

The Merits of the Paternalistic Justification for Restrictions on the Admissibility of Expert Evidence

Joseph Sanders

I. INTRODUCTION

Jennifer Bishop gave birth to a son shortly before 8 a.m. in July of 1993. Because she had decided not to breast-feed her child, that evening with dinner she received a tablet of Parlodel to prevent the production of breast milk. The next several hours were horrible. Within an hour and a half she became nauseated and vomited at least twice. Her blood pressure rose, and her temperature increased to over 102 degrees. Within three hours she became drowsy and then rigid. Approximately four hours after taking the drug, she was transferred to intensive care where she suffered respiratory arrest and lapsed into a coma. Six hours later she was pronounced dead.¹ In *Kuhn v. Sandoz Pharmaceuticals*,² Jennifer's estate and her husband sued Sandoz, the manufacturer of Parlodel, arguing the 2.5 mg she ingested was a "direct and proximate cause" of her death. In support of this position, the plaintiffs offered the opinions of a number of experts. The plaintiffs asserted that the experts' opinions were based upon the standard medical methodology of "differential diagnosis."³

¹ *Kuhn v. Sandoz Pharmaceuticals Corp.*, 14 P.2d 1170, 1173-74 (Kan. 2000). According to the court, "[T]he autopsy reported that the probable cause of the death was 'related to postpartum eclampsia' or 'possible bacteremia.'" Definitions are appropriate here to assist the reader: "postpartum" means "[a]fter childbirth;" "eclampsia" is defined as the "[o]ccurrence of one or more convulsions, not attributable to other cerebral conditions such as epilepsy or cerebral hemorrhage, in a patient with preeclampsia;" and "bacteremia" is a condition characterized by "viable bacteria in the circulating blood." *STEDMAN'S MEDICAL DICTIONARY* 1413, 540, 181 (26th ed. 1995).

² 14 P.2d 1170.

³ *Id.* at 1177. In medical dictionaries, differential diagnosis is defined as a "diagnosis based on comparison of symptoms of two or more similar diseases to determine which the patient is suffering from." *TABER'S CYCLOPEDIA MEDICAL DICTIONARY* 404 (14th ed. 1981). However, in legal usage the term is not restricted to the process of distinguishing among diseases. Rather, the term also is used to describe the process of differentiating among the possible causes of the plaintiff's ailment. That is what occurred in *Kuhn*. It is with respect to this latter, perhaps

The defendant challenged the admissibility of this testimony. The United States District Court for the District of Kansas, applying Kansas's version of the *Frye* test for admissibility of expert testimony,⁴ ruled that the plaintiffs' experts, "improperly offer medical causation opinions concerning Parlodel without general acceptance of the bases for those opinions within the relevant scientific community"⁵ The Kansas Supreme Court reversed, holding that the *Frye* test is not applicable to the type of expert testimony at issue in this case.

In *Kuhn*, the court declined the judicial gatekeeping role in cases where experts offer "pure opinion" testimony. Under the newly announced pure opinion exception, the *Frye* test only applies "when an expert witness reaches a conclusion by deduction from applying a new or novel scientific principal, formula, or procedure developed by others."⁶ Opinions that do not rely on "techniques," but rather are "developed from inductive reasoning based on the expert's own experience, observation, or research"⁷ are not to be tested by *Frye* or any other admissibility test. Rather, "[t]he validity of pure opinion is tested by cross-examination of the witness."⁸

As the court correctly notes, once this exception is created a critical question is whether the term "technique" is to be given a narrow or broad meaning.⁹ The Kansas Supreme Court adopts a narrow view of the term. According to the court, the plaintiffs' experts' opinions in *Kuhn* did not hinge on the validity of a scientific principal, device, test, or procedure developed by another but rather on the accuracy of their observation, the extent of their training, and the reliability of their interpretations.¹⁰ None of these are subject to

incorrect usage, that differential diagnosis has become controversial in legal settings. For a general discussion of differential diagnosis testimony, see Joseph Sanders & Julie Machal-Fulks, *The Admissibility of Differential Diagnosis Testimony to Prove Causation in Toxic Tort Cases: The Interplay of Adjective and Substantive Law*, 64 LAW. & CONTEMP. PROBS. 107 (2001); Jean Macchiaroli Eggen, *Clinical Medical Evidence of Causation in Toxic Tort Cases: Into the Crucible of Daubert*, 38 HOUS. L. REV. 369 (2001).

⁴ In *Frye v. United States*, 293 F. 1013 (D.D. Cir. 1923), the court held that novel expert testimony is admissible only when the scientific principle or technique from which it is deduced has gained general acceptance in the particular field in which it belongs. Kansas adopted the *Frye* "general acceptance" test in *State v. Lowry*, 185 P.2d 147 (Kan. 1947). For a more complete discussion of *Frye*, see *infra* Part II.

⁵ *Kuhn*, 14 P.2d at 1177.

⁶ *Id.* at 1179. *Kuhn* borrows the test from a Florida appellate court opinion, *Florida Power & Light Co. v. Tursi*, 729 So. 2d 995, 997 (Fla. App. 1999).

⁷ *Kuhn*, 14 P.2d at 1179.

⁸ *Id.*

⁹ *Id.* at 1180.

¹⁰ *Id.* at 1182.

Frye.¹¹ It appears likely that *Kuhn* will remove most, if not all, medical doctor differential diagnosis testimony from any judicial reliability assessment.¹²

The Kansas Supreme Court justified its holding with the following arguments:

The distinction between pure opinion testimony and testimony based on a scientific method or procedure is rooted in a concept that seeks to limit application of the *Frye* test to situations where there is the greatest potential for juror confusion The distinction would be consistent with Kansas' appellate decisions applying the *Frye* test, almost all of which have involved devices or tests surrounded by an "aura of infallibility" to which a trier of fact might tend to ascribe "an inordinately high degree of certainty."¹³

Judges generally are not trained in scientific fields and, like jurors, are lay persons concerning science. A Kansas jury has a constitutional mandate to decide between conflicting facts, including conflicting opinions of causation. The district judge . . . controls expert opinion evidence that would unduly prejudice or mislead a jury or confuse the question for resolution. Cross-examination, the submission of contrary evidence, and the use of appropriate jury instructions form a preferred method of resolving factual disputes.¹⁴

To summarize, the court justifies its result by arguing for the following five propositions: 1) jurors are relatively unconfused by expert testimony that does not involve a "technique" such as a lie detector; 2) jurors are less confused by testimony based on an expert's own investigations than they are by testimony based on the investigations of other researchers; 3) jurors are less confused by "inductive" reasoning than by "deductive" reasoning; 4) judges are no

¹¹ *Logerquist v. McVey* reaches a similar conclusion in a case involving repressed memory testimony. *Logerquist v. McVey*, 1 P.3d 113 (Ariz. 2000). The *Logerquist* court said:

[a]lthough compliance with *Frye* is necessary when the scientist reaches a conclusion by applying a scientific theory or process based on the work or discovery of others, under [Arizona Rules of Evidence 702 and 703] experts may testify concerning their own experimentation and observation and opinions based on their own work without first showing general acceptance.

Id. at 123.

¹² The plaintiff's three experts offered to testify that Parlodel caused or contributed to Bishop's death. They arrived at this result through a process of "differential diagnosis" by which they considered and ruled out other causes. *Id.* at 1176-77.

¹³ *Id.* at 1181 (internal citations omitted).

¹⁴ *Id.* at 1182 (internal citations omitted).

better than jurors in assessing the merits of a scientific argument; and 5) cross-examination, competing experts, and judicial instructions are adequate to the task of clearing up any residual jury confusion.

Underlying the specific arguments in *Kuhn* is one central idea: the jury is better able to assess arguments like those offered by the plaintiffs' experts than it is other types of expert evidence. Implicit in this argument is the premise that restrictive evidentiary rules are best justified, if they are justified at all, as a way to protect juries, litigants, and the law itself from jury shortcomings. Restrictive rules are, in a word, justified by appeals to paternalism.¹⁵

The United States Supreme Court's opinion in *Kumho Tire Co., Ltd. v. Carmichael*¹⁶ reflects this justification in the following passage:

And whether the specific expert testimony focuses upon specialized observations, the specialized translation of those observations into theory, a specialized theory itself, or the application of such a theory in a particular case, the expert's testimony often will rest "upon an experience confessedly foreign in kind to [the jury's] own."¹⁷ The trial judge's effort to assure that the specialized testimony is reliable and relevant can help the jury evaluate that foreign experience, whether the testimony reflects scientific, technical, or other specialized knowledge.¹⁸

The arguments in *Kuhn* offer a springboard for a discussion of the paternalistic justification and the empirical evidence now available to assess the merits of this justification. Section II of this article situates the *Kuhn* opinion within the context of the existing rules concerning the admissibility of expert testimony, especially as they have developed in the aftermath of the United States Supreme Court opinion in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹⁹ I note that although *Daubert* changed the criteria to be used in assessing the admissibility of expert testimony, it certainly was not the first opinion to introduce paternalistic considerations to this area of evidence. Section III sketches out the contours of a paternalistic argument developed by Alvin Goldman and others writing in the tradition of "naturalized epistemology," an epistemology that builds on and takes its direction from empirical observations about how we know things. From this perspective, if paternalism is to justify exclusionary rules,

¹⁵ See Leslie A. Lunney, *Protecting Juries From Themselves: Restricting the Admission of Expert Testimony in Toxic Tort Cases*, 48 SMU L. REV. 103 (1994).

¹⁶ *Kumho Tire Co., Ltd., v. Carmichael*, 526 U.S. 137 (1999).

¹⁷ Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40, 54 (1901).

¹⁸ *Kumho Tire*, 526 U.S. at 149.

¹⁹ 509 U.S. 579 (1993).

paternalism itself must be justified by an appeal to empirical evidence that informs us about the ability of actors in an institution to convey and understand information. Building on this discussion, Section IV assesses the available empirical research on these topics and concludes with a brief discussion of the relationship between admissibility rules and the ends of justice. Section V summarizes the existing evidence for a paternalistic approach, suggests some considerations courts should bring to the admissibility task, and briefly discusses other procedural justice concerns that may argue for the admissibility of evidence, even if this would reduce the overall accuracy of verdicts.

II. RULES CONCERNING THE ADMISSIBILITY OF SCIENTIFIC EVIDENCE

Kuhn is a counter-revolutionary opinion. It runs against the tide of heightened judicial scrutiny of expert testimony that is associated with the United States Supreme Court's opinion in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*²⁰ Indeed, *Daubert* is often said to have ushered in the "Daubert revolution."²¹ In this Section, I provide a short review of the law with respect to admissibility of scientific evidence before and after *Daubert*.

Kansas is one of a substantial number of states²² that still follows the admissibility rule first adopted in *Frye v. United States*.²³ In *Frye*, the defendant, accused of murder, offered the results of a "systolic blood pressure deception test," a precursor to the polygraph, as evidence of his innocence. Prior to *Frye*, most courts only asked whether the expert was "qualified" before admitting the expert's testimony and, in some jurisdictions, whether the subject matter in issue was beyond the range of knowledge of the average juror.²⁴ However, *Frye*'s

²⁰ *Id.* The literature on *Daubert* is voluminous. For a discussion of the ruling and its increasingly wide reach, see DAVID FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY ch. 1 (2d ed. 2002) [hereinafter FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE].

²¹ See David L. Faigman, *The Law's Scientific Revolution: Reflections and Ruminations on the Law's Use of Experts in Year Seven of the Revolution*, 57 WASH. & LEE L. REV. 661 (2000).

²² See David E. Bernstein, *Frye, Frye, Again: The Past, Present, and Future of the General Acceptance Test*, 4 JURIMETRICS J. 385, 386 (2001); see also Heather G. Hamilton, *The Movement from Frye to Daubert: Where Do the States Stand?*, 38 JURIMETRICS J. 201 (1999); Jolle Anne Moreno, *Eyes Wide Shut: Hidden Problems and Future Consequences of the Fact-Based Reliability Standard*, 34 SETON HALL L. REV. ___ (forthcoming in Fall 2003).

²³ 293 F. 1013 (D.C. Cir. 1923). My discussion here borrows heavily from FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE, *supra* note 20, at ch. 1.

²⁴ FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE, *supra* note 20, at ch. 1. For a general historical overview of the use of expert testimony, see Stephan Landsman, *Of*

expert's testimony posed special problems because he proposed to testify to a novel technique that was not currently being used by a community of experts. In a brief two-page opinion, Judge Van Orsdel placed an additional hurdle in the path of those who would introduce expert testimony. The key passage established what has come to be called the "general acceptance test."²⁵ Expert testimony is admissible when the scientific principle or technique from which it is deduced has gained general acceptance in the particular field in which it belongs.

With the adoption of the Federal Rules of Evidence in the 1970s, the *Frye* test began a slow decline in the federal courts. While new theories and techniques gained general acceptance, the *Frye* test was criticized for being too conservative because it imposed a waiting period.²⁶ In addition, others criticized the test as too liberal because of the difficulty of defining the relevant field within which general acceptance must be achieved. If the field is narrowly defined to include the proffered expert and other like-minded individuals, little will be excluded.²⁷

These criticisms and the fact that the reporter's notes accompanying the Federal Rules of Evidence did not even mention the case when discussing the admissibility of expert testimony caused a number of federal circuits to abandon the test.²⁸ Other circuits,

Witches, Madmen, and Products Liability: An Historical Survey of the Use of Expert Testimony, 13 BEHAV. SCI. & L. 131 (1995). See also David L. Faigman et al., *Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence*, 15 CARDOZO L. REV. 1799, 1803-09 (1994). The classic exploration of this historical question is Hand, *supra* note 17.

²⁵ The court in *Frye* held that:

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while the courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.

Frye, 293 F. at 1014.

²⁶ For a case praising the conservative nature of *Frye*, see *People v. Leahy*, 882 P.2d 321, 325 (Cal. 1994).

²⁷ See Paul C. Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half Century Later*, 80 COLUM. L. REV. 1197 (1980).

²⁸ The most influential early circuit court opinion rejecting *Frye* is *United States v. Downing*, 753 F.2d 1224 (3d Cir. 1985). In a case involving expert testimony on eyewitness identification, Judge Becker said that in order to be admitted the evidence must survive the trial court's preliminary inquiry. In an *in limine* proceeding, the judge should balance: (1) the reliability of the scientific principles the expert employed against (2) the likelihood that the evidence may overwhelm or

however, concluded *Frye* did survive the adoption of the Rules.²⁹

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,³⁰ the Supreme Court officially ended the debate. In a case involving the morning sickness drug Bendectin, the Court concluded that *Frye*'s rigid "general acceptance" standard is contrary to the thrust of the Federal Rules, which were intended to lower the barriers to expert opinion testimony.³¹ However, *Daubert* agreed that Federal Rule of Evidence 702 does modify Rule 402's directive to admit all relevant evidence. Rule 702 also requires reliability; evidence which is relevant but unreliable is inadmissible.³²

What constitutes reliability? In this case, where all the experts purported to be scientists, the Court turned to science for an answer. Reliable opinions are those that are arrived at using the "methods and procedures of science."³³ In a footnote, the Court added, "[i]n a case involving scientific evidence, evidentiary reliability will be based upon scientific validity."³⁴

Daubert did not offer a systematic presentation of what scientists mean when they inquire about validity, but it did propose the following four non-exclusive factors courts might consider when making a reliability/validity assessment: 1) whether the expert's theory or technique is falsifiable and has been tested;³⁵ 2) the reliability of a procedure and its potential rate of error;³⁶ 3) whether the theory or technique has been subjected to peer review³⁷ and whether the results have been published;³⁸ and 4) in a partial resurrection of the *Frye* test, whether the expert's methods and reasoning enjoy general acceptance in a relevant scientific community.³⁹

mislead the jury. In addition, the trial court should examine the "fit" between the proffered scientific testimony and the contested issues in the case. *Id.* at 1226. Concern with reliability and fit have become cornerstones of post-*Daubert* jurisprudence. The Fifth Circuit, in *Christophersen v. Allied-Signal Corp.*, set out a similar test for admissibility. *Christophersen v. Allied-Signal Corp.*, 939 F.2d 1106 (5th Cir. 1991) (en banc), *cert. denied*, 503 U.S. 912 (1992).

²⁹ *United States v. Solomon*, 753 F.2d 1522, 1526 (9th Cir. 1985).

³⁰ 509 U.S. 579.

³¹ *Id.* at 588.

³² *Id.* at 590.

³³ *Id.*

³⁴ *Id.* at 590 n.9.

³⁵ *Id.* at 593.

³⁶ *Id.*

³⁷ *Id.*

³⁸ *Id.* at 593-94.

³⁹ *Id.* at 594.

In addition, the Court noted that Rule 702 requires that the expert evidence “assist the trier of fact to understand the evidence or to determine a fact in issue.”⁴⁰ Justice Blackmun stated that,

[t]his condition goes primarily to relevance . . . [t]he consideration has been aptly described by Judge Becker as one of “fit.” “Fit” is not always obvious, and scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes.⁴¹

The “fit” requirement involves an assessment of whether the expert’s chain of reasoning contains an inferential gap that is too wide.⁴²

In another footnote, the Court expressly limited *Daubert* to scientific evidence. It noted that Rule 702 applies to “technical or other specialized knowledge” as well, but added, “[o]ur discussion is limited to the scientific context because that is the nature of the expertise offered here.”⁴³ *Daubert* left two related questions for later cases to answer: 1) does *Daubert*’s reliability requirement apply at all to non-scientific evidence; and 2) if it does apply, what role do the *Daubert* factors play in these cases? These questions frequently arose with respect to the admissibility of clinical medical testimony similar

⁴⁰ *Id.*

⁴¹ *Id.* at 591 (citing *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985)).

⁴² Courts may find a lack of fit when the studies presented by the expert simply fail to support the expert’s position. Using the “fit” requirement in this way causes courts to move close to excluding an expert’s testimony because of the expert’s conclusion. This is something the Supreme Court in *Daubert* specifically cautioned against when it said that the focus of the 702 validity inquiry “must be solely on principles and methodology, not on the conclusions that they generate.” *Daubert*, 509 U.S. at 595.

Most appellate courts downplayed the Supreme Court’s methodology-conclusion distinction. For example, in *Paoli*, an opinion following *Daubert*, Judge Becker himself stated “we think that [the distinction between principles and methods versus conclusions] has only limited practical import . . . a challenge to ‘fit’ is very close to a challenge to the expert’s ultimate conclusion about the particular case, and yet it is part of the judge’s admissibility calculus under *Daubert*.” *In re Paoli*, 35 F.3d 717, 746 (3d Cir. 1994).

In *General Electric v. Joiner*, 522 U.S. 136 (1997), the Supreme Court ratified Judge Becker’s view:

[N]othing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence which is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered. That is what the District Court did here and we hold that it did not abuse its discretion in so doing.

