Justice Brennan ignored a prevailing principle of social cognitive psychology—that stereotypes operate on a level unconscious to the decisionmaker—when he attempted to fit Dr. Fiske's testimony into a rigid legal framework.

2. *Regarding as Common Knowledge Specialized Knowledge Helpful to the Jury*

At one time, some judges regarded social framework testimony on stereotyping and bias as "common knowledge," not as specialized knowledge helpful to the jury. With Professor Krieger's work demonstrating that stereotyping and bias operate on a level unconscious to the decisionmaker, the likelihood is low that the way discrimination manifests itself is within the common knowledge of a jury.¹⁵³ Although most judges now recognize that social framework testimony on stereotyping and bias is specialized knowledge, the specter of this problem remains, because in the future, judges may be tempted to exclude social framework testimony based on new advances in social science.

In *Smith v. Colorado Interstate Gas Co.*, decided a few years after *Price Waterhouse*, the judge overlooked the role of social cognitive psychology in explaining how stereotypes facilitate discrimination, treating discrimination as an area of common knowledge.¹⁵⁴ In *Smith*, the court excluded both Dr. John M. Bermudez, who prepared testimony that plaintiff's race and gender substantially motivated her supervisor's decision to terminate her, and Dr. Lenore E.A. Walker, who prepared testimony about the role of racial and gender stereotypes, biases, and prejudice in the defendant's discrimination against the plaintiff.¹⁵⁵ The court provided two reasons for barring this testimony—that race and gender discrimination "are issues an average person can evaluate and understand without the assistance of an expert," and that by stating that discrimination played a role in the

150. Mixed motive theory, initiated in *Price Waterhouse*, allows the plaintiff to shift the burden of proof to the defendant by showing that her group status "played a role" in the employment decision. Krieger, *supra* note 91, at 1220.

151. See Bisom-Rapp, *supra* note 113, at 1043.

152. See *id.* at 1044.

153. See Krieger, *supra* note 91, at 1187.

154. See 794 F. Supp. 1035, 1044 (D. Colo. 1992).

155. *Id.* at 1044.

termination, both experts set forth legal conclusions.¹⁵⁶ While the latter reason is legitimate (at least to justify prohibiting some testimony), the former reason raises questions, especially in light of Dr. Fiske's testimony in *Price Waterhouse*, which demonstrated the specialized knowledge of social cognitive experts and challenged the notion that the "average person" can recognize discrimination.¹⁵⁷ The court in *Smith* failed to grasp the power of social framework testimony to explain employment discrimination.

3. *Relying Solely On Case Precedent And The Expert's Reputation*

Because of the judicial system's respect for precedent, judges might yield to earlier admissibility decisions without conducting a fresh analysis of the expert testimony in the case at hand, especially when dealing with an expert who has testified previously. For example, Dr. Benson Rosen testified in the age discrimination cases *Flavel v. Svedala Industries, Inc.*¹⁵⁸ and *Hurst v. F.W. Woolworth Co.*¹⁵⁹ Because Dr. Rosen had been qualified to give comparable testimony in *Flavel*, the judge in *Hurst* appeared to ignore his gatekeeping responsibility of evaluating the soundness of Dr. Rosen's methodology.¹⁶⁰ Despite the similarity of Dr. Rosen's testimony in both cases, the judge's approach in *Hurst* discounted the possibilities that the facts in *Flavel* and *Hurst* might have been different enough so as to influence the admissibility decisions, that Dr. Rosen's methodology in *Hurst* might have been flawed even if it was not in *Flavel*, or that the judge in *Flavel* did not examine Dr. Rosen's methodology properly.¹⁶¹ Any of

156. *Id.*

157. *Cf.* *Lipsett v. Univ. of Puerto Rico*, 740 F. Supp. 921 (D.P.R. 1990). In *Lipsett*, the court excluded testimony from a social worker and a social psychologist that the employer's workplace was "so heavily charged with sexism that it was intimidating, hostile, and offensive" because it believed that the jury could understand everyday common occurrence and attitudes toward sexual matters. *Id.* at 925 (quoting *Andrews v. City of Philadelphia*, 895 F.2d 1469, 1482 (1990)).