522 U.S. at 146.

⁴³ *Daubert*, 509 U.S. at 590 n.8.

to the testimony in *Kuhn*⁴⁴ and other types of “experience” testimony.⁴⁵

The Supreme Court answered these two questions in *Kumho Tire Co. v. Carmichael*.⁴⁶ The plaintiffs’ expert in *Kumho Tire* was prepared to testify that the tire failure, which led to the crash of the plaintiffs’ minivan, was the result of a manufacturing or design defect, not a result of abuse. The trial court excluded this testimony after finding that “none of the four admissibility criteria outlined by the *Daubert* court are satisfied in this case.”⁴⁷ Because the expert testimony was the plaintiff’s only evidence of defect, the district court judge then granted the defendant’s summary judgment motion.⁴⁸ The plaintiffs appealed, arguing that the district court should not have applied *Daubert*’s reliability framework because the case did not involve a “scientific” expert.⁴⁹ The United States Court of Appeals for the Eleventh Circuit agreed that *Daubert* applied only to scientific testimony. Whether the expert’s testimony was or was not scientific would have little consequence if the court invoked uniformly stringent admissibility criteria. The circuit court did assert that it was prepared to affirm a well-reasoned trial court decision to exclude the plaintiffs’ expert’s testimony on reliability grounds if, upon remand, the trial court did so without invoking the *Daubert* criteria.⁵⁰ However, in another part of the opinion, the circuit court said, “[t]hus, the question in this case is whether Carlson’s testimony is based on his application of scientific principles or theories [which we should submit to a *Daubert* analysis] or on his utilization of personal

⁴⁴ See, e.g., *Moore v. Ashland Chem., Inc.*, 126 F.3d 679 (5th Cir. 1997), *vacated*, 151 F.3d 269 (5th Cir. 1998) (en banc).

⁴⁵ See, e.g., *Compton v. Subaru of America, Inc.*, 82 F.3d 1513 (10th Cir. 1996). “The language in *Daubert* makes it clear the factors outlined by the Court are applicable only when a proffered expert relied on some principle or methodology. In other words, application of the *Daubert* factors is unwarranted in cases where expert testimony is based solely upon experience or training.” *Id.* at 1518. *But see* *Watkins v. Telsmith, Inc.*, 121 F.3d 984, 991 (5th Cir. 1997) (“[I]t seems exactly backwards that experts who purport to rely on general engineering principles and practical experience might escape screening by the district court simply by stating that their conclusions were not reached by any particular method or technique. The moral of this approach would be, the less factual support for an expert’s opinion, the better.”).

⁴⁶ 526 U.S. 137 (1999). In *General Electric Co. v. Joiner*, 522 U.S. 136 (1997), the Court concluded, *inter alia*, that trial courts’ 702 rulings should be reviewed under an abuse of discretion standard.

⁴⁷ *Carmichael v. Samyang Tire, Inc.*, 923 F. Supp. 1414, 1521 (S.D. Ala. 1996).

⁴⁸ *Id.* at 1524.

⁴⁹ *Carmichael v. Samyang Tire, Inc.*, 131 F.3d 1433, 1435 (11th Cir. 1997).

⁵⁰ *Id.* at 1436 n.9.

experience and skill with failed tires [which we would usually expect a district court to allow a jury to evaluate].⁵¹ This suggests a more lenient admissibility standard for non-science experts and echoes the *Kuhn* court's position that expert experience evidence should rarely be kept from juries.

The Supreme Court reversed the Eleventh Circuit and held that excluding Carlson's testimony was not an abuse of discretion.⁵² The reliability requirement of Rule 702 applies to all expert testimony.⁵³ As to the role of the four *Daubert* factors, the court adopted a flexible position:

We also conclude that a trial court may consider one or more of the more specific factors that *Daubert* mentioned when doing so will help determine that testimony's reliability. But, as the Court stated in *Daubert*, the test of reliability is "flexible," and *Daubert*'s list of specific factors neither necessarily nor exclusively applies to all experts or in every case. Rather, the law grants a district court the same broad latitude when it decides how to determine reliability as it enjoys in respect to its ultimate reliability determination.⁵⁴

It would be a mistake to read *Kumho Tire* to say that the trial

⁵¹ *Id.* at 1436.

⁵² *Kumho Tire*, 526 U.S. 137.

⁵³ The Court provides four reasons why *Daubert*'s general reliability requirement applies to all expert testimony. First, the language of Rule 702 makes no relevant distinction between "scientific" knowledge and "technical" or "other specialized" knowledge. Second, although the *Daubert* opinion did restrict itself to "scientific" knowledge, that was only because the issue presented in the case involved scientific expertise. Third, the evidentiary rationale that underlies the gatekeeping requirement is that Rules 702 and 703 give wide latitude to all experts to offer their opinions, latitude that is unavailable to other witnesses. This latitude is premised on the "assumption that the expert's opinion will have a reliable basis in the knowledge and experience of his discipline." 526 U.S. 137, 147 (quoting *Daubert*, 509 U.S. at 589). Because the Rules grant this latitude to all experts, they all must meet the reliability standard. Fourth, a rule that distinguishes between scientific experts and other experts would be very difficult if not impossible to administer.

Federal Rule of Evidence 702 now has been revised to incorporate the main components of *Daubert* and *Kumho Tire*. Amended Rule 702 provides as follows:

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 (as modified Dec. 1, 2000).

⁵⁴ 526 U.S. at 141-42.

court may simply ignore the *Daubert* factors in non-science cases.⁵⁵ The Court noted that “a trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of expert testimony.”⁵⁶ Implicit in the *Kumho Tire* opinion is the belief that the justifications for restricting expert testimony are as valid when expertise is based on experience as when it is based on science, a point with which the Kansas Supreme Court presumably would disagree.

Because there are many differences between the old *Frye* test and the *Daubert-Kumho Tire* line of cases, it is easy to lose sight of the fact that, when compared to an opinion like *Kuhn*, they share much in common. Both the *Frye* approach and that of *Daubert* ask the judge to exclude relevant evidence if it is unreliable. They differ in how to assess reliability and, perhaps, the height of the reliability hurdle over which the expert must jump.

One justification for the reliability requirements is paternalistic. Restrictions on the admissibility of expert evidence shelter jurors from their own shortcomings. Absent such rules, jurors will be more likely to reach an incorrect conclusion. The *Kuhn* opinion is instructive precisely because it rejects the paternalistic justification for reliability standards.

III. PATERNALISTIC JUSTIFICATIONS IN LAW

Paternalistic arguments are not restricted to the law of evidence. To the contrary, they are staples of law. The state frequently constrains behavior on paternalistic grounds. Laws, which dispense food stamps rather than cash, forbid gambling, regulate interest rates, bar dueling, and make suicide a crime, along with the Eighteenth Amendment were all designed to prevent people from acting in ways contrary to what others think are against their own self interest.⁵⁷ Of course, paternalistic arguments do not always prevail. Other considerations may cause us to reject them either because we think they are misplaced—that people will act in their own best interest—or because of other considerations that outweigh the force of paternalistic considerations. The pull of different considerations

⁵⁵ See *Black v. Food Lion, Inc.*, 171 F.3d 308, 311-12 (5th Cir. 1999) (“In the vast majority of cases, the district court first should decide whether the factors mentioned in *Daubert* are appropriate. Once it considers the *Daubert* factors, the court then can consider whether other factors, not mentioned in *Daubert*, are relevant to the case at hand.”).

⁵⁶ 526 U.S. at 152.

⁵⁷ Much of this list comes from Gerald Dworkin, *Paternalism*, in *PATERNALISM* 19, 20 (Rolf Satorius ed., 1983).

can be seen in one high-profile area: Supreme Court opinions concerning restrictions on commercial speech.⁵⁸

A. *Paternalism in Speech Cases*

Paternalistic arguments frequently arise in discussions regarding restrictions on commercial speech. Typically, the legislature justifies restrictive enactments by arguing, *inter alia*, that the consumers of some messages would be better off without the communication because the information contained in the message would cause them to act against their own best interest and also, perhaps, against the best interests of the community. As Justice Thomas noted in his concurring opinion in *Greater New Orleans Broadcasting Assn. v. United States*,⁵⁹ the primary governmental justification for such restrictions are paternalistic; that is they are designed “to keep legal users of a product or service ignorant in order to manipulate their choices in the marketplace.”⁶⁰

First Amendment free speech jurisprudence is an area where anti-paternalistic instincts are perhaps the strongest. It is not surprising, therefore, that the Supreme Court generally rejects these justifications.⁶¹ In the area of commercial speech, however, such arguments have not met with universal failure.⁶²

⁵⁸ Paternalistic arguments also arise with respect to restrictions on obscenity, hate speech, and speech directed toward minors. See, e.g., Catherine J. Ross, *Anything Goes: Examining the State's Interest in Protecting Children From Controversial Speech*, 53 VAND. L. REV. 427, 495 (2000).

⁵⁹ 527 U.S. 173.

⁶⁰ *Id.* at 197.

⁶¹ See Geoffrey R. Stone, *Content Regulation and the First Amendment*, WM. & MARY L. REV. 189, 231-33 (1983); see also David L. Shapiro, *Courts, Legislatures, and Paternalism*, 74 VA. L. REV. 519, 543 (1988); Martin H. Redish & Howard M. Wasserman, *What's Good For General Motors: Corporate Speech and the Theory of Free Expression*, 66 GEO. WASH. L. REV. 235, 267 (1998).

⁶² Anti-paternalism was an important rationale in *Virginia State Bd. of Pharmacy v. Virginia Citizens Consumer Council, Inc.*, 425 U.S. 748 (1976), a 1976 opinion often cited as the beginning of the court's renewed interest in protecting commercial speech. The Court expressly critiqued the paternalistic justifications offered by Virginia for restricting price advertising of prescription drugs. *Id.* at 769-70. On the other hand, the Court adopted a more paternalistic position in *Posadas de Puerto Rico Assoc. v. Tourism Co.*, 478 U.S. 328 (1986), permitting the government to restrict casino advertising addressed to Puerto Rican citizens. Then, in *Florida Bar v. Went For It, Inc.*, 515 U.S. 618 (1995), the Court exhibited some sympathy toward a paternalistic argument for a bar association rule prohibiting lawyers from using direct mail to solicit personal injury clients within thirty days of an accident. More recently, however, in *44 Liquormart, Inc. v. Rhode Island*, 517 U.S. 484 (1996), the Court rejected Rhode Island's restriction on alcoholic beverage price advertising that the state defended on the ground that it was designed to promote temperance by reducing the consumption of alcohol by its citizens. In the process it rejected the

Why have paternalistic arguments in this area been more central to, and met with slightly greater success, than in other areas of First Amendment jurisprudence? Professor Daniel Halberstam argues that the centrality of paternalistic arguments in the commercial speech arena is due to a basic difference between these cases and other First Amendment opinions.⁶³ Most First Amendment doctrine focuses on preserving the ability of speakers to communicate their views, but “commercial speech doctrine disclaims significant reliance on the speaker-based model, and instead focuses on the listener.”⁶⁴ This “focus on the listener” reflects the fact that this area of First Amendment law does not view communicative interactions “as abstract exchanges of views between persons about whom nothing is known, but instead, as context-dependent interactions with purposes that can be judicially ascertained with a reasonable degree of confidence.”⁶⁵ Professor Halberstam argues that we should assess

argument advanced in *Posadas* that restrictions should be given more leeway if they regulate “vice” activities from which the state has a particularly strong interest in shielding its citizens. 44 *Liquormart*, 517 U.S. at 514. This position was reaffirmed in *Greater New Orleans Broadcasting Assn. v. United States*, 27 U.S. 173 (1999), where the Court held that a prohibition on broadcasting lottery information could not be applied to advertisements of lawful private casino gambling that were broadcast where such gambling is legal. In *Lorillard Tobacco Co. v. Reilly*, 533 U.S. 525 (2001), the Court struck down Massachusetts’ restrictions on outdoor and point of sale advertising of tobacco products. In *Thompson v. Western States Med. Ctr.*, 535 U.S. 357 (2002), the Court struck down an advertising restriction on compound drugs. The government’s justification for the restriction was not based on paternalistic arguments, however. 535 U.S. at 370.

44 *Liquormart*, *Greater New Orleans*, and *Reilly* have generated the usual volume of law review commentary both about these cases and their implication for related issues such as professional speech, restrictions on tobacco advertising, advertising prescription drugs, and ads extolling the beneficial effects of alcohol consumption. See Sean P. Costello, *Strange Brew: The State of Commercial Speech Jurisprudence Before and After 44 Liquormart, Inc. v. Rhode Island*, 47 CASE W. RES. L. REV. 681 (1997); see also Arlen W. Langvardt, *The Incremental Strengthening of First Amendment Protection for Commercial Speech: Lessons from Greater New Orleans Broadcasting*, 37 AM. BUS. L.J. 587 (2000); Daniel Halberstam, *Commercial Speech, Professional Speech, and the Constitutional Status of Social Institutions*, 147 U. PA. L. REV. 771 (1999); Nat Stern, *In Defense of the Imprecise Definition of Commercial Speech*, 58 MD. L. REV. 55 (1999); Martin H. Redish, *Tobacco Advertising and the First Amendment*, 81 IOWA L. REV. 589 (1996); Margaret Gilhooley, *Constitutionalizing Food and Drug Law*, 74 TULANE L. REV. 815 (2000); Nancy K. Plant, *Prescription Drug Promotion on the Internet: Tool for the Inquisitive or Trap for the Unwary?*, 42 ST. LOUIS U. L.J. 89 (1998); Erik Bierbauer, *Liquid Honesty: The First Amendment Right to Market the Health Benefits of Moderate Alcohol Consumption*, 74 N.Y.U. L. REV. 1057 (1999); Mitchell N. Berman, *Commercial Speech and the Constitutional Conditions Doctrine: A Second Look at “The Greater Includes the Lesser,”* 55 VAND. L. REV. 693 (2002).

⁶³ See Halberstam, *supra* note 62.

⁶⁴ *Id.* at 831.

⁶⁵ *Id.* at 830.

restrictions on commercial speech in terms of whether they sustain or improve the integrity of communication in the buyer-seller relationship.⁶⁶ Using this qualitative approach, he would approve of the position set forth in *Central Hudson Gas & Electric Corp. v. Public Service Commission*,⁶⁷ that speech restrictions prohibiting false or deceptive advertising are permissible because they impede communication between a buyer and seller.⁶⁸

More difficult are the situations where the speech is not misleading, and where it could not be said that a perfectly rational and fully-informed consumer would be unable to assess the information and give it the attention it deserved. Ultimately, restrictions in this situation must search for their justification in alleged listener shortcomings.⁶⁹

I find that I am in general agreement with the Halberstam position. Restrictions on paternalistic grounds are most easily justified within the context of a specific set of communications directed at a defined purpose where parties play defined roles.

B. Paternalism in Evidence Law

When a case is tried to a jury, jurors, litigants, and judges play roles not unlike the role of consumer, advertiser, and government in commercial speech cases. When the parties to the litigation wish to communicate their message to jurors, judges are asked to limit or restrict what the speaker should be permitted to say. These limitations are justified in part by the legal system's assessment of the capacity of the jury, and what will be best for them with respect to the evidence they receive.⁷⁰

⁶⁶ *Id.* at 866.

⁶⁷ 447 U.S. 557 (1980). *Central Hudson* involved a challenge to a New York agency regulation that barred utilities from advertising to promote the use of electricity. It remains the most completely articulated constitutional test for commercial speech. Part one of the four part *Hudson* test is whether the affected speech is non-misleading and concerns a lawful activity. If it fails on either prong, it receives no First Amendment protection. *Id.* at 563, 566, n.9. When speech passes this threshold, the next three parts of the *Hudson* test come into play. Together, they require the government to show a substantial interest to be served by its restriction on commercial speech (part two), the restriction directly advances the underlying government interest (part three), and is no more extensive than necessary to further it (part four). *Id.* at 563-64, 566; *see also* Langvardt, *supra* note 62, at 599-601.

⁶⁸ Halberstam, *supra* note 62, at 857.

⁶⁹ *See* David A. Strauss, *Persuasion, Autonomy, and Freedom of Expression*, 91 COLUM. L. REV. 334, 335 (1991).

⁷⁰ There are, of course, many differences between restrictions of commercial speech and restrictions of testimony. One might note, for example, that rarely does the judge act on her own. Rather, restrictions only occur when the opposing party

This raises the question of how to determine what is best, or to put the question a different way, “best” with what end in view? Here one must make a fundamental choice. If the reader disagrees with the next paragraph, he or she may question much of what follows. I return to this point in Section IV.

The primary goal in the view of the law of evidence, and of the trial itself, is to uncover the truth. Ronald Allen and Brian Leiter argue:

it is striking and important that the vast majority of the rules of evidence have as their primary rationale their (alleged) truth-conductive virtues. Competency of witnesses, authentication of evidence, relevancy, expert testimony, and hearsay (including the exceptions) all, at bottom, rest on the thought that inclusion and exclusion of evidence in line with these rules will increase the frequency with which truth is ascertained.⁷¹

Other non-evidentiary rules governing jury behavior appear to be motivated by the same consideration. These rules include, for example, the prohibition of note-taking and pre-deliberation discussion of the case by the jury.⁷²

If ascertaining the truth is the primary goal of a trial, a central question becomes what evidence should one consider in order to maximize the likelihood of arriving at a correct answer? Alvin Goldman notes that, in the philosophy of science, a popular

asks for them. In this sense the judge may be more passive than the legislature. We should not overemphasize this distinction, however. It is rare for either courts or legislatures to act *sua sponte*.

⁷¹ Ronald Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491, 1501 (2001).

More formally, Federal Rule of Evidence section 102 lists the ascertainment of truth as one of the Rules' primary objectives:

Rule 102. Purpose and Construction:

These rules shall be construed to secure fairness in administration, elimination of unjustifiable expense and delay, and promotion of growth and development of the law of evidence to the end that the truth may be ascertained and proceedings justly determined.

FED. R. EVID. 102. As the rule suggests, however, ascertaining the truth is not the sole objective of the rules.

Commentators vary widely in the belief that the search for truth is the primary goal of trials or the primary function of the jury. Compare Franklin Strier, *Making Jury Trials More Truthful*, 30 U.C. DAVIS L. REV. 95 (1996), with Mark Cammack, *In Search of the Post-Positivist Jury*, 70 IND. L.J. 405 (1995). In addition, see Edward J. Imwinkelried, *The Worst Evidence Principle: The Best Hypothesis as to the Logical Structure of Evidence Law*, 46 U. MIAMI L. REV. 1069, 1070 (1992), for a useful discussion of “jury control” and “best evidence” explanations of the structure of evidence law.