158. 875 F. Supp. 550 (E.D. Wis. 1994).

159. No. 95 CIV. 6584(CSH), 1997 WL 685341 (S.D.N.Y. Nov. 3, 1997).

160. *Id.* at *1-2.

161. A similar pattern occurred in the hostile work environment cases *Robinson v. Jacksonville Shipyards, Inc.* and *Jenson v. Eveleth Taconite Co.*, both of which predated *Daubert*. The plaintiffs in both cases presented experts on gender stereotyping, but the defendants—and perhaps as a result, the judges—failed to question the soundness of the methodology of the expert testimony. See *Robinson*, 760 F. Supp. 1486 (M.D. Fla. 1991); *Jenson*, 824 F. Supp. 847 (D. Minn. 1993). Dr. Fiske (in *Robinson*) and Dr. Eugene Borgida (in *Jenson*) reviewed depositions and other documents, identified the preconditions that enhance stereotyping in the workplace, and concluded that those preconditions were present in the workplaces of the defendants. See *Robinson*, 760 F. Supp. at 1502-05; *Jenson*, 824 F. Supp. at 880-83. In *Robinson*, the judge concluded that Dr. Fiske had provided a "sound, credible theoretical framework" from which to determine that the defendants had a sexually hostile work environment abusive toward women. See *Robinson*, 760 F. Supp. at 1505. *Jenson* cited *Robinson* in its assessment of gender stereotyping and mirrored *Robinson* in its analysis of Dr. Borgida's testimony, going so far as to use identical language. See *Jenson*, 824 F. Supp. at 883.

these possibilities might have led the judge in *Hurst* to exclude Dr. Rosen's testimony, yet the judge did not explain whether he analyzed the soundness of Dr. Rosen's methodology.

4. *Treating Sound Methodology as an Issue of Weight and Not Admissibility*

At times, judges have treated the soundness of the methodology as an issue of weight instead of as an issue of admissibility. Such an approach may be appropriate at the class certification stage, when the judge need only decide whether the class meets the requirements of Rule 23,¹⁶² but to ensure the soundness of the methodology, judges still must use the *Daubert* standard in an admissibility hearing before trial.

In an individual trial, the judge must always apply the full *Daubert* admissibility standard. In *Mukhtar v. California State University*, the Ninth Circuit overturned the district court because the district judge did not assess the methodology of the expert testimony on racial discrimination.¹⁶³ When denying the defendant's motion in limine, the district judge stated, "[w]ell . . . I would prefer that none of [the experts] express their own opinion about whether this decision [not to grant tenure] was right or wrong. But if any of them are going to, then I guess all of them have to." With this statement, the district judge indicated that he was not troubled by the possibility that inappropriate testimony might be admitted.¹⁶⁴ Presumably, the district judge assumed that admitting inappropriate testimony would have little effect on the outcome because it would be given little weight by the jury. Research suggests, however, that procedural safeguards such as cross-examination may not work as well as judges expect;¹⁶⁵ thus, the district judge's decision to admit this testimony was based on a faulty premise. The district judge's failure to review the soundness of the methodology of the expert testimony tainted the district court's decision.

5. *Failing to Substantiate the Decision to Admit or Exclude Testimony*

Finally, judges simply may fail to substantiate their admissibility decisions. In the racial discrimination class action *Robinson v. Metro-North Commuter Railroad Co.*, the judge casually dismissed Dr. Bielby's testimony, even though the testimony closely resembled what had been accepted in *Butler v. Home Depot* just a few days earlier.¹⁶⁶ Stating that Dr. Bielby's testimony "consist[ed] on its face of little more than rank

162. See *supra* Part II.G.

163. 299 F.3d 1053 (9th Cir. 2002).

164. See *supra* Part IV.A.

165. *Id.* at 1064.

166. See 175 F.R.D. 46, 48 (S.D.N.Y. 1997).

conclusion and gross speculation,” the judge rejected the testimony as “facially suspect” without applying *Daubert* or Federal Rule of Evidence 702.¹⁶⁷ Because the decision was made in the context of a determination whether the plaintiff class had satisfied the Rule 23 requirement of commonality, the lower *Daubert* standard presumably applied, making the judge’s hasty dismissal of Dr. Bielby’s testimony especially surprising. Regardless, the judge’s failure to explain the admissibility decision left great uncertainty about the grounds for and correctness of the decision.¹⁶⁸

One reason that judges may not engage in an analysis of methodological soundness was revealed in *McKnight v. Dormitory Authority*, where the judge noted that “the factors outlined in *Daubert* are not particularly helpful.”¹⁶⁹ Before the 2000 amendment to Rule 702, judges had little guidance in examining the methodology of expert testimony. Whether judges rely on the guidelines set forth in this amendment largely remains to be seen, as too few decisions have been issued to draw any conclusions yet.¹⁷⁰ At any rate, the 2000 amendment still does not distinguish reliability from validity, casting doubt on whether judges have the necessary tools to examine expert testimony. Part V addresses this deficiency and puts forth other suggestions on how the legal community can establish a more solid foundation on which to examine social science evidence.

V.

BRIDGING THE GAP BETWEEN LAW AND SOCIAL SCIENCE

These indications that judges, not to mention attorneys and jurors, may not be particularly skillful at handling social framework testimony raise the inevitable question: How can the legal community become more sophisticated in its treatment and use of scientific knowledge? Based on existing research and the cases involving admissibility decisions on social framework testimony, this Comment offers three suggestions: (1) improving scientific training for law students, attorneys, and judges, (2) distinguishing reliability and validity in the 2000 amendment to Federal

167. *Id.*

168. *See id.* In fact, appellate courts increasingly require district courts to create a sufficient record of the basis for admissibility decisions. 1 FAIGMAN, *supra* note 32, at § 1-3.1.1.

169. *See* 189 F.R.D. 225, 229 n.2 (N.D.N.Y. 1999).

170. However, in a discouraging sign, judges rarely referred to the 2000 amendment and never used it in a step-by-step approach to decide whether social framework testimony was admissible in the small number of employment discrimination cases decided after the 2000 amendment. In the cases collected for this Comment, just a few were decided after the 2000 amendment, and only *EEOC v. Dial Corp.* included the language of the 2000 amendment. *See Dial*, No. CIV.A. 99-C-3356, 2002 WL 31061088, at *1 (N.D. Ill. Sept. 17, 2002). Perhaps judges have avoided using the 2000 amendment in light of the unhelpfulness of the *Daubert* factors.

Rule of Evidence 702, and (3) providing in the advisory note to Federal Rule of Evidence 702 a recommended approach to the admissibility standard for expert testimony that filters the testimony through a series of steps. While the first suggestion is backed up by the research discussed in Part IV.A, the latter two suggestions hinge on the first and should be viewed as possible avenues for reform. Before implementing them, further research is necessary to consider their effects on litigation and policy outside the context of the employment discrimination cases that were the focus of this Comment.

A. Strengthening Scientific Training for Attorneys and Judges

One of the recommendations of Kovera et al. is to provide better scientific training to law students, attorneys, and especially judges.¹⁷¹ First, they propose requiring an interdisciplinary scientific evidence class in law school, preferably team-taught by scientists and law professors, which would especially focus on guiding attorneys in their research for cross-examining opposing witnesses.¹⁷² Such a class also would prepare future litigators and judges to distinguish studies based on poor methodology from those based on sound methodology.¹⁷³ Second, they propose that those with scientific training produce “white papers” to assist attorneys and judges in recognizing the common methodological pitfalls associated with social science evidence, reference manuals devoted to scientific evidence,¹⁷⁴ and case books that offer some background in the methodology of scientific evidence.¹⁷⁵ However, they point out the potential limitations of training for cases where social science testimony is based less on methodological research and more on personal or work experience.¹⁷⁶ This Comment brings to light this distinction through the discussion of the differences between social framework evidence and employment practices evidence.