⁷² See Joseph Sanders, *Scientifically Complex Cases, Trial by Jury, and the Erosion of Adversarial Processes*, 48 DEPAUL L. REV. 355, 385 (1998).

epistemological answer to this question is that an individual should base his judgments and subjective probabilities on the total evidence available. Therefore, when confronted with a decision, individuals should use all available evidence that can be collected and used at a reasonable cost.⁷³ He calls this the “requirement of total evidence.”⁷⁴ When the evidence that might be made available to the individual is under the control of another, a corollary of this requirement is that the other person should make available to the individual all of the evidence relevant to the question at hand. Goldman calls this the control version of the requirement of total evidence.⁷⁵ In the present context, this control version argues that judges should not impose admissibility rules that keep relevant, but relatively unreliable, evidence from juries.⁷⁶

Goldman argues, however, that the legal system’s failure to adopt a control version of the requirement of total evidence is not necessarily unwise, because such a rule may not produce the most accurate results if jurors give too much weight to relatively unreliable evidence.⁷⁷ To the degree this is so, an admissibility rule that excludes such evidence may be justified on the grounds of what Goldman calls “epistemic paternalism.” “The general idea is that the indicated rules of evidence are designed to protect jurors from their own ‘folly,’ just as parents might keep dangerous toys or other articles away from children, or might not expose them to certain facts.”⁷⁸ For example, Rule 403 permits the exclusion of evidence that is relevant if its “probative value is substantially outweighed by the danger of

⁷³ Alvin Goldman, *Epistemic Paternalism: Communication Control in Law and Society*, 88 J. OF PHILOSOPHY 113 (1991).

⁷⁴ For a justification of this rule, see PAUL HORWICH, *PROBABILITY AND EVIDENCE* (1982). Goldman notes that there may be a number of non-epistemological grounds for excluding some evidence, for example, if collecting the evidence invades the privacy rights of some individual. Goldman, *supra* note 73, at 114.

⁷⁵ Goldman, *supra* note 73, at 114. Again, Goldman notes there may be many non-epistemic reasons for excluding testimony, for example, that it is the result of an illegal search and seizure.

⁷⁶ Evidence that is completely unreliable is, by definition, irrelevant. That is, it cannot alter the probability that some fact is more or less likely. An evidentiary restriction that admits only relevant evidence does not, therefore, violate the requirement of total evidence as it is ordinarily understood. See Goldman, *supra* note 73, at 113 n.2.

⁷⁷ See Richard Lempert, *Modeling Relevance*, 65 MICH. L. REV. 1021 (1977).

⁷⁸ Goldman, *supra* note 73, at 118. Specific empirical examples of situations where restricting information to jurors lead to more accurate conclusions are rare. But see Dale Nance & Scott Morris, *An Empirical Assessment of Presentation Formats for Trace Evidence With a Relatively Large and Quantifiable Random Match Probability*, 42 JURIMETRICS J. 403, 442-44 (2002).

unfair prejudice, confusion of the issues, or misleading the jury.” As Brian Leiter describes the issue:

Epistemic paternalism substitutes the rulemaker’s judgment about what is epistemically best for agents for their own judgment. Assuming that the primary epistemic value is truth, epistemic paternalism entails designing rules of evidence that are epistemically best for jurors, i.e. that lead them to form true beliefs about disputed matters of fact.⁷⁹

Goldman admits that a paternalistic concern for the “welfare” of jurors is a bit odd. If we focus solely on the well-being of jurors themselves, the paternalistic impulses in the law of evidence seem almost trivial. Paternalism is generally thought of as an intervention in a person’s freedom aimed at furthering his own good. The good to be achieved here, however, is not central to the well-being of the individual juror. Indeed, one might say that the intervention is not paternalistic at all, if by the term we only mean that the court is operating in a manner contrary to the jury’s preferences. It might well be that were we to ask jurors many would say, “Sure, keep the unreliable evidence from me. I do not choose to be bothered by it.”

However, Donald VanDeVeer proposes a second definition of paternalistic behavior that, comes closer to describing what courts do, and closer to what legislatures say they are doing when they restrict commercial speech. By this definition, an act is also paternalistic when one deliberately acts to shape another’s preferences in ways that bypass the other’s capacity to resist.⁸⁰ Keeping people from information that they never knew to exist may be paternalistic in this sense. Moreover, the apparently trivial nature of admissibility paternalism disappears if we agree with Gerald Dworkin⁸¹ and John Kleinig⁸² that paternalism exists even when the class of persons whose good is involved is not the same as the class of persons whose freedom is restricted. Requiring medical doctors to be licensed or restricting the ability of individuals to obtain drugs without a prescription are paternalistic in this sense. They are designed not to protect physicians or drug manufacturers, rather they are intended to protect consumers.⁸³ Admissibility restrictions that are justified

⁷⁹ Brian Leiter, *The Epistemology of Admissibility: Why Even Good Philosophy of Science Would Not Make For Good Philosophy of Evidence*, 1997 BYU L. REV. 803, 814 (1997).

⁸⁰ DONALD VANDEVEER, PATERNALISTIC INTERVENTION: THE MORAL BOUNDS ON BENEVOLENCE 19 (1986).

⁸¹ Gerald Dworkin, *Paternalism*, in MORALITY AND THE LAW 107, 110 (Richard A. Wassertrom ed., 1971).

⁸² JOHN KLEINIG, PATERNALISM 11 (1983).

⁸³ Dworkin calls these examples of “impure” paternalism. Dworkin, *supra* note

because they are best for the parties to the litigation, or for the legal system itself, are paternalistic in the same way.⁸⁴

Thus far, the issues have been presented as if they simply involve the ability of individual jurors. However, the acquisition of knowledge has a social as well as an individual dimension. As Alvin Goldman notes, we should be concerned not only with individual knowers but also with the social processes and practices which inculcate belief.⁸⁵ The individual ability of the juror must be understood within the context of the trial and the way in which evidence is typically presented to the factfinder. From this wider perspective, Goldman suggests five factors when considering the effect of paternalistic rules on decisions, that are relevant to whether paternalistic communication control policies will lead to veritistic outcomes.⁸⁶ They are as follows: 1) the characteristics of the audience; 2) the characteristics of the speakers; 3) the characteristics of the controller (or gatekeeper); 4) the controller's criterion of selection among speakers or messages; and 5) the availability of other alternate channels of communication that address the same topic.⁸⁷

It is a relatively straightforward undertaking to apply Goldman's factors to the admissibility of expert testimony. For example, a number of Goldman's considerations relate to arguments advanced in *Kuhn*. Characteristics of the audience (jurors) and the speakers (the experts) relate to *Kuhn's* first three arguments: 1) jurors are relatively unconfused by expert testimony that does not involve a "technique" such as a lie detector; 2) jurors are less confused by testimony based on an expert's own investigations than they are by testimony based on the investigations of other researchers; and 3) jurors are less confused by inductive reasoning than by deductive reasoning. Characteristics of the controller (the judge), and the

81, at 22.

⁸⁴ Goldman, *supra* note 73, at 119.

⁸⁵ ALVIN I. GOLDMAN, KNOWLEDGE IN A SOCIAL WORLD 4-5 (1999). Goldman is a leading advocate of "naturalized epistemology," an approach to knowledge that wishes to craft norms to guide our acquisition of knowledge based in large part on empirical information about how the human cognitive apparatus works. His perspective has a strong social dimension. Social epistemology focuses on social paths or routes to knowledge, examines the spread of information (or misinformation) across groups, and may consider collective or corporate entities such as juries or legislatures to be the relevant unit of analysis. The core commitment of naturalized epistemology is methodological: an examination of the social mechanisms and practices that inculcate belief. See Allen & Leiter, *supra* note 71.

⁸⁶ Goldman, *supra* note 73, at 124.

⁸⁷ *Id.* I have rearranged the order in which Goldman discusses the variables.

controller's criterion of selection, relate to *Kuhn's* fourth argument: judges are no better than jurors in assessing the merits of a scientific argument. The availability of alternative channels touches on *Kuhn's* fifth argument: cross-examination, competing experts, and judicial instructions are adequate to the task of clearing up any residual jury confusion.

The *Kuhn* opinion cites no research as authority for its assertion about the abilities of judges and juries. In this regard, however, *Kuhn* is no different from most opinions. The next Section discusses what evidence we have in support of *Kuhn's* view, or the opposite view that restrictions on the admissibility of expert evidence may improve outcome accuracy.

The reader should be warned, however, that the existing research rarely addresses the exact questions for which we need answers. There is no research that addresses the ultimate question of whether, in the aggregate, cases are more frequently decided correctly under various admissibility regimes. This does not mean the research results are irrelevant, for even if they do not answer the ultimate question, they do narrow the range of reasonable disagreement.

IV. EMPIRICAL EVIDENCE RELATING TO PATERNALISTIC ARGUMENTS

Although there are many ways one might organize the research relevant to the paternalistic argument, this article employs Goldman's five factors as a template.⁸⁸

A. *The Characteristics of the Audience*

The heart of the argument for restrictions turns on the characteristics of the audience, that is the jurors. As Goldman and others recognize, these characteristics are shaped not only by qualities that individuals bring to the courthouse, e.g. their educational attainment, demographic characteristics, and life experiences, but also by the forum itself. They are not simply a set of individuals, rather they are a jury. As a jury, they are constrained in many ways. They cannot go to the library at night to do independent reading on a topic in the trial. They cannot ask the parties for a standard textbook that they might examine. In most jurisdictions, they cannot even ask for clarification of ambiguous points. As Mirjan Damaska notes "jurors have no proof initiative and are usually not even permitted to ask questions of witnesses. While evidence is being

⁸⁸ *Id.* at 124. I have rearranged the order in which Goldman discusses the factors.

adduced, they sit silent, cast—one might say—into the role of potted courtroom plants.”⁸⁹ Therefore, while research on the abilities of individuals is of some relevance,⁹⁰ better evidence on the characteristics of the audience must come from research about people placed in the role of a juror.⁹¹

If there is one universal finding in jury research, it is that juries take their job very seriously. If they fail to arrive at appropriate results, normally it is not due to a lack of effort.⁹² Shortcomings, therefore, are best thought of as caused by some combination of lack of ability among jury members and by the way in which the jury receives information.

The literature on jury performance in general is voluminous.⁹³ I focus on articles that at least tangentially touch on issues related to expert witnesses. In this regard, it may be helpful to break the

⁸⁹ MIRJAN R. DAMASKA, EVIDENCE LAW ADRIFT 90 (1997). Recent efforts in Arizona and elsewhere to empower jurors are slowly changing this pattern. See Nancy Marder, *Juries and Technology: Equipping Jurors for the Twenty-First Century*, 66 BROOK. L. REV. 1257 (2001); see also Paula L. Hanaford et al., *Permitting Jury Discussions During Trial: Impact of the Arizona Reform*, 24 L. & HUMAN. BEHAV. 359 (2000); Larry Heuer & Steven Penrod, *Juror Notetaking and Question Asking During Trials*, 18 L. & HUMAN BEHAV. 121 (1994).

⁹⁰ In the survey context, the public often expresses views contrary to those held by experts. For an example of this phenomenon, see Nancy Kraus et al., *Intuitive Toxicology: Expert and Lay Judgments of Chemical Risk*, 12 RISK ANALYSIS 215 (1992), reporting on expert and lay opinions regarding a number of matters concerning risk. One question asked respondents whether they agreed or disagreed with the following statement:

Residents of a small community (30,000 people) observed that several malformed children had been born there during each of the past few years. The town is in a region where agricultural pesticides have been used during the past decade. It is very likely that these pesticides were the cause of the malformations?

Id. at 221. The results were as follows: among toxicologists only 5% agreed, while 48% of lay people agreed. *Id.*

⁹¹ Even within the area of jury research, there is wide variation in the quality of a study design. Most jury research is done in the laboratory and ranges from full blown trial simulations to paper and pencil exercises conducted with college sophomore subjects. Generally, the more realistic the simulation the more faith we may have that it captures the actual jury experience and, therefore, suffers from fewer external validity threats.

⁹² Exceptions to this rule include the very short deliberation of the jury in the O.J. Simpson criminal trial.

⁹³ For useful discussions of jury research, see Robert MacCoun, *Inside the Black Box: What Empirical Research Tells Us about Decisionmaking by Civil Juries*, in VERDICT: ASSESSING THE CIVIL JURY SYSTEM 137, 152 (Robert E. Litan ed., 1993); Michael J. Saks, *What do Jury Experiments Tell Us About How Juries (should) Make Decisions?*, 6 S. CAL. INTERDISC. L.J. 1 (1998); Neil Vidmar, *The Performance of the American Civil Jury: An Empirical Perspective*, 40 ARIZ. L. REV. 849 (1998); NEAL FEIGENSON, LEGAL BLAME: HOW JURORS THINK AND TALK ABOUT ACCIDENTS (2000).

research into two parts: first, the ability of juries to deal with complex cases; and second, juries' ability to assess expert testimony uninfluenced by other, non-epistemic factors that might affect judgment.⁹⁴

1. The Jury and Complex Cases

Evidence suggests that, while juries are competent in sorting out facts in simple cases,⁹⁵ their ability in more complex cases is open to question.⁹⁶ Complexity may arise due to the difficulty of the issues being considered, or due to the sheer volume of information the jury is asked to consider.⁹⁷ Both types of complexity appear to give juries difficulty.

Laboratory research reveals that mock jurors have trouble with statistical and probabilistic evidence and tend to underutilize statistical information in the sense that they fail to weigh it properly.⁹⁸

⁹⁴ These issues overlap in actual trials. Cases which present complex factual questions inevitably are cases with expert witness testimony.

⁹⁵ REID HASTIE ET AL., *INSIDE THE JURY* (1983); see also Phoebe C. Ellsworth, *Are Twelve Heads Better Than One?*, 52 LAW & CONTEMP. PROBS. 205, 217-18 (1989); Vidmar, *supra* note 93, at 853-54.

⁹⁶ See Joe S. Cecil et al., *Citizen Comprehension of Difficult Issues: Lessons From Civil Jury Trials*, 40 AM. U. L. REV. 727, 750-64 (1991); see also Richard Lempert, *Civil Juries and Complex Cases: Taking Stock After Twelve Years*, in VERDICT: ASSESSING THE CIVIL JURY SYSTEM 181 (Robert E. Litan ed., 1993); JOSEPH SANDERS, *BENEDICTIN ON TRIAL* 117-30 (1998).

⁹⁷ The sheer volume of information may make a case complex, or at least complicated. A second dimension of complexity is the technical difficulty of the evidence. The most "complex" trials are those that contain large quantities of various types of technical and scientific evidence. However, even short trials that contain substantial amounts of technical, scientific, or statistical evidence may be difficult for lay persons. Indeed, in such situations, a trial may be too short to convey a full understanding of the evidence. Irwin Horowitz et al., *The Effects of Complexity on Jurors' Verdicts and Construction of Evidence*, 86 J. OF APPLIED PSYCHOL. 641, 649 (2001), found that information load (the number of facts in the case) adversely affected simulated juries' ability to sort out those facts having the greatest probative value.

⁹⁸ See David L. Faigman & A.J. Baglioni, Jr., *Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence*, 12 LAW & HUM. BEHAV. 1 (1988); see also Brian C. Smith et al., *Jurors' Use of Probabilistic Evidence*, 20 LAW & HUM. BEHAV. 49 (1996); Thompson & Schumann, *Interpretation of Statistical Evidence in Criminal Trials*, 11 LAW & HUM. BEHAV. 167 (1987); William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence?*, 52 LAW & CONTEMP. PROBS. 9 (1989); Jane Goodman, *Juror's Comprehension and Assessment of Probabilistic Evidence*, 16 AM. J. TRIAL ADVOC. 361 (1992); Jonathan Koehler et al., *The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?*, 37 JURIMETRICS J. 425 (1995); Jason Schklar & Shari S. Diamond, *Juror Reactions to DNA Evidence: Errors and Expectations*, 23 LAW & HUM. BEHAV. 159 (1999); Nance & Morris, *supra* note 78. In general, laypersons have difficulty applying statistical concepts to everyday behavior. Richard Nisbett et al., *Teaching Reasoning*, 238 SCIENCE 625 (1987).

For example, in one study by Jane Goodman, jurors were asked to assess guilt of a defendant where his blood type was said to match that found at the crime scene. Jurors in some versions of the experiment were told the blood type matched 5% of the people in town.⁹⁹ In other versions, they were told it matched 1%, and in still other versions 0.1%. A control group was provided no frequency information.¹⁰⁰ The mean estimate of guilt was higher in groups that heard the probabilistic evidence. However, the jurors' estimates "failed to make fine distinctions between probability estimates that were mildly incriminating, moderately incriminating, and strongly incriminating."¹⁰¹ Mock jurors with prior mathematical experience gave more weight to this evidence, lending some credence to the idea that juries comprised of more knowledgeable individuals, sometimes called "blue-ribbon" juries, would do better in complex cases.¹⁰²

Molly Treadway Johnson conducted a set of experiments examining the ability of members of a jury pool to make correct causal judgments based on their understanding of fictitious epidemiological studies.¹⁰³ In one experiment, subjects were given a pair of two-by-two tables presenting typical epidemiological findings. In Table 1, there was a statistically significant association between exposure and disease where the relative risk was 2.8—indicating that if one is exposed, one has slightly less than three times as great a chance of having the disease than if one is not exposed. In Table 2, the relationship was not statistically significant, the relative risk was 1.01—indicating that if one is exposed one has a just slightly greater chance of having the disease than if one is not exposed (if the relative risk were 1.0, the likelihood of having the disease would be unaffected by exposure). For each table, Johnson asked the following two questions:

1) do the results of the study indicated that being exposed to [substance x] increases a person's risk of developing [a certain abnormality], and 2) for any particular person who was exposed to [substance x] and now has [the abnormality], is it more likely than not that it was the [substance], rather than something else, that caused the [abnormality]?¹⁰⁴

⁹⁹ Goodman, *supra* note 98, at 361.

¹⁰⁰ *Id.* at 369.

¹⁰¹ *Id.* at 372. In the 5% condition, the mean estimate of guilt was 40%, in the 1% condition it was 45%, and in the 0.1% condition it was 47%. *Id.* at 371.

¹⁰² *Id.* at 389.

¹⁰³ Molly T. Johnson, *An Investigation of Juror Comprehension of Statistical Proof of Causation* (1990) (unpublished Ph.D. dissertation) (on file with author).

¹⁰⁴ *Id.* at 51.

Twenty-five respondents answered the question, so that there were a total of 100 possible correct answers (two “yes’s” for the first table and two “no’s” for the second table). Altogether, the respondents gave forty-one correct answers; only two subjects answered all four questions correctly.¹⁰⁵

In this experiment, the subjects were unassisted in their task. In a second experiment, some subjects viewed an eighteen-minute videotape of an epidemiologist explaining the tables and how epidemiologists analyze and interpret epidemiological data followed by jury instructions.¹⁰⁶ A control group heard only the instructions. Surprisingly, subjects did about as well in this second experiment, and there was no statistically significant difference in the frequency of correct responses between the subjects who heard the expert testimony and those who did not.¹⁰⁷

Johnson conducted a third experiment in which Experiment I was replicated using college students rather than people from the local jury pool as subjects. The college students did substantially better.¹⁰⁸ Altogether, twenty-five college students gave sixty-four out of a possible 100 correct answers.¹⁰⁹ Again, this result is consistent with other research indicating that the quality of jury performance depends, in part, on the abilities the jurors bring to their deliberations. Both education and occupation are corollaries of juror competence.¹¹⁰

¹⁰⁵ *Id.* at 54. All ambiguous answers, e.g. “maybe,” were coded as incorrect. Some people might argue that the “correct” answer to the second question for each Table is not obvious. If we only consider answers to the first question for each Table, then out of fifty possible correct answers the respondents answered correctly twenty-three times. *Id.* at 55. Generally, respondents tended to answer “No” and the distribution of answers to the two Tables was nearly identical. *Id.* For Table 1 there were eight “yeses,” fourteen “noes,” and three “others.” For Table 2 there were eight “yeses,” fifteen “noes,” and two “others.” *Id.*

¹⁰⁶ For an interesting study of how factors influence the odds that epidemiologists themselves will draw a causal inference based on epidemiological research, see C.D. Holman et al., *A Psychometric Experiment in Causal Inference to Estimate Evidential Weights Used by Epidemiologists*, 12 EPIDEMIOLOGY 246 (2001). Factors with the strongest influence include: level of statistical significance, refutation of alternative explanations, strength of association, number of supporting studies, and information concerning biological and theoretical coherence.

¹⁰⁷ *Id.* at 80.

¹⁰⁸ This result is consistent with research that indicates training in statistics or economics improves statistical reasoning ability. See Richard Nisbett et al., *Teaching Reasoning*, in RULES FOR REASONING (Richard Nisbett ed., 1993).

¹⁰⁹ Johnson, *supra* note 103, at 94. With respect to the first question for each table, the college students provided thirty-two out of fifty possible correct answers. *Id.* at 96.

¹¹⁰ See Fred Strodtbeck et al., *Social Status in Jury Deliberations*, 22 AM. SOC. REV. 713

It is important to note that the Johnson study did not involve a deliberation, and there is evidence that if some members of the jury are correct in their understanding of some piece of information they will be able to communicate it to other jurors. On the other hand, the Johnson study does suggest that evidence from a relatively simple epidemiological study may be “complex” in the eyes of many jurors.

Irwin Horowitz, Lynn Forester Lee, and Ian Broly examined complexity in a setting where jury-eligible adults saw a videotape of a complex civil trial.¹¹¹ The experiment varied both the information load (the quantity of evidence to be processed) and the complexity (comprehensibility), of the testimony of the witnesses. These two dimensions paralleled the trial length and scientific complexity dimensions of other studies. Horowitz, *et al.* manipulated information load by varying the number of plaintiffs and witnesses who testified. They found that high information load negatively affected juror evaluation of liability. Jurors were less able to distinguish effectively among differentially liable plaintiffs.¹¹²

The results of these experiments are substantiated by survey data and by case studies of actual trials. Daniel Shuman, *et al.* sent questionnaires to judges who presided over cases in which expert witnesses testified. Only forty-seven percent rarely thought the expert testimony was too technical for jurors to understand.¹¹³ Self-reports by jurors also indicate that some, but not all, jurors report having trouble. A Federal Judicial Center study reports that forty-six percent of jurors in long trials rated difficult or very difficult, although most said that they were able to comprehend the evidence.¹¹⁴

Steven Austin interviewed jurors in a complex predatory pricing case under the Robinson-Patman Act. He reports that the jurors were “overwhelmed, frustrated and, confused by testimony well beyond their comprehension.”¹¹⁵ Similarly, in another study, Steven Friedland describes several complex cases in which juries

(1957); *see also* HASTIE ET AL., *supra* note 95.

¹¹¹ Horowitz et al., *supra* note 97, at 757.

¹¹² *Id.* at 764.

¹¹³ Daniel W. Shuman et al., *An Empirical Examination of the Use of Expert Witnesses in the Courts—Part II: A Three City Study*, 34 JURIMETRICS J. 193, 198 (1994).

¹¹⁴ Cecil et al., *supra* note 96, at 751-53. The FJC research tends to equate complexity with trial length.

¹¹⁵ Arthur Austin, *The Jury System at Risk From Complexity, The New Media, and Deviancy*, 73 DENV. U. L. REV. 51, 54 (1995). Relevant to the issue of central versus peripheral processing discussed below, he noted that the jurors began to focus on how the lawyers were dressed and other such matters.

experienced comprehension problems.¹¹⁶

The Special Committee on Jury Comprehension of the American Bar Association's (ABA) Litigation Section conducted an in-depth study of jury decision-making in four complex cases where complexity was a matter of both length and technical difficulty.¹¹⁷ In some of the cases, the jury reported it was able to sort out the evidence, but in others jurors told researchers they had trouble deciding claims because of the large volume of data.¹¹⁸

Richard Lempert provided a useful discussion of actual jury behavior in complex cases based on a review of thirteen cases that could be considered complex either because of length or subject matter.¹¹⁹ Lempert used a three-point "defensibility scale" (high, moderate, and low) to rate each verdict on the merits. The defensibility of the decision on the merits was rated as high in seven

¹¹⁶ Steven I. Friedland, *The Competency and Responsibility of Jurors in Deciding Cases*, 85 NW. U. L. REV. 190 (1990).

¹¹⁷ SPECIAL COMMITTEE ON JURY COMPREHENSION OF THE AMERICAN BAR ASSOCIATION SECTION OF LITIGATION, JURY COMPREHENSION IN COMPLEX CASES (1989) [hereinafter SPECIAL COMMITTEE]. The cases involved antitrust, sexual harassment, misappropriation of trade secrets, and insurance fraud. *Id.* at 9-23. Although the researchers could not tape the actual jurors in the cases, they did tape simultaneous alternate jury deliberations with the objective of gaining some insight into what happened in the real jury deliberations. Unfortunately for the researchers, but revealing in its own right, in three of the four cases the alternate jury reached a different verdict than the actual jury. *Id.* at 59.

¹¹⁸ *Id.* at 25-26.

¹¹⁹ Lempert, *supra* note 96, at 181. Cases 1-4 were part of the ABA study: (1) sexual harassment, (2) antitrust, (3) insurance fraud related to an arson, and (4) misappropriation of trade secrets. The research was conducted by Elizabeth Loftus, Jane Goodman, and Edith Green. Case 5 reports on the famous *Pennzoil v. Texaco* case which involved a tortious interference with contract. Stephen Adler, *How to Lose the Bet-Your-Company Case*, AM. LAW. 27-30, 107-10 (Jan./Feb. 1986). Cases 6 and 7 are the trial and retrial of the antitrust case, *Cleveland v. Cleveland Electric Illuminating Co.*, studied by Austin. See ARTHUR AUSTIN, COMPLEX LITIGATION CONFRONTS THE JURY SYSTEM: A CASE STUDY (1984). Cases 8-11 are reports of notable trials in *The American Lawyer* by a number of different authors. Case 8, Stephen Brill, *Inside the DeLorean Jury Room*, AM. LAW. 94-105. (Dec. 1984), reviews the criminal conspiracy trial of John DeLorean. Case 9, Allison Frankel, *He Bombed as a Stool Pigeon*, AM. LAW. 35-89. (May 1989), discusses *United States v. GAF Corp.*, a stock manipulation case. Case 10 Gay Jerve, *Charlie Keating, Meet Your Peers*, AM. LAW. 101-09. (March 1992), reports on *California v. Keating*, the prosecution of Charles Keating for stock fraud. Case 11, Mitchell Pacelle, *Contaminated Verdict*, AM. LAW. 75-80 (Dec. 1986), writes about a toxic tort case against W.R. Grace & Co and Beatrice Corp. for leukemia allegedly caused by their dumping toxic substances that seeped into the ground water used as drinking water by plaintiffs. Case 12 is a Rand report of an asbestos case. MOLLY SELVIN & LARRY PINKUS, THE DEBATE OVER JURY PERFORMANCE: OBSERVATIONS FROM A RECENT ASBESTOS CASE (1987). Case 13, Claudia Weinstein, *Losing Big in Computer Land*, AM. LAW. 123-30 (Oct. 1983), discusses *Micro/Vest v. Computerland Corp.*, a breach of contract case.

cases, moderate in four cases, and low in two cases.

In an attempt to separate cases that are “complex” because of their subject matter, Lempert rates each case according to his own sense of the inherent difficulty of the evidence, once again on a three point scale: high, moderate, and low.¹²⁰ Of the six cases scored low on inherent difficulty, four juries reached highly-defensible verdicts and two reached moderately-defensible verdicts. Of the seven trials rated moderate or high on difficulty, three juries reached highly defensible verdicts, two reached moderately defensible verdicts, and two scored low on defensibility. The one jury for which the evidence difficulty was rated high, but whose verdict on the merits was rated as highly defensible, found damage amounts that Lempert rated low on defensibility. Overall, one can conclude from this study that jury performance is more likely to be a problem in those cases that are “complex” because of the technical nature of the evidence.

I reached a similar conclusion after interviewing jurors about the testimony in a complex evidence case involving the drug Bendectin. I concluded that the jurors had a weak grasp of the science resulting in an indefensible verdict.¹²¹

These studies are consistent with the observation of Joe Cecil and his colleagues, that the most difficult type of complex evidence is that containing statistical and technical information.¹²² There is some research to support the suggestion that the difficulty jurors have in assigning appropriate weight to this evidence is exacerbated when the evidence is of low probative value, that is when it does not change the probability of a causal relationship by very much.¹²³

¹²⁰ A case was scored high on difficulty “where a large amount of hard to understand, unfamiliar scientific information bore on the central issue in the case.” Lempert, *supra* note 119, at 189. It was scored moderate or low where “technical or specialized information seemed somewhat easier to understand or where full understanding seemed less crucial to correct decision making, because the evidence was not so central, or because it was redundant with other easier to understand evidence or because vaguer understanding would suffice.” *Id.*

¹²¹ Joseph Sanders, *The Jury Deliberation in a Complex Case: Havner v. Merrell Dow Pharmaceuticals*, 16 JUST. SYS. J. 45 (1993).

¹²² See Cecil et al., *supra* note 96, at 757-58 (1991). See, e.g., Thompson & Schumann, *supra* note 98, at 167; Thompson, *supra* note 98, at 9.

¹²³ See Paul Slovic et al., *What Should We Know About Making Risk Comparisons?*, 10 RISK ANALYSIS 389 (1990); see also Brandon B. Johnson, *Stability and Inoculation of Risk Comparisons' Effects Under Conflict: Replicating and Extending the "Asbestos Jury" Study by Slovic et al.*, 22 RISK ANALYSIS 777 (2002). A similar finding indicates the differences in low probabilities has little impact on decisions. Cass R. Sunstein, *Probability Neglect: Emotions, Worst Cases, and Law*, 112 YALE L.J. 61, 71-73 (2002); see also W. Kip Viscusi, *Jurors, Judges, and the Mistreatment of Risk by the Courts*, 30 J. LEGAL STUD. 107 (2001). The ability to assess risk is further eroded in situations that are emotionally highly

An examination of the cases Lempert scored as “high” on difficulty reveals that when confronted with complex technical and scientific evidence from competing experts, jurors have a difficult time assessing the merits of the testimony.¹²⁴ Jurors are often confronted with a “battle of the experts,” each of whom is chosen for the ability to be convincing and appear credible.¹²⁵ This point leads to the second body of data that directly addresses how juries respond to experts.

2. The Jury and Expert Witnesses

It is a common misperception that juries are overwhelmed by experts simply because they are experts.¹²⁶ The ABA study found, for example, that the jurors in their four cases were not overly impressed with the experts, and dismissed many of them as “hired guns.”¹²⁷ In fact, the Committee concluded that a witness who is perceived to be a “hired gun” can do positive harm to a party’s case.¹²⁸

charged. See Reid Hastie, *Emotions in Juror’s Decisions*, 66 BROOK. L. REV. 991 (2001) (summarizing the relevant psychology literature).

¹²⁴ The three were the ABA trade secrets case, the asbestos exposure case, and the ground water contamination case. Pacelle, *supra* note 199, at 75-80; SELVIN & PINKUS, *supra* note 119, at 4.

¹²⁵ See Anthony Champagne et al., *An Empirical Examination of the Use of Expert Witnesses in American Courts*, 4 JURIMETRICS J. 375 (1991); see also Samuel Gross, *Expert Evidence*, 1991 WISC. L. REV. 1113 (1991).

¹²⁶ One might be tempted to conclude, as apparently the *Kuhn* court did, that if a jury is not overwhelmed by a piece of testimony there is no need for a reliability filter. This does not logically follow. A hypothetical example may help to demonstrate this point. Assume that the reliability of expert evidence could be measured on a scale of one to ten, with the most reliable evidence scored ten. Also assume the jury is presented with a piece of evidence with a score of one on our reliability scale. If jurors are overwhelmed by all expert testimony, this would mean, one presumes, that they would believe that this evidence is much more reliable, say eight on our scale. Clearly, this would cause them to mis-estimate the evidentiary value of the evidence. Moreover, if jurors were routinely overwhelmed by all expert testimony, they could make no distinctions between any evidence. Evidence with a score of four on our hypothetical reliability scale would be considered as good as evidence actually scoring an eight and no better than the evidence with a score of one. However, even if jurors are not overwhelmed, this does not mean they will properly weigh the reliability of evidence. They might simply disregard all expert evidence, effectively giving it a reliability score of one, or they could give all evidence a five, and thus still fail to distinguish between evidence scoring eight, evidence scoring four, and evidence scoring one.

¹²⁷ SPECIAL COMMITTEE, *supra* note 117, at 40. This finding agrees with other research on the impact of experts. See studies summarized in Neil Vidmar, “Assessing the Impact of Statistical Evidence, A Social Science Perspective,” in THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS 279 (Stephen Fienberg ed., 1989).

¹²⁸ SPECIAL COMMITTEE, *supra* note 117, at 42; see also Scott Sundby, *The Jury as*

Daniel Shuman and his colleagues conducted a group of studies and concluded that jurors do not mechanically defer to experts because of their expertise. Rather, they are far more skeptical in their assessments.¹²⁹

Molly Selvin and Larry Pinkus report a general skepticism, if not a negative disposition, by jurors toward the experts in the asbestos case.¹³⁰ Apparently, a frequent jury response to difficult scientific issues is to downplay the importance of the experts and their testimony. For example, Shari Diamond and Jonathan Casper, reporting on a laboratory study of jury decision-making in a complex antitrust case, noted that "lack of clarity, that is, perceived complexity and difficulty, discourages the jurors from accepting an expert's position, rather than inducing them to accept it."¹³¹ Mitchell Pacelle's report of the ground water contamination case suggests that jurors answered some questions with little reference to the science introduced to address the issue.¹³² Perhaps the clearest statement of this view is to be found in the following comment made by a juror in an asbestos case studied by Goodman, Green, and Loftus: "[t]he expert testimony was not a real factor in our decision, except in the very backhanded sense that it lent medical credence to any result."¹³³

The evidence that jurors do not overrate experts does not mean that jurors always understand experts. In the face of difficult, conflicting expert testimony, jurors have three options. First, as the quote above suggests, is simply to disregard the testimony and decide the case on other grounds. Second, is to slog through the testimony and assess the testimony on the merits. Third, is to rely on other indicia of whom to believe. Determining when individuals will choose between the second and third alternatives is the subject of a

Critic: An Empirical Look at How Capital Juries Perceive Expert and Lay Testimony, 83 VA. L. REV. 1109 (1997) (providing a concurring opinion).

¹²⁹ Daniel Shuman & Anthony Champagne, *Removing the People from the Legal Process: The Rhetoric and Research on Judicial Selection and Juries*, 3 PSYCHOL. PUB. POL'Y & LAW 242, 258 (1997).

In a valuable article published too late for me to include, Ivkovic and Hans discuss juror evaluations of experts in seven cases. Sanja Kutnjak Ivkovic & Valerie P. Hans, *Jurors' Evaluations of Expert Testimony: Judging the Messenger and the Message*, 28 LAW & SOC. INQUIRY 441 (2003).

¹³⁰ SELVIN & PINKUS, *supra* note 119, at 77.

¹³¹ Shari Diamond & Jonathan Casper, *Blindfolding the Jury to Verdict Consequences: Damages, Experts and the Civil Jury*, 26 LAW & SOC'Y. REV. 513, 543 (1992).

¹³² Mitchell Pacelle, *Contaminated Verdict*, AM. LAW. 75, 80 (Dec. 1986).

¹³³ Jane Goodman et al., *What Confuses Jurors in Complex Cases*, TRIAL 65, 68 (Nov. 1985); *see also* Sanja Ivkovich & Valerie Hans, *Jurors and Experts*, 16 ADVOCATE: THE MAGAZINE FOR DELAWARE TRIAL LAWYERS 17 (1994).

body of research on decision making in settings that involve persuasion attempts. This research indicates that people employ two basic cognitive processes to assist them in decision making: central processing and peripheral processing.¹³⁴

In central or systematic processing, people examine the content of a communication to assess its validity. Persuasion is primarily a function of the quality of the arguments presented.¹³⁵ In peripheral or heuristic processing, people do not attend to the quality and validity of arguments. Rather, they adopt shortcuts to determine the value of a message. People rely on factors such as the number of arguments (rather than their quality), the attractiveness of the communicator, and the communicator's credentials.

Joel Cooper, Elizabeth Bennett, and Holly Sukel conducted a laboratory experiment on the effects of trial complexity on juror assessment of expert testimony in a laboratory setting.¹³⁶ The subjects viewed a one-hour videotape of a civil case involving a person who claimed that exposure to polychlorinated biphenyls (PCBs) caused his cancer. In the video, the defendant conceded that the plaintiff had been exposed, thus the only factual issue was whether the exposure caused the illness.¹³⁷ The evidence on this question was presented by two expert witnesses, one for each side. The experiment manipulated the linguistic complexity of the testimony of one of the experts.¹³⁸ The experiment also manipulated the expert's

¹³⁴ See Alice Eagly & Shelly Chaiken, *Attitude Structure and Function*, in 1 HANDBOOK OF SOC. PSYCHOL. 269 (Daniel Gilbert et al. eds., 1998); see also Richard Petty & Duane Wegener, *Attitude Change: Multiple Roles for Persuasion Variables*, in 1 HANDBOOK OF SOC. PSYCHOL. 323 (Daniel Gilbert et al. eds., 1998); RICHARD E. PETTY & JOHN T. CACIOPPO, COMMUNICATION AND PERSUASION: CENTRAL AND PERIPHERAL ROUTES TO ATTITUDE CHANGE (1986); Shelly Chaiken, *Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion*, 39 J. PERS. SOC. PSYCHOL. 752 (1980).