Researcher Melvin Mark provides other recommendations for assisting judges in their gatekeeping function. Expanding upon the proposals of other researchers to enlist the aid of court-appointed experts, special masters, or science courts,¹⁷⁷ Mark envisions a nonpartisan research institute

171. Kovera et al., *supra* note 129, at 197-98.

172. *Id.*

173. *See id.* at 198.

174. The most prominent is FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (2d ed. 2000). *See* Kovera et al., *supra* note 129, at 198.

175. Currently, the leading casebook is DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE (2001). *See* Kovera et al., *supra* note 129, at 198.

176. *See id.*

177. *See* Melvin M. Mark, *Social Science Evidence in the Courtroom: Daubert and Beyond?*, 5 PSYCHOL. PUB. POL'Y & L. 175, 189 (1999) (citing Daniel A. Krauss & Bruce D. Sales, *The Problem of "Helpfulness" in Applying Daubert to Expert Testimony: Child Custody Determinations in Family Law as an Exemplar*, 5 PSYCHOL. PUB. POL'Y & L. 78 (1999)).

that would offer general continuing education, literature reviews, and technical assistance for judges dealing with specific evidentiary issues.¹⁷⁸ Mark cautions that this institute only would be necessary if the admissibility threshold rises, forcing judges to consider the methodology of testimony more closely.¹⁷⁹ However, this Comment demonstrates that such an institute may be needed now, because judges already fall back on proxies for trustworthiness in their admissibility decisions, instead of fulfilling the purpose of *Daubert* by taking on their gatekeeping role.

B. Modifying the Language Added in the 2000 Amendment to Federal Rule of Evidence 702

After *Daubert* and the 2000 amendment to Federal Rule of Evidence 702, judges appear to have more tools at their disposal when assessing the methodology of expert testimony. Based on the cases reviewed in this Comment, however, judges rely heavily on the original Rule 702 in their evaluations of social framework testimony, hardly mentioning these other devices, let alone using them to determine the outcome of their analyses. With some slight revisions to language added in the 2000 amendment to Rule 702, judges would have more helpful guidelines to consider when evaluating the methodology of social science testimony, perhaps convincing them to rely more on the rule and less on proxies for trustworthiness.

1. The Current Role of the Four Daubert Factors and the 2000 Amendment

Research suggests that the function of the four *Daubert* factors in judges' assessments of expert testimony—especially the post-*Daubert* role of general acceptance—remains in question. One hypothesis proposes that while judges look at the four *Daubert* factors when analyzing the soundness of the expert's methodology, in their final evaluations, they return to general acceptance as the most dependable factor because this factor is the only one that allows for comparisons with similar evidence already accepted as trustworthy.¹⁸⁰ A second hypothesis assumes that with general acceptance relegated to a minor role by *Daubert*, the factors that require a more exacting analysis, like testability and error rate, contribute to a more exclusive approach to social science testimony.¹⁸¹ The truth may lie

178. Mark, *supra* note 177, at 189.

179. *See id.*

180. Michael C. Polentz, *Post-Daubert Confusion with Expert Testimony*, 36 SANTA CLARA L. REV. 1187, 1206 (1996).

181. Faigman, *supra* note 122, at 970. Perhaps offering support for this hypothesis is a psychological study that indicates that the peer review/publication "proxy" factor does not affect judges' interpretation of social science research. Kovera et al., *supra* note 129, at 187 (citing M.B. Kovera &

somewhere in between, depending on the judge and the type of testimony, as *Kumho* did emphasize the flexibility of the admissibility standard.¹⁸²