¹³⁵ Joel Cooper et al., *Complex Scientific Testimony: How Do Jurors Make Decisions?*, 20 LAW & HUM. BEHAV. 379, 381 (1996).

¹³⁶ *Id.*

¹³⁷ *Id.* at 384.

¹³⁸ For example, in the low complexity condition the expert made statements such as this: "In the rats and mice, PCBs caused not only liver disease but also cancer of the liver. In additional [sic] to the liver damages, [the study] found diseases of the immune system as well." *Id.* at 385. In the high complexity condition, the expert made statements such as this: "[The study] reported tumor induction in rats and mice. [The study] also reported that not only rats and mice, but in monkeys as well, there was hepatomegaly, hepatomegalocytosis and lymphoid atrophy in both spleen and thymus." *Id.* The testimony of the other expert was "complex" in all conditions. *Id.* The manipulation of linguistic complexity rather than actual scientific or technical complexity is a shortcoming of this study. In actual trials, experts are coached to reduce their arguments to simple terms that can be understood by lay

credentials.¹³⁹ In the high complexity condition, ninety-one percent of the jurors voted for the plaintiff when the case was presented by an expert with high credentials, but only sixty-four percent of the jurors voted for the plaintiff when the case was presented by an expert with low credentials. This difference is statistically significant.¹⁴⁰ In the low-complexity condition, the effect of expert credentials was not significant. However, jurors were more likely to vote for the plaintiff when the testimony was presented by the expert with low credentials, and there was a significant interaction between the level of complexity and the strength of credentials.¹⁴¹ This interaction was also significant when the jurors were asked to estimate the probability that PCBs were the cause of the plaintiff's illness.¹⁴² In the high-complexity case, jurors who heard the highly-credentialed expert concluded, on average, that there was a ninety-six percent probability that the PCBs caused the cancer. On the other hand, jurors who heard the low credentialed expert concluded, on average, that it was only forty-nine percent probable that this was the case.¹⁴³

In a second article, Joel Cooper and Isaak Neuhaus reported three experiments that assessed the effect of high pay and frequent testifying on jury judgments of experts at trial.¹⁴⁴ In the first experiment, a highly credentialed expert who was paid \$4,800 a day had less influence on jurors than either an expert who was paid less, an expert with lesser credentials, or both. A second experiment

people. Thus, the circumstances of this manipulation are unlikely to occur in actual trials. The study is also limited by the fact that the "jurors" did not deliberate.

¹³⁹ In the "high" condition, the expert had advanced degrees from highly prestigious universities, was currently teaching and conducting research at a similar institution, and had published many articles on cancer in peer-reviewed journals. In the "low" condition, the expert received a degree from a relatively obscure institution, taught at a large state university, and had published far fewer articles. Subjects perceived the individual with high credentials to have more expertise. Cooper et al., *supra* note 135, at 386.

¹⁴⁰ *Id.* at 387.

¹⁴¹ *Id.*

¹⁴² *Id.* at 388.

¹⁴³ *Id.* Vidmar and Diamond challenge Cooper's interpretation of this result. Neil Vidmar & Shari S. Diamond, *Juries and Expert Evidence*, 66 BROOK. L. REV. 1121, 1154-55 (2001). They note that the Cooper study tested juror comprehension of the testimony and jurors in all four experimental conditions indicated good comprehension. They argue that the strong and consistent performance on comprehension is important because it suggests that jurors centrally processed the testimony and did not rely simply on the more impressive educational and professional background of the highly credentialed expert.

¹⁴⁴ Joel Cooper & Isaak M. Neuhaus, *The "Hired Gun" Effect: Assessing the Effect of Pay, Frequency of Testifying, and Credentials on the Perception of Expert Testimony*, 24 LAW & HUM. BEHAV. 149 (2000).

manipulated the rate of pay and expert witness testifying experience independently. The highly paid expert was found to be less persuasive when he also had a substantial history of testifying. In a third experiment, Cooper and Neuhaus manipulated expert pay and the complexity of the expert's language. All the experts used in this study were experienced and highly credentialed. Again, there was an interaction effect. The expert using highly complex language and paid the top rate was less persuasive than the experts in the other conditions (less complex testimony and/or paid less). The jury also saw this expert as the least honest. Cooper and Neuhaus concluded that experts who are highly paid and who testify frequently are perceived as hired guns who are neither liked nor believed.¹⁴⁵ As in their earlier study, Cooper and Neuhaus interpret these results as supporting the shift to peripheral processing when faced with complex, cognitively challenging testimony.

As they did with respect to Cooper's earlier study, Vidmar and Diamond question whether these results are best explained as an example of peripheral processing.¹⁴⁶ They argue that a competing explanation is that the jurors centrally processed the testimony, but rejected it when the expert's motives were suspect. Vidmar and Diamond's disagreement with the Cooper and Neuhaus interpretation does highlight two facts. First, central processing and peripheral processing are not mutually exclusive methods of thinking. Juries, like other decision makers, may engage in a combination of central and peripheral strategies when making up their mind.¹⁴⁷ Indeed, the competing explanation offered by Vidmar and Diamond suggest just such a process. Jurors may have attempted to centrally process the expert testimony, but discounted the testimony of some experts based on the peripheral information that they were highly paid and, therefore, their motive was suspect.

Second, peripheral processing can be a better or worse decision making strategy depending on the peripheral cues used to come to a decision because some cues are better indicia of reliability than others. For example, the first Cooper and Neuhaus study argued that an expert's credentials counted as a positive factor in assessing the

¹⁴⁵ *Id.*

¹⁴⁶ Vidmar & Diamond, *supra* note 143, at 1156 (2001).

¹⁴⁷ For example, Greenberg and Wursten found that simulated jurors were influenced more by medical expert testimony than by psychological expert testimony even when the testimony presented by the experts was identical in form and complexity. J. Greenberg & A. Wursten, *The Psychologist and the Psychiatrist as Expert Witnesses: Perceived Credibility and Influence*, 19 PROF. PSYCHOL.: RES. & PRAC. 373 (1988).

validity of their argument whereas in the second study, the comparatively excessive compensation paid to the expert acted as a negative factor. Most students of expert testimony would predict that within a given case, differences in credentials and fees are a weak predictor of validity. However, they may believe other cues are even worse, e.g. the gender of the expert.

On the other hand, they might believe that other factors such as the *Daubert* factors of general acceptance, and peer review and publication are better because they are more probative as to the reliability of expert testimony. In fact, as the following study indicates there is some research indicating that jurors may use this type of peripheral information as well.

In a laboratory experiment using undergraduate subjects to examine “juror” reasoning skills about the scientific validity of expert evidence in a gender discrimination case, Margaret Kovera and her colleagues manipulated four variables.¹⁴⁸ The first manipulation involved the publication status of the underlying research that formed the basis of the expert’s testimony. Some jurors heard that the research had been published in a peer-reviewed journal and had been cited in major psychology texts while other jurors heard the research had been submitted for publication, but not published or cited. In a second manipulation, “ecological validity” was manipulated.¹⁴⁹ Some jurors heard that the subjects in the underlying study had been college students and others that they had been blue collar employees in a company similar to the defendant company. A third manipulation involved the “construct validity” of the underlying study.¹⁵⁰ Some jurors heard that there was but a single measure of sexual harassment while others heard that there were multiple measures. A fourth manipulation varied the quality of the cross-examination. In the scientifically uninformed version, the cross was restricted to general questions such as whether the expert was qualified to make a legal determination regarding the occurrence of sexual harassment. In the informed version, the cross drew attention

¹⁴⁸ Margaret B. Kovera et al., *Reasoning About Scientific Evidence: Effects of Juror Gender and Evidence Quality on Juror Decisions in a Hostile Work Environment Case*, 84 J. APPLIED PSYCHOL. 362 (1999).

¹⁴⁹ Ecological validity is a species of external validity. External validity involves the ability to generalize conclusions to *particular* persons, settings, and times and to *types* of persons, settings, and times. See THOMAS D. COOK & DONALD T. CAMPBELL, *QUASI-EXPERIMENTATION: DESIGN AND ANALYSIS ISSUES FOR FIELD TESTING* 71 (1979).

¹⁵⁰ Construct validity problems may arise when one has a single operationalization of a cause or an effect: a mono-operation bias. More reliable research includes multiple operationalizations. COOK & CAMPBELL, *supra* note 149, at 67.

to the issues of ecological and construct validity.

The subjects rated the expert evidence as more valid if it had been published.¹⁵¹ None of the other manipulations affected validity judgments, but the jurors did judge the expert as more credible if she studied the responses of individuals who resembled the plaintiff's coworkers.¹⁵² Jurors were not sensitive to variations in the construct validity manipulation. This was the case whether or not the jurors heard the scientifically sophisticated cross-examination.¹⁵³

In my judgment, the weight of the experimental research suggests that jurors do engage in peripheral processing when assessing expert testimony and that the peripheral cues take on added significance as the scientific issues in the case become more complex.¹⁵⁴

Trial lawyers tend to agree with at least the first part of this conclusion. Sanders, Diamond, and Vidmar conducted two separate focus groups with groups of Texas trial lawyers who concentrate on products liability cases.¹⁵⁵ One focus group was with plaintiff lawyers and one with defense counsel. Defense lawyers expressed concerns about the impact of many personal factors on the effectiveness of expert testimony. These factors included, inexperience on the stand,¹⁵⁶ high fees,¹⁵⁷ where the expert is from,¹⁵⁸ presentation style,¹⁵⁹ and other personal factors.¹⁶⁰ For example, one defense lawyer made the following comment:

¹⁵¹ Kovera et al., *supra* note 148, at 370.

¹⁵² *Id.*

¹⁵³ *Id.* at 372.

¹⁵⁴ See also Daniel A. Krauss & Bruce D. Sales, *The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Sentencing*, 7 PSYCHOL. PUB. POL'Y & L. 267 (2002).

¹⁵⁵ Joseph Sanders et al., Trial Lawyer Perceptions of Expert Knowledge (the Law and Society Association Meeting in Miami, Florida) (on file with author).

¹⁵⁶ Transcript of Defense Lawyer Focus Group, at 43 (March 1999) (on file with author).

¹⁵⁷ *Id.* at 14, 26.

¹⁵⁸ *Id.* at 25.

¹⁵⁹ *Id.* at 6.

¹⁶⁰ *Id.* at 57. One lawyer commented on an expert who:
is such a good doctor, he is so analytical, he lays the case out, he goes through the medical records with a fine tooth comb. He gives you every variation, every theory, but the guy, I mean I hate to say it, he's the most unattractive person in the world. You just wouldn't put him in front of the jury. He's got a bookish appearance and just an unattractive way, he talks in a monotone and just for all of those reasons, he has no appeal whatsoever as a jury expert but he's probably the best analytical expert.

Id. at 36.

I tried a case in Sierra Blanca, which is down on the Rio Grande River a little while back and you have to carefully pick your witnesses down there so that the jury can identify with them and get along with them and their manner of speech and their manner of presentation.¹⁶¹

These comments reflect the attorneys' perceptions that jurors are likely, perhaps more likely than a judge, to rely on peripheral rather than central processing in assessing the testimony of an expert.

Some defense lawyers also expressed a concern that jurors may not understand the content of expert testimony in complex cases. One lawyer noted:

We all settle cases that we strongly believe in the defense of a case from a technical standpoint because there is that lingering doubt that even though you're able to hopefully translate the technical aspect, the medical aspect to the jury, there is still that lingering doubt that the emotional aspects of the case is going to overwhelm the jury's ability to understand and process a complex case.¹⁶²

Plaintiff lawyers also discussed both the ability of jurors to properly discount extraneous factors and their ability to understand expert testimony. They, too, were concerned about how an expert will appear before the court. As one attorney noted, "there have been cases where somebody, after I met them, I said 'you know, they are just not going to look too shiny and I need to go get a show dog and do a handoff.'"¹⁶³ Generally, the lawyers agreed that if one has a bad witness, it is rarely worth the effort to send him to "charm school." One lawyer noted, however, that this is a more common practice on the defense side. At an earlier point in his career, when he represented large corporations, the lawyer noted:

it was not unusual at all to send an expert that's being used nationwide to witness school in California and they became very polished. They learned that when you answer a question you look at the jury when you answer it. And they learned how you can occasionally slip something into your testimony and how to sound sincere.¹⁶⁴

¹⁶¹ Transcript of Defense Lawyer Focus Group, at 6.

¹⁶² *Id.* at 73-74.

¹⁶³ Transcript of Plaintiff Lawyer Focus Group, at 36.

¹⁶⁴ *Id.* at 42. These comments suggest that lawyers are concerned that jurors might be influenced by factors such as presentation style. But when asked directly whether jurors can understand expert testimony, the plaintiff lawyers responded with an adamant "yes." A few comments provide a flavor of this fervor: "Around this table everybody will give you the same answer. Yeah, we trust juries, we believe in juries even when they pour us out." *Id.* at 64. ("Pour us out" is a Texas lawyer expression

There is one other body of research that addresses an issue similar to, but not quite the same as peripheral processing. That is the tendency of jurors (and potentially other fact finders), to trade off the elements of a tort in arriving at a verdict. For example, the jury may believe that the plaintiff's causation argument is weak but that the defendant's behavior was particularly egregious. Considered together, the breach of duty "makes up for" the weak causal evidence.¹⁶⁵ The evidence that this occurs comes in the form of experiments on bifurcation.

Zeisel and Callahan conducted a field experiment on the effects of bifurcation of liability and damages in the Northern District of Illinois in the early 1960s. They found that defendants prevailed in fifty-six percent of bifurcated trials, but in only thirty-four percent of unitary trials.¹⁶⁶ In a laboratory experiment, Horowitz and Bordens found that juries hearing a unitary trial were significantly more likely to find for the plaintiff (85%) than were juries that heard bifurcated trials (68%). This tendency was strongest when the bifurcated trial juries heard the general causation testimony first. If these juries did find for the plaintiff, however, their compensatory damages awards were significantly larger.¹⁶⁷

Horowitz and Bordens used tape transcripts of their juries' deliberations to explain these results. Juries that heard bifurcated trials independently decided each element of the cause of action. In

for a jury defense verdict.) "They have phenomenal ability to cut through the bullshit." *Id.* at 64.

As with the opinions of defense attorneys, it is not easy to completely reconcile these statements about jury ability with a concern that the personal attributes of an expert may cause the rejection of an otherwise sound analysis. Clearly, however, the plaintiff attorneys we spoke with view the abilities of juries more favorably than defense attorneys.

¹⁶⁵ The tendency to make this tradeoff is partly the result of a heuristic of culpable causation, i.e., the tendency to weigh the causal impute of a factor more heavily if the factor is the result of moral blameworthiness. For example, in one study a person is speeding home and collides with another car. The driver of the second car is injured. Study participants were asked to assess the degree to which the speeder's driving causally contributed to the accident. In one condition of the experiment the driver was speeding to hide an anniversary present before his spouse got home and in a second condition he was speeding to hide drugs. Subjects assigned more causation to the person in the latter condition than the former condition even though the reason for speeding has no logical connection to the degree to which speeding caused the accident. Mark D. Alicke, *Culpable Causation*, 63 J. PERSONALITY & SOC. PSYCHOL. 368 (1992).

¹⁶⁶ Hans Zeisel & Thomas Callahan, *Split Trials and Time Saving: A Statistical Analysis*, 76 HARV. L. REV. 1606, 1612 (1963).

¹⁶⁷ Irwin A. Horowitz & Kenneth S. Bordens, *An Experimental Investigation of Procedural Issues in Complex Tort Trials*, 14 LAW & HUM. BEHAV. 269, 277-78 (1990).

contrast, unitary juries failed to separate elements:

[E]vidence for matters not directly related to the issue under consideration intruded on the decision-making process. For example, when deciding general causation, unitary juries appeared to use aspects of the damages evidence. At each decision point, juries searched in other areas, especially evidence concerning damages, to buttress their decisions. This was evident when juries were faced with the most ambiguous trial issue, general causation Separated trials are structured so that the reinterpretation of the general causation evidence is less likely because many of these juries do not hear damages evidence. . . . In fact, only 25% of the juries in the separated condition hearing only causation evidence . . . found for the plaintiffs, whereas 87.5% of the unitary trial juries, which only decided causation (but heard all of the evidence), found for the plaintiffs.¹⁶⁸

Other jury experiments have generated similar results and researchers offer a similar interpretation.¹⁶⁹

Diamond and Vidmar note that reasoning skills that might improve jury performance can be taught.¹⁷⁰ It is not clear, however, how jurors would be given these skills absent the use of a court appointed master or court appointed expert. Under normal circumstances, jury errors are likely. The question remains, of course, whether restrictive admissibility rules have any beneficial affect on the error rate. The answer to that question turns in part on the evidence concerning the characteristics of the speakers (witnesses) and the characteristics of the controller (the judicial gatekeeper).

B. The Characteristics of the Speakers

Juries do not hear cases in a vacuum. The evidence comes from expert witnesses. Their characteristics interact with those of jurors in affecting the likelihood that jurors will arrive at true verdicts. Just as juror decision making is shaped by the rules of evidence and other rules surrounding deliberation processes, expert testimony is shaped by the way it is presented in American trials. In this Section, I review

¹⁶⁸ *Id.* at 282.

¹⁶⁹ Jonathan D. Casper et al., *Juror Decision Making, Attitudes, and the Hindsight Bias*, 13 LAW & HUM. BEHAV. 291 (1989). Recent articles arguing for bifurcation as a way to focus jury decision making include Steven S. Gensler, *Bifurcation Unbound*, 75 WASH. L. REV. 705 (2000), and Lars Noah, *Civil Jury Nullification*, 86 IOWA L. REV. 1601 (2001).

¹⁷⁰ Diamond & Vidmar, *Juries and Expert Evidence*, *supra* note 143, at 1135 (2001); *see also* Nisbett et al., *supra* note 108.

empirical evidence on the effect of the expert's role on expert testimony and then turn to evidence concerning the effect of the expert's demeanor.