Interestingly, based on the post-*Daubert* cases surveyed in this Comment, judges appeared to cite the original Rule 702—but not the *Daubert* factors—as the foundation for their decisions. When they did cite the *Daubert* factors, they referred most often to the “proxy” factors for trustworthiness—peer review/publication and general acceptance—in their evaluations of social framework testimony. Perhaps the reason that judges generally have not depended on the *Daubert* factors is that they have not found these factors helpful. After all, the *Daubert* factors do not assess validity, which is a significant measure of whether a methodology is sound.

Consistent with an approach that may not fulfill the purpose of *Daubert*, judges who mentioned the *Daubert* factors emphasized their flexibility. One researcher notes that in this respect, *Kumho* may have sown more confusion about how judges should go about applying the *Daubert* admissibility standard:

Given the complexity of the gatekeeping task and given the paucity of relevant training for judges, it is perhaps not too cynical to wonder whether judges’ implementation of *Daubert* will make psychologists think instead of the irrational decision makers in the “Dilbert” cartoon and whether the judges’ application of *Kumho Tire* will leave observers tired of asking “how come?”¹⁸³

Without more concrete measures of reliability and validity, this “flexibility” enthusiastically promoted in *Kumho* raises the questions whether judges fall back on the *Frye* general acceptance standard, on other proxies for trustworthiness, or even on factors such as their sociopolitical attitudes in their admissibility decisions.¹⁸⁴

The purpose of the 2000 amendment to Federal Rule of Evidence 702 was to provide a more thorough and structured approach to admissibility decisions¹⁸⁵ without defining a one-size-fits-all procedure for evaluating expert testimony.¹⁸⁶ The advisory committee’s note to Rule 702 stressed that the 2000 amendment is as functional when applied to testimony based

B.D. McAuliff, *The Effects of Peer Review and Evidence Quality on Judge Evaluations of Psychological Science: Are Judges Effective Gatekeepers?*, 85 J. APPLIED PSYCHOL. 574 (2000)).

182. See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 150 (1999).

183. Mark, *supra* note 177, at 188.

184. One study supports the proposition that judges’ sociopolitical attitudes affect whether they give credence to social science research. See Kovera et al., *supra* note 129, at 187 (citing R.E. Redding & N.D. Reppucci, *Effects of Lawyers’ Socio-Political Attitudes on Their Judgments of Social Science in Legal Decision Making*, 23 L. & HUM. BEHAV. 31 (1999)). In other words, a judge may be more likely to admit testimony that is consistent with his or her political orientation. *Id.*

185. KENT SINCLAIR, PRAC. L. INST., TRIAL HANDBOOK ch. 4, art. 7, S153, S154 (2001).

186. *Id.* at S158.

on experience as it is for testimony based on research.¹⁸⁷ However, this Comment suggests that while the 2000 amendment offered more concrete reliability standards, by failing to add validity standards, it did not address the most significant limitation of the four *Daubert* factors.

2. *Revising the 2000 Amendment*

Federal Rule of Evidence 702 includes tests for fit, sound methodology, and qualifications. The original Rule 702 requires that testimony assist the trier of fact (measuring fit), consist of specialized knowledge (measuring the soundness of the methodology generally), and come from an expert with distinctive knowledge, skill, experience, training, or education (measuring qualifications). The 2000 amendment added three additional measures of reliability: (1) testimony must be based upon sufficient facts or data, (2) testimony must be the product of reliable principles and methods, and (3) the witness must apply the principles and methods reliably to the facts of the case.¹⁸⁸