1. The Effect of the Expert's Role on Expert Testimony

Any discussion of speaker characteristics must address the adversarial status of experts in the American system. In inquisitorial systems such as Belgium, France, Germany, and Japan, the judge plays a large role in the production of evidence. Experts are almost always court appointed and are asked to submit written reports. Parties may be given the opportunity to object to a particular expert, question the expert about the opinion rendered, or hire their own expert to rebut the court-appointed expert, but the process is relatively non-adversarial.¹⁷¹ By way of contrast, our system is one in which the parties generally select, prepare, and present experts. Court appointed experts are rare. The vast majority of experts are party witnesses.¹⁷²

A substantial body of psychological literature suggests that adopting a role affects attention to details, memory retrieval, and decision thresholds.¹⁷³ Some research on witnesses confirms this effect. In one study, Shepard and Vidmar, conducted an experiment where undergraduates viewed a slide show and heard an audio tape depicting a fight. The "witnesses" then were interviewed by an adversary or non-adversary lawyer and a week later testified about what they saw. Witnesses interviewed by the adversary lawyer biased their testimony in favor of the lawyer's client and this affected the impressions of the factual evidence and the responsibility judgments of "naive" adjudicators who did not know who had interviewed the witness.¹⁷⁴

¹⁷¹ CAROL A.G. JONES, *EXPERT WITNESSES: SCIENCE, MEDICINE, AND THE PRACTICE OF LAW* (1994); see also John Langbein, *The German Advantage in Civil Procedure*, 52 U. CHI. L. REV. 823 (1985); Howard M. Erichson, *Mass Tort Litigation and Inquisitorial Justice*, 87 GEO. L.J. 1983 (1999).

¹⁷² See Joe S. Cecil & Thomas E. Willging, *Accepting Daubert's Invitation: Defining a Role for Court-Appointed Experts in Assessing Scientific Validity*, 43 EMORY L.J. 995 (1994); see also Joe S. Cecil & Thomas E. Willging, *The Use of Court-Appointed Experts in Federal Courts*, 78 JUDICATURE J. 41 (July-Aug. 1994); Carol Krafka et al., *Judge and Attorney Experiences, Practices, and Concerns Regarding Expert Testimony in Federal Civil Trials*, 8 PSYCHOL. PUB. POL'Y & L. 309 (2002).

¹⁷³ See D. Michael Risinger et al., *The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CAL. L. REV. 1, 18-19 (2002).

¹⁷⁴ Blair H. Sheppard & Neil Vidmar, *Adversary Pretrial Procedures and Testimonial Evidence: Effects of Lawyer's Role and Machiavellianism*, 39 J. PERS. & SOC. PSYCHOL. 320 (1980).

In a follow-up study, Vidmar and Laird had students witness the same fight stimulus.¹⁷⁵ This time the experimenters manipulated the students' role simply by telling the students they would appear either as a witness of the court or as a witness for either the plaintiff or the defendant. Student "judges" who were blind to this manipulation were asked to rate whether the evidence provided by the witness favored the plaintiff or the defendant. Separately, a set of raters heard the testimony of the witnesses and rated it as more or less pro-plaintiff. For both "judges" and raters, when compared to witnesses who testified from a neutral role, witnesses who testified for an adversary party produced testimony favorable to that party.¹⁷⁶ Interestingly, when the witnesses themselves were asked to rate the evidence as pro-plaintiff or pro-defendant, their assigned role did *not* influence their judgment, i.e. the ratings of "plaintiff," "defendant," and neutral witnesses did not significantly differ from one another.

One must be careful in drawing conclusions from a study that is so weak on ecological validity. However, the results do suggest that very weak "role" manipulations can produce biasing effects even among "witnesses" who themselves have no psychological or economic interest in a given outcome. In general, experts are more likely to come to the stand with "hot biases."¹⁷⁷ Such biases are not necessarily intentional but they are directionally motivated. The experts are more likely to want a certain outcome to prevail.¹⁷⁸ Justice Breyer expressed this concern with respect to the plaintiff's expert's testimony in *Kumho*.¹⁷⁹

¹⁷⁵ Neil Vidmar & Nancy MacDonald Laird, *Adversary Social Roles: Their Effects on Witnesses' Communications of Evidence and the Assessments of Adjudicators*, 44 J. PERS. & SOC. PSYCHOL. 888 (1983).

¹⁷⁶ *Id.* at 893.

¹⁷⁷ Some biases are intentional, e.g. those that are the result of fraud or advocacy. Other biases may be thought of as "hot." They are often unintentional and even unconscious but they are directionally motivated because the individual expects or wants an outcome to prevail. Still other biases are "cold." They occur even in the absence of a desire for a certain outcome and in spite of a desire to achieve accuracy. Robert J. MacCoun, *Biases in the Interpretation and Use of Research Results*, in ANN. REV. OF PSYCHOL. 259, 268 (Janet T. Spence et al. eds., 1998).

¹⁷⁸ *Id.* at 268.

¹⁷⁹ Indeed, no one has argued that Carlson himself, were he still working for Michelin,

would have concluded in a report to his employer that a similar tire was similarly defective on grounds identical to those upon which he rested his conclusion here. Of course, Carlson himself claimed that his method was accurate, but, as we pointed out in *Joiner*, "nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the ipse dixit of the expert." 522 U.S. at 146.

These hot biases may in turn increase the likelihood of cold biases.¹⁸⁰ Indeed, the two may merge together to form what MacCoun describes as “warm” biases, influenced both by motivation and cognition.¹⁸¹ When federal judges are asked about problems they encounter with expert testimony, the most frequently mentioned problem is that experts abandon objectivity and become advocates for the side that hired them.¹⁸²

Kumho Tire, 526 U.S. at 157.

¹⁸⁰ Meadow and Sunstein provide a useful example of cold biases in the context of medical doctor judgements concerning bacterial meningitis in children. They note that the disease is an infection of the brain that can usually be treated with antibiotics and, in general, the sooner the antibiotics are begun the better the outcome. In a malpractice case, therefore, an important question is whether the treatment was unreasonably delayed.

The authors interviewed doctors with relevant specialties to determine how long, in their opinion, it would take to administer antibiotics to a child brought to the emergency room with relevant symptoms, and compared their response with the actual time it took to administer antibiotics to ninety-three children treated at two large university-associated pediatric centers in Chicago. The doctors, like most individuals, exhibited an optimism bias. The median estimate of fifty-four pediatric emergency room specialists was forty-six minutes and that of twenty-three pediatric infectious disease specialists was eighty minutes. But the actual median time to treatment in the two hospitals was 120 minutes, a time nearly identical to times reported in hospitals in South Carolina and California. The authors note that this bias operates independently of the pressures imposed by the adversary system. William Meadow & Cass R. Sunstein, *Statistics, Not Experts*, 51 DUKE L.J. 629, 636-39 (2001).

Anchoring effects are another example of a cold bias. Anchoring effects occur when estimates people make are influenced by arbitrary starting positions. For example, estimates of the length of the Mississippi River, the number of countries in the U.N., and the annual meat consumption of the average American, all are affected by asking people whether the number is above or below some arbitrary starting point. Michael Risinger et al., *supra* note 173, at 17.

¹⁸¹ MacCoun, *supra* note 177, at 268.

¹⁸² Carol Krafska et al., *supra* note 172, at 328.

One solution to this latter difficulty would be the greater use of court appointed experts. Some are concerned, however, that such testimony would overwhelm and unduly influence the jury. There is a certain irony in the fact that this objection is itself a paternalistic argument. It argues that juries will engage in peripheral processing by focusing on the neutral status of the expert rather than centrally processing the merits of the expert's argument. Brekke et al., conducted a laboratory experiment designed to assess this potential. Registered voters who volunteered to participate were shown a videotape re-enactment of a sexual assault trial. Nancy J. Brekke et al., *Of Juries and Court - Appointed Experts: The Impact of Nonadversarial versus Adversarial Expert Testimony*, 15 LAW & HUM. BEHAV. 451 (1991). In adversarial versions of the trial the testimony was presented by an expert introduced as a prosecution witness and examined by attorneys, while in non-adversarial versions the expert was introduced as a court-appointed expert and questioned only by the judge. In a fully crossed design, the content of the expert's testimony was also manipulated. Half the jurors heard one-sided testimony (favoring the prosecution) while others heard balanced testimony. Some subjects deliberated

Jurors are not stupid, however. They know that the experts are testifying for a party. There is some evidence that jurors give greater scrutiny to adversarial testimony than non-adversarial testimony.¹⁸³ However, witness biases are often difficult to detect. This is especially so when the expert offers only a holistic, summary judgment without a detailed discussion of the factors that produce the judgment. The use of objective, rational admissibility criteria increases the likelihood that unreliable opinions will be excluded.¹⁸⁴

in groups of six, while others did not deliberate. All subjects were asked to assess the expert's credibility and to rate the expert's testimony. The results provide mixed support to those who are concerned about the effects of court appointed experts. On the one hand, fears that a court-appointed expert would simply overwhelm jurors were not supported. Court-appointed status did not boost the expert's credibility. *Id.* at 468. On the other hand, deliberating jurors were less responsive to the content of the non-adversarial experts. For example, conviction rates varied with the content of the expert testimony in the adversarial condition, but not in the non-adversarial condition. *Id.* at 470. One explanation for this result is that non-adversarial expert testimony is subjected to less central processing, but the authors note that their study is not well-designed to test this explanation.

Note that in the above study, the subjects heard either a court-appointed expert or a party expert, but the experts were not pitted against each other. In actual trials, it would be more likely that jurors would hear both types of experts. Cooper and Hall conducted a study that did exactly this. Joel Cooper & Joan Hall, *Reactions of Mock Jurors to Testimony of a Court Appointed Expert*, 18 BEHAV. SCI. & L. 719 (2000). Undergraduates, playing the role of jurors heard testimony about a plaintiff's injury in an automobile accident. In some conditions, medical testimony was presented by party experts for each side and in other conditions an additional, court-appointed expert testified. In the cells with a court-appointed expert, half the time the expert sided with the plaintiff and half the time the expert sided with the defendant. Sometimes, the defendant was an individual, and sometimes a corporation. There was no deliberation. The jurors sided with the court-appointed expert in every condition except when the expert favored a corporate defendant.

¹⁸³ Brekke et al., *supra* note 182, at 457-58.

¹⁸⁴ See Joseph Sanders, *Kumho and How We Know*, 64(2 & 3) LAW & CONTEMP. PROBS. 373 (2001). For example, in *Hall v. Baxter Healthcare*, 947 F. Supp.1387, 1404-05 (D. Or. 1996), the district court ultimately excluded the testimony of an epidemiological expert who, at an initial admissibility hearing, stated that he was not willing to testify, based on the then existing sixteen epidemiological studies, that silicone implants more likely than not could cause systemic autoimmune disease in women. However, later, with the release of one additional abstract of an unpublished epidemiological study, the expert reported that he was prepared to change his testimony and say that it is more likely than not that implants cause systemic autoimmune disease. The abstract itself reports that it included only three women with implants and the authors of the abstract reached a different conclusion than the expert. The Judge's response to this change of position was perhaps predictable. He said in a footnote, "I find this change in so-called 'scientific opinion' not only suspect but shocking, with no scientific basis to support it. This is exactly the type of 'junk science' that the Supreme Court in *Daubert I* commanded courts to exclude." *Id.* It is not my point to argue whether this expert's testimony should be admitted, but only to note that, had the expert been asked simply to present a summary opinion about the relationship of silicone implants and autoimmune

Role effects have the potential to affect not only the testimony of the individual expert but the total body of testimony presented to the jury. A number of commentators have observed that because the experts are chosen by the parties, the system favors the selection of experts with extreme views, rather than views that are representative of the scientific community.¹⁸⁵ This may give the jurors the impression that there is less consensus in a field than actually exists.¹⁸⁶ For example, I interviewed jurors in one of the Bendectin cases. Most would agree that the research on the question of whether the drug caused birth defects in the offspring of the mothers who took it to control morning sickness clearly points in the direction of no causal relationship, and this is the position of the great majority of knowledgeable scientists. Nevertheless, the jurors generally concluded that scientists were divided equally on the issue, or that most scientists thought Bendectin was a teratogen.¹⁸⁷ Moreover, they tended to discount the epidemiological evidence and rate epidemiology in general as less probative than animal studies, and in vitro research.¹⁸⁸ This result is not surprising within the context of the trial. The jurors heard approximately an equal number of experts on each side of the issue, lending an impression that the scientific community was closely divided on the causal question.¹⁸⁹ Moreover, plaintiff's experts stressed the importance of other types of

disease, it would have been much more difficult to assess his potential bias.

¹⁸⁵ MICHAEL J. SAKS & RICHARD VAN DUIZEND, *THE USE OF SCIENTIFIC EVIDENCE IN LITIGATION* (1983).

¹⁸⁶ SANDERS, *supra* note 96, at 130.

¹⁸⁷ *Id.* at 126. This may be an example of an anchoring effect. When asked to make numerical estimates, people are strongly influenced by the initial value presented to them. Insofar as the jurors were "presented" with a similar number of plaintiff and defense experts, this anchor may have influenced their final estimate of the distribution of opinion on the causal question.

¹⁸⁸ *Id.* at 129.

¹⁸⁹ As Saks and Wissler note,

In civil litigation . . . all manner of experts are found to testify opposite their colleagues.

Whether such "balancing" of expert witnesses helps the fact-finder evaluate their testimony is another matter. The search for witnesses that is driven by the adversary process may result in a distortion of knowledge when applied to expert witnesses. For example, if 999 of every 1000 experts in a given field hold one view of a question and one holds an alternate view, the two experts who appear in court will have been detached from the extremely skewed distribution of opinion from which they were drawn. The fact finder has no way of knowing this.

Michael J. Saks & Roselle L. Wissler, *Legal and Psychological Bases of Expert Testimony: Surveys of the Law and of Jurors*, 2 BEHAV. SCI. & L. 435, 439-40 (1984).

evidence and frequently denigrated epidemiology. Consider the following exchange between the plaintiff's lawyer and a plaintiff expert testifying, on the basis of *in vitro* studies, that is a teratogen in humans.

Q: Well, Doctor, the kind of work that we've been talking about, is that what's known as hard science?

A: Yes.

Q: As opposed to soft science?

A: Yes.

Q: What is hard science?

A: Hard science is science that's experimentally based, where the data that's collected is based on procedures, protocols that have been designed to have groups that are treated with something, groups that act as controls, all of the variables are under control by the experimenter so that you administer the drugs at the same time, you administer them to animals of a given age, the timing and the environment of the experiment is under control, and, therefore, you can rely on the data that comes out of a set of experiments like this. The experiments can be done by somebody else in another laboratory and they can be confirmed if they were correct in the first place and so on

Q: What would, for example, soft science be?

A: A soft science would be thing like taking polls of people depending on their memory of circumstances.

Q: What about epidemiology; is that a soft science?

A: That's a soft science, yes.¹⁹⁰

In the face of such testimony, it is not surprising that some jurors may have a hard time coming to an appreciation about the balance of scientific opinion on an issue. I described these jurors as one-eyed fact-finders. "Far from being blind, they can see everything around them. What they lack is depth perception. All experts appear similarly qualified, all evidence of similar value and relevance."¹⁹¹

2. The Witness's Demeanor

Because the truth of a witness's statements is sometimes impossible to verify, the law has long encouraged jurors to use the witness's demeanor as a clue to veracity. This is, of course, a type of

¹⁹⁰ Testimony of Dr. Stuart Newman, at 12, *Havner v. Merrell Dow Pharmaceuticals, Inc.*, No. 88-3915-F. 214th Judicial District, Nueces County, Texas (Sept. 6, 1991) (on file with author).

¹⁹¹ *Id.* at 130.

peripheral processing. Nevertheless, in the absence of any alternative way to assess veracity, the law is simply bowing to the inevitable in approving the use of demeanor evidence as an aid to decision making. In fact, what evidence we have suggests that demeanor cues often reduce accuracy in detecting witness deception.¹⁹² Even when deception is not an issue, aspects of a witness's demeanor may mislead the jury. For example, it is well documented that jurors tend to believe that eyewitness accuracy is strongly correlated with eyewitness certainty, even though the actual correlation is quite weak.¹⁹³

In the domain of expert witness testimony, the problem is exacerbated by selection effects. Sam Gross makes this point in the following passage:

expert witnesses can become expert courtroom performers; they can learn by repeated practice to present their testimony to achieve maximum effect. Attorneys, for their part, can select expert witnesses by the same criteria—they can (and do) shop around for those experts with the best testimonial manner and the most appealing credentials, and they avoid those experts (however knowledgeable) who look bad, speak poorly, or have insufficiently impressive diplomas.¹⁹⁴

The lawyers in the focus groups discussed above report choosing experts in part because of their demeanor. All in all, as Brewer notes,

[d]emeanor is an especially untrustworthy guide where there is what we might call a lucrative “market” for demeanor itself. . . . Judges, lawyers, and commentators are thoroughly aware that lawyers choose expert witnesses at least as much because they will appear to a jury to be competent as because (in the lawyer's judgment) the experts actually are competent.¹⁹⁵

Similarly, in most cases involving complex scientific and technical

¹⁹² See Saks, *supra* note 93, at 21; see also Gerald R. Miller & Judee K. Burgoon, *Factors Affecting Assessments of Witness Credibility*, in THE PSYCHOLOGY OF THE COURTROOM 169, 182-90 (Norbert L. Kerr & Robert M. Bray eds., 1982); Joseph W. Rand, *The Demeanor Gap: Race, Lie Detection, and the Jury*, 33 CONN. L. REV. 1, 7-16 (2000) (discussing the psychological literature on the ability to detect lying by the use of demeanor cues).

¹⁹³ See DAVID FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE, *supra* note 20, ch. 15.

¹⁹⁴ Gross, *supra* note 125, at 1133. Gross cites a number of “how to” books recommending the selection of witness based in part on their demeanor. He cites a 1967 survey of judges, lawyers and doctors in the Los Angeles area that found that “[o]ver three-quarters of the attorneys responding . . . indicated that some factor other than medical expertise—usually an impressive ‘courtroom manner’—often determines the choice of an expert witness.”

¹⁹⁵ Scott Brewer, *Scientific Expert Testimony and Intellectual Due Process*, 107 YALE L.J. 1535, 1622-23 (1998).

evidence, the academic and professional credentials of the various experts are sufficiently impressive and sufficiently similar that they offer little useful information to a juror who might wish to assess the merits of an argument based on the qualifications of its proponent.¹⁹⁶ In sum, the credentials and demeanor of an expert witness are unlikely to be of much probative value to jurors in cases involving complex issues.

C. The Characteristics of the Controller (Gatekeeper)

Articles about jury competence often note that it would be wrong to assume that judges, acting as triers of fact, would do much better. Anecdotal evidence is sometimes to the contrary. For example, Gross cites *Wells v. Ortho Pharmaceutical Corp.*¹⁹⁷ By agreement of the parties, the case was tried by the judge rather than a jury, who found for the plaintiff in her claim that her mother's use of a contraceptive spermicide caused her birth defects. In his opinion, Judge Shoob explained his verdict as follows:

The Court's decision, therefore, turned on the oral testimony of a variety of expert witnesses whose opinions often were diametrically opposed on the major issues presented in the case. In assessing the credibility of these witnesses, the Court considered each expert's background, training, experience, and familiarity with the circumstances of this particular case; and the Court evaluated the rationality and internal consistency of each expert's testimony in light of all the evidence presented. The Court paid close attention to each expert's demeanor and tone. Perhaps most important, the Court did its best to ascertain the motives, biases, and interests that might have influenced each expert's opinion.