However, even after the 2000 amendment, Rule 702 still lacks specific measures for validity, a factor that clearly affects the soundness of expert testimony. With a few simple revisions, Rule 702 easily could measure two basic forms of validity, internal and external validity. While principles or methods may be reliable, or consistently produced, they may not be internally valid, measuring what they were purported to measure.¹⁸⁹ And while specialized knowledge may be helpful in resolving an issue in dispute, that specialized knowledge may not be externally valid, generalizing to that particular issue.¹⁹⁰ The concepts of internal and external validity, originally formulated by the psychologist Donald Campbell in the 1950s, would better address the concerns the Court intended to tackle in proposing the four *Daubert* factors.¹⁹¹ Therefore, this Comment proposes revising Rule 702 to include the following italicized changes:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable *and internally valid* principles and methods, and (3) *the principles and methods have been applied reliably and have external validity to the facts of the case.*

187. FED. R. EVID. 702 advisory committee's note.

188. See FED. R. EVID. 702.

189. See *supra* note 23.

190. See *id.*

191. See Denbeaux, *supra* note 23, at 39 n.79.

Accompanying this revision with an advisory committee note is vital to ensure that judges, attorneys, and experts understand that the revision's purpose is to restore the scientific usages of the terms "reliability" and "validity" in order to provide distinct tools helpful in assessing whether an expert's methodology is sound.

Note that when combined with the *Daubert* "proxy" factors for trustworthiness, the revised 2000 amendment somewhat resembles the proposed guidelines for evaluating social science research set forth even before *Daubert* by interdisciplinary "law and social science" pioneers John Monahan and Laurens Walker.¹⁹² These guidelines advised the courts to put faith in research that (1) has survived the critical review of the scientific community (like the peer review/publication and general acceptance *Daubert* factors), (2) has employed valid research methods (like this Comment's revision of the 2000 amendment's second subsection, which requires that testimony is the product of reliable and *internally valid* principles and methods), (3) is generalizable to the case at issue (like this Comment's revision of the 2000 amendment's third subsection, which requires that the principles and methods have been applied reliably and have *external validity* to the facts of the case), and (4) is supported by a body of other research (like the 2000 amendment's first subsection, which requires that testimony is based upon sufficient facts or data).¹⁹³

C. *A Recommended Approach to the Admissibility Standard for Expert Testimony*

This Comment recommends an alternative approach to the admissibility standard for expert testimony that could be added to the advisory note to Federal Rule of Evidence 702. This approach establishes relevance and fit first, to provide context for the inquiry into whether the expert's methodology is sound, and it refers to reliability and validity in a manner consistent with their scientific usages. The virtue of this proposal, however, is that it is a step-by-step approach that provides a checklist of crucial factors that judges should at least contemplate when making an admissibility decision. This checklist includes internal and external validity, which only appear implicitly in the current standard.

First, the judge should establish the relevance of the specialized knowledge offered, and especially the "fit" of the testimony, to ensure that

192. Landon Summers, *The Justices and Psychological Research: But is it Really Science?*, 21 L. & PSYCHOL. REV. 93, 131 (1997) (citing John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. PA. L. REV. 477, 499 (1986)).

193. Summers, *supra* note 192, at 131. As Monahan and Walker point out, these guidelines are not unlike the considerations courts take into account when evaluating the precedential weight of cases. Monahan & Walker, *supra* note 192, at 488-89.

the testimony will assist the trier of fact.¹⁹⁴ If the case is at the class certification stage, the question of relevance and fit should focus on whether the testimony helps establish the Rule 23 requirement or requirements at issue.¹⁹⁵ Beginning with an analysis of relevance and fit aligns the admissibility standard for expert testimony with the admissibility standard for other evidence, ensures that the judge performs the inquiry into the expert's methodology in the context of the relevance and fit of the testimony, and prevents the judge from undergoing a technical and often complex analysis of the expert's methodology if the evidence is irrelevant or unhelpful.

Second, as a foundation, the judge should review the expert's qualifications to offer specialized knowledge. This evaluation also is less technical and should rely on the plain meaning of Rule 702, accounting for the expert's "knowledge, skill, experience, training, or education" in the relevant field. However, the judge should be cautious not to rely simply on the expert's reputation for providing trustworthy testimony in other cases.