With few exceptions, the Court found the testimony of plaintiffs' experts generally to be competent, credible, and directed to the specific circumstances of this case. The testimony of defendant's experts, in contrast, often indicated bias or inconsistency.¹⁹⁸

The judge then proceeded to provide a bill of particulars concerning his judgment about individual plaintiff and defense experts. Unfortunately, the judge was wrong. Relying on the same considerations that may send a jury astray, demeanor and prejudice, the judge ruled for the plaintiff even in the face of an FDA report

¹⁹⁶ See SANDERS, BENEDICTIN ON TRIAL, *supra* note 96, at 120-22.

¹⁹⁷ 615 F. Supp. 262 (N.D. Ga. 1985), *aff'd and modified in part*, 788 F.2d 741 (11th Cir. 1986).

¹⁹⁸ *Id.* at 266-67.

concluding that the spermicide did not cause birth defects. The judge discounted the FDA report because one of the defense experts had served as a consultant to the FDA panel of scientists who conducted the investigation.¹⁹⁹ A single case does not resolve the issue of judicial competence, of course, but it is far from the only evidence that judges, too, may have trouble assessing complex expert testimony, especially testimony involving statistical evidence.²⁰⁰

Judges are also susceptible to the usual cognitive biases that afflict most people. Guthrie, *et al.*, asked 167 United States magistrate judges to respond to a five page questionnaire that presented them with five items designed to assess the influence of five common cognitive biases: (1) anchoring (the misuse of an artificial initial value on a numerical estimate); (2) framing (the effect of framing a decision in terms of “gains” or “losses” on risky decisions); (3) hindsight bias (overestimating the probability of a known past incident); (4) the representative heuristic (underutilization of base rate information); and (5) egocentric biases (believing oneself to be above average on various dimensions).²⁰¹ To a greater or lesser extent, the magistrates fell prey to each of the biases. However, the authors concluded that judges did better than other groups with respect to the framing and the representativeness heuristic biases.²⁰² They concluded that judges are likely to be better decision-makers in circumstances where decision-making experience blunts the effects of illusions. For example, even though both judges and jurors are prone to anchoring effects, the deleterious effects of this problem is a product of the reasonableness of the initial value. Insofar as judges have better starting values, the adverse effects of anchoring will be minimized. This might occur, for example, were the judge to assess damages based on prior damage awards in similar cases rather than on some value suggested by one of the litigants to the present lawsuit.²⁰³ Similarly, if judges are better than juries with respect to the representativeness heuristic, they may be in a better position to assess the relevance of a piece of evidence. On the other hand, jurors might be better in minimizing the effects of hindsight bias.

All of this data is quite instructive, but it is not entirely relevant

¹⁹⁹ Gross, *supra* note 125, at 1123.

²⁰⁰ See Gary Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 J. PERSON. & SOC. PSYCHOL. 739 (1992); see also THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS (Stephen E. Fienberg ed., 1989).

²⁰¹ Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 817 (2001).

²⁰² *Id.* at 817.

²⁰³ *Id.* at 826.

to the admissibility question per se. The question is not whether we should substitute fact finders. Rather, it is whether judges can perform a gatekeeping role with sufficient skill that on balance they improve the probability of a correct outcome at the end of the case. This question leads us to the fourth variable, the controller's criterion of selection among speakers or messages.

D. The Controller's Criterion of Selection Among Speakers or Messages

Data on how judges actually decide admissibility decisions is quite limited. Sophia Gatowski and her colleagues conducted telephone interviews with a sample of 400 state court trial judges from all fifty states that was designed to assess their understanding of the *Daubert* admissibility criteria.²⁰⁴ Most of the judges reported having some CLE training about scientific evidence, but ninety-six percent reported no training about scientific methods and principles. Therefore, it is unsurprising that the judges had difficulty explaining the *Daubert* falsifiability and error rate criteria. The authors concluded that only four percent of the judges offered an explanation that involved a clear understanding of falsifiability and thirty-five percent gave answers that were clearly wrong. The results were similar with respect to error rate. The judges did much better when asked to explain peer review and general acceptance. Seventy-one percent clearly understood the former and eighty-two percent clearly understood the latter. Based on this study, one would conclude that judges would do better by peripherally processing the adequacy of expert testimony than by attempting to centrally process the evidence with the first two *Daubert* criteria as a guide.

Unfortunately, it is unclear from this study whether the problems judges had as a group in providing definitions translates into poor admissibility decisions. Another study by Margaret Kovera and Bradley McAuliff suggests that it may.²⁰⁵ They surveyed circuit court judges in Florida to examine whether they were able to identify research flaws in proposed expert testimony in a hostile work environment sexual harassment case. The expert proposed to testify concerning a study the expert had conducted on whether exposure to sexually suggestive materials influenced participants' subsequent

²⁰⁴ Sophia Gatowski et al., *Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-Daubert World*, 25 LAW & HUM. BEHAV. 433 (2001).

²⁰⁵ Margaret Kovera & Bradley 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).

interactions with a female confederate. Some judges heard a version without methodological flaws. Others heard versions that were altered to include one of three flaws: a missing control group,²⁰⁶ a confound,²⁰⁷ or potential experimenter bias due to a non-blind

²⁰⁶ In the correctly done study, men exposed to sexually explicit materials were compared with men who were not. In the first flawed study, the second group was absent. Without a control group it is usually difficult to know what to make of any results. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE, *supra* note 20, § 4-4.1.2, offers the following example:

[O]ccasionally people have suggested that marijuana, even if not harmful in itself, is dangerous because it leads (by some pharmacological or psychological or sociological route) to the use of harder drugs. They have offered the suggestion that if a substantial number of heroin addicts were found to have used marijuana when they were younger, that would confirm the hypothesis. Figure 3 depicts the pattern of data such commentators have in mind. The hypothetical data in the figure show sixty percent of a sample of 500 heroin addicts to have used marijuana at an earlier time.

Figure 3
Single row (or column) missing data pattern

Heroin Addict	Smoke Marijuana at Time-1	
	Yes	No
at Time-2		
Yes		300
No		200
	—	—

Because these data [do not have a control group], they cannot reveal whether or not a relationship exists. More specifically, without comparison data we cannot know whether fewer than sixty percent, about the same sixty percent, or more than sixty percent of people who are not heroin addicts earlier smoked marijuana. And it is on that comparison that the existence or non-existence of a relationship depends.

Suppose, for example, that someone had proposed that drinking milk as a child led to heroin addiction as an adult. The same table with marijuana replaced by milk would reveal that more than ninety-nine percent of heroin addicts drank milk as children. Would that reveal that milk was to blame? Filling in the data for the rest of the table would make clear that ninety-nine percent of non-heroin addicts drank milk as children, and therefore no relationship between milk drinking and heroin addiction existed. Until the rest of the data were supplied for Figure 3, one could not tell whether a relationship existed between marijuana smoking and heroin addiction, other than by speculating on what the missing cells contained.

²⁰⁷ In a confounding situation, some other variable, often introduced as part of the methodology of the study, may be what explains one's result. Because of the

confederate.²⁰⁸

Judges' ratings of the quality of the study did not differ based on these manipulations. Judges were also asked to make an admissibility decision concerning this expert. Again, the quality of the study did not affect this decision. Judge rating of the reliability of the evidence revealed some effect. Judges who had received training in scientific methods rated the valid study significantly more positively than other judges who had not undergone training.²⁰⁹ However, training seems not to have sensitized judges to the specific methodological problems associated with missing control groups or the lack of a blind condition.

It is unfortunate that these studies did not include a sample of federal judges. State court trial judges rarely write opinions and, therefore, there is no opportunity to examine whether their inability to provide a definition translates into poor judgments or, on the other hand, whether when forced to rule on admissibility in a case involving complex scientific data, judges are able to provide a clear written explanation for their opinion.

Because federal judges do often write opinions, especially when their admissibility decision has the effect of non-suiting one of the parties, it is theoretically possible to assess judicial understanding by looking at the opinions themselves. I am not aware of any systematic effort to do this. My own reading of the vast majority of *Daubert* opinions in toxic tort cases, leads me to believe that the opinions themselves display a range of comprehension.²¹⁰ Judging by opinions, far more than four percent of the cases exhibit a clear understanding

study design, there is no way to separate out the independent effects of the variable of interest and the confounder. In this study the confound was introduced by using two research assistants. One of the assistants only interacted with men exposed to sexual material and the other research assistant only interacted with men who were not exposed. The problem is that we cannot now ascertain whether the effects we observe are due to the presence or absence of sexually explicit materials or, on the other hand, are due to differences in the attributes of the two research assistants. Of course, the more similar the assistants are to each other, the less we may be concerned by this threat to internal validity.

²⁰⁸ Best experimental practice "blinds" researchers (and often subjects as well) to the treatment. Here, a confederate who observed participant interactions with women employees knew whether the participants had or had not been in the experimental condition. The danger is that this knowledge may skew the confederate's observation in favor of the underlying hypothesis (i.e. that exposure alters behavior).

²⁰⁹ Margaret Kovera et al., *Assessment of the Commonsense Psychology Underlying Daubert*, 8 PSYCHOL. PUB. POL'Y & L. 180, 186 (2002).

²¹⁰ This nearly masochistic act is the result of required annual updates to a treatise.

of falsifiability and nowhere near thirty-five percent are clearly wrong. The same might be said of the error rate. Moreover, many opinions do not turn on either of these criteria, but rather on an analysis of internal and external validity threats to the research underlying an expert's testimony and the fit between the findings in this research and the conclusions drawn by the expert.²¹¹

The fact that the federal district court opinions paint a more favorable picture of judicial conceptual understanding than the Gatowski and Kovera results is not surprising. Many of the federal opinions follow a Rule 104A *Daubert* hearing in which the court has heard the parties and their experts discuss and brief the merits of the expert knowledge in question. Moreover, more federal judges have enjoyed some education in research methodology through Federal Judicial Center programs. Unfortunately, fewer state court judges have this opportunity nor do they have the luxury of lengthy admissibility hearings or a clerk to assist them in their decision making.

Even if a survey of federal judges demonstrated that they could provide a clear definition of the *Daubert* criteria, this does not mean that judges employ the criteria wisely. Knowledge is not the same thing as judgment. Unfortunately, at the federal level there does not exist any individual level data on the merits of judicial decisions to admit or exclude. What is available is aggregate data from a recent Rand study on the effects of *Daubert* on civil litigation.²¹² In this study, Dixon and Gill analyze 399 federal district court opinions between January 1980 and June 1999 that addressed challenges to expert evidence in civil cases.²¹³ Many of the opinions analyzed more than one element of evidence and the researchers separately coded 601

²¹¹ See FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE, *supra* note 21, ch. 34-35. This intuition that non-*Daubert* factors have played an increasingly important role is supported by a recent Rand study on the effect of *Daubert*. Dixon and Gill note that:

after *Daubert*, challengers and judges initially focused on the *Daubert* factors when challenging and evaluating reliability. . . . As time passed, however, and judges gained experience in evaluating reliability and appellate court opinions reinforced their authority, challengers and judges would have felt less compelled to address each *Daubert* factor and instead paid increasing attention to more general factors important to assessing reliability.

Lloyd Dixon & Brian Gill, *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the Daubert Decision*, 8 PSYCHOL. PUB. POL. & L. 251, 284-85 (2002).

²¹² Dixon & Gill, *supra* note 211.

²¹³ This group consisted of a thirty-three percent random sample of 1,345 civil cases that a Westlaw search identified as dealing with admissibility of expert testimony. *Id.* at 264.

elements. Nearly half of the cases involved personal injury. Contracts and business torts (16%) was the next largest category. Fifteen percent of the cases were lumped into a general “other and unknown” category.²¹⁴ Approximately eighty percent of the evidence discussed in the opinions was proposed by plaintiffs, and a similar percentage of the challenges came from defendants.

The researchers concluded that during the period of their study, judges did scrutinize expert testimony more carefully and applied stricter admissibility standards, not only with respect to the reliability factor but also with respect to relevance and expert qualifications.²¹⁵ This trend began even before the *Daubert* opinion and continued after that opinion. Moreover, challenges to expert evidence increasingly resulted in a summary judgment.²¹⁶

Dixon and Gill are especially interested in how litigants responded to this tightening in standards. They observe that up until 1997, the proportion of evidence excluded increases consistent with the stricter scrutiny under *Daubert*. After 1997, however, there is a decline in the percentage of evidence elements excluded, controlling for type of case.²¹⁷ The authors assume that the admissibility standard did not change in this period and, therefore, concluded that this declining rate is consistent with one or both of the following factors. It could be that the declining rate occurred because the parties, emboldened by past successes, extend their challenges to relatively better evidence and their success rates on the margin decline. On the other hand, it could be that challenges are less successful, because the parties have responded to heightened scrutiny by offering, on average, better evidence. In the authors’ judgment, both processes may have been at work.²¹⁸

Dixon and Gill are careful to note what their study cannot do: it cannot specifically address the question of how well judges are performing the gatekeeping function. There are some indicia that federal judges were becoming more sophisticated in their explanations for exclusion. In the first few years after *Daubert*, judges seem to have been feeling their way and their opinions hewed closely to the *Daubert* criteria.

Over time, however, as judges gained experience in evaluating reliability and as appellate court opinions reinforced their

²¹⁴ *Id.* at 267.

²¹⁵ *Id.* at 274, 291.

²¹⁶ *Id.* at 294-96.

²¹⁷ *Id.* at 292-93.

²¹⁸ Dixon & Gill, *supra* note 211, at 299.

authority, they appear to have felt less compelled to address each *Daubert* factor and to have paid increasing attention to more general issues important to addressing reliability. Of particular note was the rapid rise in the frequency with which judges addressed the clarity and coherence of the expert's explanation of the theory, methods, and procedures underlying the evidence.²¹⁹

But greater sophistication does not necessarily mean better judgment.²²⁰ The authors suggest a research design that would have a panel of experts assess the reliability of admitted and excluded evidence.²²¹

In sum, the data available to address Goldman's fourth factor is limited and somewhat contradictory. At the state level, the Gatowski, Kovera, and McAuliff research should give pause as to the ability of courts to apply some *Daubert* factors. At the federal level, the Dixon and Gill paper offers persuasive evidence that expert evidence has been subjected to greater scrutiny and the discussion of the evidence in judicial opinions exhibits greater sophistication than in the years immediately following *Daubert*. However, we have only impressionistic evidence about the "correctness" of these opinions. Indeed, we probably do not have much agreement about what constitutes a "correct" outcome. Dixon and Gill's proposal would constitute a first step, but only a first step. Even if one could reduce concepts such as reliability to a single dimension, and then rank order opinions along such a dimension, one would still be left with the question of the optimal degree of scrutiny.

E. The Availability of Other Alternate Channels of Communication That Address the Same Topic

For the purposes of this paper, Goldman's final criterion might

²¹⁹ *Id.*

²²⁰ Psychologists are likely to make this point by distinguishing between coherence and correspondence. Most psychological research on decision making involves coherence theories and tries to explain the process by which a person's judgements are logical and rational, or not. Correspondence theories, on the other hand, are designed to explain why or why not a person's judgments achieve empirical accuracy. See Gregory Mitchell, *Taking Behavioralism Too Seriously? The Unwarranted Pessimism of the New Behavioral Analysis of Law*, 43 WM. & MARY L. REV. 1907, 1941 (2002). Of course there may be a correlation between coherence and correspondence, but one does not perfectly predict the other. Judges may "talk a good *Daubert* game" and still get it wrong in much the same way that Judge Shoob talked a good demeanor game in the spermicide case, but got it wrong.

²²¹ The authors also note that their study did not address the question of costs nor could it assess *Daubert's* overall affect on case outcomes. Dixon & Gill, *supra* note 211, at 301-02.

be put as follows: are there other, perhaps better, ways to assist jurors than restricting the information they hear? The Kansas Supreme Court, for example, argued in *Kuhn* that the traditional tools of attorney argument—opposing experts and vigorous cross-examination—are sufficient to guard against decisions based on unreliable evidence.

Diamond and her colleagues examined this proposition within the context of a criminal case.²²² The stimulus involved the testimony of an expert modeled after a Dr. James Grigson, a psychiatrist who regularly testified for the prosecution in death penalty cases on the issue of future dangerousness.²²³ Typically, Dr. Grigson would testify that the defendant constituted an ongoing danger. For example, in *Barefoot v. Estelle*, he stated that there was a “one-hundred percent and absolute chance” the defendant would commit future crimes of violence.²²⁴ In the experiment, the jury, drawn from the Cook County, Illinois jury pool, watched a seventy-five minute videotape of a death penalty hearing involving an armed robbery and murder of a stranger who the defendant robbed in order to buy beer.²²⁵

In three conditions of the experiment, the jurors heard the prosecution’s expert state that he had diagnosed the defendant as a sociopath, based solely on an examination of records of prior court proceedings, pre-sentence reports, and prison records. The expert concluded that the defendant was “certain to kill again” if he was not executed. The expert asserted that he had extensive prior experience in making such predictions, and his predictions were generally accurate.²²⁶

In the first, “weak cross-examination” condition, the defense only brought out the fact that the witness usually testified for the state, but did not challenge the future dangerousness prediction.²²⁷ In the second, “strong cross-examination” condition, the defense

²²² Shari Se. Diamond et al., *Juror Reactions to Attorneys at Trial*, 87 J. CRIM. L. & CRIMINOLOGY 17 (1996).

²²³ In Texas, where Dr. Grigson most frequently testified, the jury could not impose the death sentence unless they concluded that the defendant was likely to “commit criminal acts of violence that would constitute a continuing threat to society.” TEX. CODE CRIM. PROC. ANN. art. 37.071(2)(b)(1) (Vernon 2002).

²²⁴ *Barefoot v. Estelle*, 463 U.S. 880, 919 (1983).

²²⁵ Diamond et al., *supra* note 222, at 19.

²²⁶ *Id.* at 38.

²²⁷ There is some evidence that this is a typical cross in many circumstances. My experience in reading the trial transcripts of seven Bendectin trials is that the vast majority of cross is devoted to expert qualifications and potential sources of bias. Margaret Kovora et al., *Assessment of the Commonsense Psychology Underlying Daubert*, 8 PSYCHOL. PUB. POL’Y & L. 180, 192 (2002) (reporting a similar result).

added to the cross in the weak condition by pointing out at length that the expert's prediction of future killing was inconsistent with prior research and that the expert has not employed the standard methods for diagnosing future dangerousness. The witness admitted on the stand that the best scientific literature indicates that two-thirds of dangerousness predictions prove to be incorrect. The cross-examination also brought out the fact that the expert had never published his findings in peer-reviewed journals. The expert responded that he was focused on clinical diagnosis, not publication, and that he was confident he was correct.²²⁸ In the third, "strong cross-examination plus defense expert" condition, the defense lawyer conducted the same cross-examination as in the strong-cross condition. In addition, a defense expert, who was also a psychiatrist, testified that the defendant coped reasonably well but on rare occasions excessive drinking interacted with a personality disorder to produce violence. The defense expert testified that predictions about future violence could not be made with any certainty, but that in his view, the likelihood of future similar violence was not great and the defendant was a good candidate for an alcohol abuse program.²²⁹

In a fourth "control" condition, the plaintiff expert made a realistic prediction, basically agreeing with the defense expert that predictions of future dangerousness are accurate only about one-third of the time, but warned about the defendant's potential for future violence. The cross-examination was identical to the cross in the first condition.²³⁰

The dependent variables in the study included a question about the persuasiveness of the state's expert, the jury verdict preference (death or life), and a verdict confidence index. The first condition, with a strong prediction of future dangerousness, no opposing expert, and a weak cross should produce the highest percentage of death penalty verdicts. If cross-examination is an effective prophylactic against unreliable testimony, the second condition should produce lower persuasiveness scores and a lower percentage of death penalty verdicts. And the combination of a powerful cross and an opposing expert should produce still lower persuasiveness scores and even fewer death penalty verdicts. Ideally, this version

²²⁸ Diamond et al., *supra* note 222, at 39.

²²⁹ *Id.*

²³⁰ My discussion of this condition and the results come from a chapter of an as yet unpublished book by Diamond and Casper. SHARI S. DIAMOND & JONATHAN D. CASPER, UNDERSTANDING JURIES, ch. 4: The Influence of Experts (forthcoming in 200_).

would produce jury judgment indistinguishable from the fourth version in which the expert reported a one in three chance of being correct.²³¹

In fact, neither the strong cross, nor the strong cross plus the opposing expert had a significant affect on the plaintiff expert's persuasiveness, percentage of juries opting for death, or verdict confidence. For example, in the weak cross condition forty-seven percent of the juries gave a death verdict, in the strong cross condition fifty-one percent recommended death, and in the strong cross plus opposing expert fifty percent recommended death. The only condition with a different result was in the fourth, "control" condition where the plaintiff's expert testified that predictions were wrong two-thirds of the time. In this condition, thirty-nine percent of the juries recommended the death penalty.²³² Diamond and Casper note that one possible interpretation of these results is that the jurors simply did not care about future dangerousness. However, based on evidence from their deliberations, this is not the case. Most juries explicitly discussed the issue, and there was a strong correlation between predictions of future dangerousness and verdict preferences.²³³ However, jury estimates of future dangerousness if released did not vary significantly across the three conditions where the plaintiff's expert testified the defendant would kill again.²³⁴

Another recent future dangerousness experiment using undergraduates as subjects by Krauss and Sales also explored the effects of cross-examination and opposing experts on juror evaluations.²³⁵ The jurors heard two types of experts give testimony. The "clinical opinion expert" based his judgment that the defendant was a severe sociopath and represented a future danger to society on his interview with the defendant and his years of experience. The "actuarial expert" had identical experience and training as the clinical expert. This expert used a dangerousness prediction actuarial instrument called the Violence Risk Assessment Guide (VRAG) to assess future dangerousness. The expert explained the instrument and testified that based on his interview with the defendant and the VRAG results, he believed the defendant

²³¹ DIAMOND & CASPER, *supra* note 230, at 41.

²³² *Id.* at 41 t.4.

²³³ *Id.* at 43.

²³⁴ *Id.* at 43 t.5.

²³⁵ Krauss & Sales, *supra* note 154. Unfortunately, this study did not have the subjects deliberate, minimizing its ecological validity.

represented a future danger to society.²³⁶ The results indicated that the clinical expert was more effective than the actuarial expert in changing the mock juror's dangerousness ratings.

The testimony of both types of experts was met with opposing expert testimony and an effective cross-examination that pointed out the high error rates of predictions of future dangerousness. In this experiment, the cross-examination and opposing expert reduced mock juror assessment of future dangerousness from where it was after hearing the first expert's testimony. However, Krauss and Sales note that, "although adversary procedures had an impact on both types of expert testimony, their influence was significantly less on the clinical opinion expert testimony. Adversary procedures failed to return mock jurors who received clinical opinion expert testimony to their initial dangerousness rating levels."²³⁷ These studies suggest that, perversely, the *Kuhn* court may have had it backward when it suggested that clinical pure opinion testimony could be effectively countered through cross-examination and opposing expert testimony. The limited research available suggests that this type of testimony is more impervious to cross-examination and opposing experts than technique evidence.²³⁸

These studies do not resolve the question of the effectiveness of the "battle of the experts" and cross-examination. Seidman and Casper offer the possibility that future dangerousness testimony is particularly difficult to overcome because it is consistent with beliefs and expectations already held by the jurors.²³⁹ However, the results are consistent with a study by Kovera and her colleagues.²⁴⁰ They varied the strength of the defense's cross-examination of an expert witness. Although manipulation checks revealed that jurors were sensitive to the relative strength of the cross-examination of the expert, this did not affect participant's perceptions of the quality of the evidence nor did it affect the verdict. This result was replicated in

²³⁶ *Id.* at 286-87.

²³⁷ *Id.* at 302.

²³⁸ This may be due in part to the fact that jurors confronted with holistic experience testimony are not provided with the decompositional decision strategies that have been shown to improve complex decision making. See Osvaldo F. Morera & David Budescu, *Random Error Reduction in Analytic Hierarchies: A Comparison of Holistic and Decompositional Decision Strategies*, 14 J. BEHAV. DECISION MAKING 223 (2001).

²³⁹ *Id.* at 53. There is research supporting the proposition that mock jurors hold strong beliefs concerning the ability of clinicians to predict future dangerousness and that they overestimate clinician accuracy. Krauss & Sales, *supra* note 154, at 276.

²⁴⁰ Margaret Kovera et al., *Expert Testimony in Child Sexual Abuse Cases: Effects of Expert Evidence Type and Cross-Examination*, 18 LAW & HUM. BEHAV. 653 (1994).

a second study by the same authors.²⁴¹ Together, these experimental findings should give pause to the *Kuhn* court and others who believe that the traditional tools of the adversarial process are a full substitute to restrictions on the admissibility of unreliable expert testimony.

These results raise another issue as well. They lend support to the argument that rulings excluding unreliable evidence promote jury accuracy even if we assume jurors are as good as judges in assessing reliability. Edward Imwinkelried sets forth this justification for *Daubert* in the following passage:

The criterion is not whether the judge is more competent to decide the issue than the jury. Rather, the test is whether there is a significant risk that the jurors' exposure to the foundational testimony and the proffered evidence will distort their deliberations even when they make a conscious decision that the item of evidence is technically inadmissible.²⁴²

If any opinion evidence is to be excluded because of an inadequate foundation, these experiments support the idea that it is better to separate the admissibility decision from the decision as to how much weight to give some testimony.

CONCLUSION: SUMMARIZING THE RESEARCH

Goldman's five criteria offer one roadmap through the empirical literature on juries, judges, and the admissibility of expert testimony. Jurors have trouble understanding expert testimony, especially in complex cases that involve statistical and probabilistic evidence. This difficulty affects how they relate to expert witnesses. When the nature of the testimony is such that it is difficult for jurors to understand, they are more likely to engage in peripheral processing; that is they rely on factors other than the merits of the argument itself to determine the truth value of a message. Such factors may include the expert's credentials, the expert's demeanor, how much the expert is paid, whether the expert's findings have been published, and the number of experts (rather than the quality of their testimony).

Expert opinion that rests solely on the experience of the expert may create similar problems for jurors. The Supreme Court's *ipse dixit* comment in *Joiner*²⁴³ can be understood in part as a recognition that whether or not an expert's bare assertion based on professional

²⁴¹ Kovera et al., *supra* note 148, at 362.

²⁴² Edward J. Imwinkelried, *Commentary: Logerquist v. McVey: The Majority's Flawed Procedural Assumptions*, 33 ARIZ. ST. L.J. 121, 132-33 (2001).

²⁴³ *General Electric v. Joiner*, 522 U.S. 136, 146 (1997).

judgment is correct, such statements offer the jury no opportunity to engage in central processing and force it to accept or reject the assertion, presumably on other, often peripheral grounds.²⁴⁴

Juror difficulties potentially are made worse by the status of the experts. American expert witnesses are almost exclusively employed by the parties within the context of an adversarial proceeding. In most contexts, when one is embedded in an adversarial process, one's testimony is biased in favor of the side one is representing. However, the effect of these biases on jurors is unclear. There is evidence that jurors give greater scrutiny to adversarial versus non-adversarial expert witnesses. Whether this greater scrutiny is sufficient to permit jurors to detect and discount expert bias is a question for which we have little empirical data.

The fact that the parties choose the experts has another potentially misleading aspect. The system favors selection of experts with extreme views, rather than views that are representative of the scientific community. There is some evidence that this gives jurors the impression that there is less consensus in a field than actually exists. Moreover, when we assess who to believe in situations where we are confronted with different versions of the truth, we may turn to demeanor as a clue to veracity. This type of peripheral processing is in fact encouraged by the legal system. Unfortunately, demeanor is often a poor cue, and in the case of expert testimony, the usefulness of demeanor cues is further attenuated by selection effects.

If it is true that jurors will have a difficult time assessing the merits of the experts' arguments in complex cases, and the process of selection and presentation of evidence is likely to diminish the usefulness of many peripheral cues to veracity, close scrutiny of testimony is potentially most beneficial in complex cases where peripheral processing is most likely. Jurors may still engage in this type of processing, but the elimination of the least reliable evidence should reduce the egregiousness of errors that are made.

But if jurors and experts are the problem, are admissibility rules imposed by judges the answer? If the judge is to act as a gatekeeper, what evidence do we have that he or she is up to the task? This question has two parts. First, do judges have the skills to judge expert evidence, and second, in the context of a trial can they use the skills they have to sort out reliable and unreliable testimony. Empirical research on these questions is limited and somewhat contradictory. A survey of state judges indicates a shallow understanding of some

²⁴⁴ See Sanders, Kumho *and How We Know*, supra note 184, at 373, 408.

Daubert criteria, but a study of published federal court opinions suggests that as federal judges gain experience in evaluating reliability, their opinions became more scientifically sophisticated. Unfortunately, we do not have systematic evidence as to whether greater sophistication produces better decisions, that is, decisions that reliably distinguish between more and less reliable expert testimony.

Even if judges are no better than jurors in assessing expert evidence, the Diamond and Casper research support the position of some evidence scholars that final outcomes will be better if the task of assessing admissibility is separated from the task of assigning the proper weight to each piece of evidence.

In sum, I believe on balance, the empirical research does lend some support to the paternalistic justification for restrictions on the admissibility of unreliable expert testimony. We are still left with many questions, however. The following two seem to me to be particularly important. First, is reliability a good principle upon which to base decisions? Second, how stringent should admissibility criteria be?

As to the first question, in my judgment reliability is a serviceable criterion. Reliability itself is a complex, multidimensional set of considerations.²⁴⁵ Its very complexity can be a source of trouble, as seen in the survey of state court judges, but its complexity also allows judges to make sophisticated judgments.

Evidence from the Rand study suggests that courts have used the reliability test to exclude the most unreliable evidence. If this is the case, the result is consistent with the adoption of what Nance calls a “worst evidence principle:” “evidence law seeks to prevent jury error by filtering out the really bad evidence that is likely to lead the jury astray.”²⁴⁶ There is also some evidence that courts have been sensitive to the quality of the available evidence when making admissibility

²⁴⁵ See Joseph Sanders, *Scientific Validity, Admissibility, and Mass Torts after Daubert*, 78 MINN. L. REV. 1387 (1994).

²⁴⁶ Dale A. Nance, *Naturalized Epistemology and the Critique of Evidence Theory*, 87 VA. L. REV. 1551, 1555 (2001). In a recent opinion adopting the *Daubert* standard, the Nebraska Supreme Court expressed a similar view:

We are convinced that by shifting the focus to the kind of reasoning required in science—empirically supported rational explanation—the *Daubert/Joiner/Kumho Tire Co.* trilogy of cases greatly improves the reliability of the information upon which verdicts and other legal decisions are based. Because courts and juries cannot do justice in a factual vacuum, the better information the fact finders have, the more likely that verdicts will be just.

Schafersman v. Agland Coop., 631 N.W.2d 862, 876 (Neb. 2001).

decisions. As better evidence becomes available, courts may not let experts base opinions on less reliable evidence that would be admissible in another context.²⁴⁷ This phenomenon, which might be called a “better evidence principle,” suggests that judges are using admissibility rulings to control advocates as well as juries. Nance describes the advocate control approach as a way in which the judge, as an agent for the jury, is able “to protect juries from the epistemic consequences of third-parties’ choices—namely, the choices of the advocates about what evidence to present and how to present it.”²⁴⁸ Whether or not judges are doing a good job in individual cases, it does appear that they are using admissibility decisions in a way that is consistent with a focused paternalistic justification. The federal opinion study of Dixon and Gill is consistent with the possibility that in face of heightened judicial scrutiny, the parties themselves have improved the overall quality of evidence.

The second question is how stringent should the courts be?²⁴⁹ How good is good enough? Insofar as this is a question that solely concerns overall trial accuracy, the theoretical answer is up to the point where improvements in jury accuracy are offset by judicial errors in excluding evidence. A more practical response might be that courts should be especially vigilant in monitoring weak evidence in cases involving complex questions, because jurors may have special difficulty assigning proper weight to such testimony. This argument applies at least as strongly to experience evidence (what the *Kuhn* court called “pure opinion evidence”), as it does to other types of testimony. As cases get closer on the merits, of course, the role of admissibility decisions as a means of assisting juries diminishes, if, for

²⁴⁷ David L. Faigman et al., *How Good is Good Enough?: Expert Evidence Under Daubert and Kumho*, 50 CASE W. RES. L. REV. 645 (2000).

²⁴⁸ Nance, *supra* note 246, at 1557.

²⁴⁹ It is quite unlikely that the law of evidence will abandon restrictive evidentiary rules, either in general or with respect to the admissibility of expert evidence. Rule 702 and Rule 403 embody such restrictions. The choice, therefore, does not include a control version of the requirement of total evidence approach, but rather the nature and extent of restrictions on expert testimony. My sense is that most litigants are more concerned with the extent of the restrictions than they are with the legal formula employed to discuss the restriction. If plaintiff lawyers are disturbed by developments following the Supreme Court opinion in *Daubert*, I believe it is not primarily because they have a conceptual preference for the admissibility criteria of *Frye*. At the time *Daubert* was decided many plaintiff experts considered it to be a conceptual victory. See Kenneth J. Chesebro, *Taking Daubert’s “Focus” Seriously: The Methodology/Conclusion Distinction*, 15 CARDOZO L. REV. 1745 (1994). Current objections exist primarily because decisions under the *Daubert* standard are thought to be more restrictive. See Sanders et al., *Trial Lawyer Perceptions of Expert Knowledge*, *supra* note 155.

no other reason, than the fact that as cases are closer and closer on the merits, it is not clear whether an outcome is more accurate.²⁵⁰

Moreover, there are competing considerations. As I noted above, some may disagree that the primary end in view of the law of evidence and of the trial itself is to uncover the truth. They may argue that procedural justice considerations should trump substantive considerations and that both the parties and the society will be more supportive of outcomes that are the result of a jury verdict. This paper is not the place for a full discussion of the relationship between substantive and procedural justice. However, a few words are in order. Procedures that are perceived to be fair, help to produce acquiescence even in the face of perceived outcome unfairness. One well accepted theory advanced by Tyler and Lind points to three factors that are important to the belief that procedures are fair: (1) neutrality (the authority engages in evenhanded treatment), (2) trust (the authority tries to be fair), and (3) status recognition (the authority treats one politely, with dignity, and with respect for one's rights and opinions).²⁵¹ Adverse admissibility decisions, especially if they result in a directed verdict for the other party, may well be perceived to be unfair with respect to one or more of these factors. For example, refusal to permit an expert to testify might be perceived as a lack of respect for one's opinion.

Undoubtedly, the plaintiff's personal injury bar feels strongly that cases decided on the basis of the exclusion of expert opinion evidence are less legitimate. Although most explain this position in substantive terms, e.g. jury verdicts are more accurate, it may also be true that their objection is partly procedural. Pushed, they might say that jury judgments are "fairer," and we should prefer them even if, on average, juries reach more erroneous results in the absence of admissibility rules that require reliable expert evidence. The question, of course, is how erroneous? Institutions that routinely fail to achieve substantive justice are likely to lose political and social

²⁵⁰ However, admissibility decisions do not have to be the only arrow in the judge's quiver. The jury's difficulty in weighting complex testimony does not exist only in the weakest cases. Court appointed experts, comments on the weight of the evidence, and other devices may have a role to play in assisting the jury, especially in the face of an adversarial process that often obscures as much as it clarifies. Erichson, *supra* note 171. In addition, see Phoebe Ellsworth & Alan Reifman, *Juror Comprehension and Public Policy*, 6 PSYCHOL. PUB. POL. & L. 788 (2000), for a review of research on the effectiveness of methods designed to assist jury comprehension.

²⁵¹ See Tom R. Tyler & E. Allan Lind, *Procedural Justice*, in HANDBOOK OF JUSTICE RESEARCH IN LAW (Joseph Sanders & V. Lee Hamilton eds., 2001).

support even if they are perceived to be procedurally just. The purpose of just procedures is not solely a way to “cool out” the losers in disputes. It is also a way to arrange things so as to come as close as possible to achieving substantive justice. Unfortunately, procedures designed to maximize substantive justice may conflict with procedures that are designed to maximize the perception of procedural justice.²⁵² As is so often the case in law, the question is one of balance. Some may strike the balance more in favor of procedural justice.

However one may weigh these competing concerns, we should understand that the question of how high the admissibility hurdle should be is not only a narrow question of evidence law. If the paternalistic instinct that underlies *Daubert* has any merit when we ask how high the bar should be, we are in part asking about the proper balance between procedural and substantive justice. Unlike many quarrels in law, this one is about something important.

²⁵² What we sorely need is more research on how people integrate their assessment of these different dimensions. See Joseph Sanders & V. Lee Hamilton, *Justice and Legal Institutions*, in HANDBOOK OF JUSTICE RESEARCH IN LAW (Joseph Sanders & V. Lee Hamilton eds., 2001).