Third, the judge should examine the principle's methodology, focusing on its reliability to confirm that the principle can be consistently reproduced,¹⁹⁶ its internal validity to determine whether the principle measures what it purports to measure,¹⁹⁷ and its external validity to ensure that the principle generalizes to the issue in dispute. The reliability analysis might rely in part on some combination of the first and third *Daubert* factors (testability and rate of error) and the original 2000 amendment, while the validity analysis is codified in the language this Comment adds to the 2000 amendment.

Finally, in a less scientific and more instinctive approach, the judge might take into account several proxies for "trustworthiness," including the second and fourth *Daubert* factors (peer review/publication and general acceptance) and the expert's reputation. Note that under this recommended approach, these proxies for trustworthiness would be relegated to a minor role.

Significantly, judges should remember that conclusions about the soundness of a principle are best drawn in terms of degree of strength rather than as an absolute division between what is "trustworthy" and what is "untrustworthy."¹⁹⁸ While in the end, judges must make a categorical decision to admit or to exclude expert testimony, they need not approach this testimony with a categorical view of science.¹⁹⁹

194. See Moreno, *supra* note 56, at 1054.

195. See *supra* Part II.G.

196. See Moreno, *supra* note 56, at 1069-70.

197. See *id.* at 1069.

198. *Id.* at 1069.

199. 1 FAIGMAN, *supra* note 32, § 1-3.4.

VI.

CONCLUSION

Like much scientific evidence, social framework evidence often is challenged for whether its methodology is sound. Instead of focusing primarily on whether social framework evidence generally should be admissible as a *category of testimony*, this Comment used the treatment of social framework testimony as a vehicle to explore the *admissibility standard itself*. This Comment concludes that to improve the approach to social framework evidence, the legal community would benefit from upgraded scientific training and a reformed admissibility standard.

Better scientific training is the only solution to some challenges in the evaluation of social science testimony, especially in circumstances that require judges to recognize the value of or deficiencies in this evidence. For example, the revision of the 2000 amendment to Rule 702 cannot ameliorate shortcomings in judicial decisions that sharply limit scientific principles to fit a legal doctrine or that treat specialized knowledge as if it were common knowledge. Also, without some source of scientific guidance, judges hardly can assess the more complex methodologies that provide the bases for some categories of social science evidence.

However, a revised approach to the admissibility standard can address other weaknesses in judicial decisions. The added language in the 2000 amendment to Federal Rule of Evidence 702 might provide judges with more helpful analytical tools than the vague concepts of “evidentiary reliability” (still used interchangeably with “scientific validity”) or the four *Daubert* factors currently available. This Comment’s recommended approach to the admissibility standard discourages judges from treating sound methodology as an issue of weight or from dismissing evidence without any explanation of the basis for its inadmissibility. Such an approach also requires judges to examine the expert’s methodology in all cases where the expert testimony appears to be relevant and helpful to the jury.

But judges are not the only parties targeted by the suggestions in this Comment. These developments might influence experts to refine their testimony to pay more attention to addressing whether it is based on sound methodology. They also should encourage attorneys to prepare their experts to meet the new criteria and to refine their understanding of scientific concepts in preparation for oral arguments and cross-examination. The hope is that the ultimate effect of these changes is a shift away from the dependence on superficial indicators of an expert’s trustworthiness, such as the expert’s reputation or the level of acceptance of his or her work. If implemented, the more substantive gatekeeping approach promised by *Daubert* might finally replace these superficial indicators.

As the legal community increasingly draws upon scientific knowledge to guide its decisions, it must make changes in the law to keep in step with this trend. This Comment's proposals to improve scientific training for law students, attorneys, and judges, to revise the 2000 amendment to Federal Rule of Evidence 702, and to provide a recommended approach to expert testimony might begin to bridge the gap between the legal and social science communities in this area of the law.

