

Reliability Under Rule 702: A Specialized Application of 403

*Calvin William Sharpe**

The concern for witness reliability is not new to Evidence law. Since the Anglo-American adversarial system relies for accurate fact-finding upon jurors who have no independent knowledge of the facts at issue, it is important for witnesses supplying evidence that will form the basis of such fact-finding to be reliable.¹ Evidentiary submissions that raise issues of hearsay,² first-hand knowledge,³ original documents,⁴ and others invoke rules that reflect a concern about reliability. Expert testimony raises the same concern under Rule 702.⁵

* John Deaver Drinko—Baker & Hostetler Professor of Law, Case Western Reserve University Law School. The author is grateful to Margaret Skarbek for helpful research assistance on this paper. This research was made possible by a grant from Case Western Reserve University Law School.

¹ RONALD J. ALLEN ET AL., EVIDENCE: TEXT, PROBLEMS, AND CASES § 204 (3d ed. 2002).

² See FED. R. EVID. art. VIII, advisory committee's introductory note (noting as a purpose of the hearsay rule "to encourage the witness to do his best with respect to each of [the factors of perception, memory, narration, and sincerity], and to expose any inaccuracies which may enter in").

³ See FED. R. EVID. 602 advisory committee's note (citing McCormick for the following passage in describing the justification for the rule: "[T]he rule requiring that a witness who testifies to a fact which can be perceived by the senses must have had an opportunity to observe, and must have actually observed the fact' is a 'most pervasive manifestation' of the common law insistence upon 'the most reliable sources of information.'").

⁴ See FED. R. EVID. 1001 advisory committee's note (noting that the best evidence "afforded substantial guarantees against inaccuracies and fraud"); see also 2 MCCORMICK ON EVIDENCE 229 (John W. Strong ed., 5th ed. 1999) (noting "the danger of mistransmitting critical facts which accompanies the use of written copies or recollection").

⁵ See *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 591-92 (1993) (noting the concern of the common law with reliable sources of information as manifested in the personal knowledge and hearsay rules and explaining that the reliability requirement in 702 grows out of a concern about the wide latitude permitted the expert witness whose opinions need not be based on firsthand knowledge or observation).

In any of these contexts, perfectly reliable evidence would justify the jury's absolute confidence in the truth of its contents and provide complete support for reasonable inferences to be drawn from the evidence.⁶ Some cross-examination and all impeachment are efforts to diminish the perceived reliability of evidence in the hope that a fact-finder will discount its value.⁷

THE LAW

The Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,⁸ *General Electric Co. v. Joiner*,⁹ and *Kumho Tire Co. v. Carmichael*,¹⁰ and Congress in Rule 702 as amended in 2000, have attempted to delineate the parameters of reliable expert testimony. Even before the 2000 amendments the concept of reliability was embedded in Rule 702. That rule permitted an expert to give helpful scientific, technical, or other specialized knowledge. It is the concept of knowledge in Rule 702 that embodied the idea of reliability as truth. The Supreme Court unearthed the buried idea of reliability in *Daubert*.¹¹ In dispelling the notion of general acceptance as a prerequisite to the admissibility of scientific evidence, the Court expressed the assurance that admissibility of scientific evidence was not unlimited.¹² It spelled out a judge's screening function in such cases as involving the dual determination of relevancy and reliability.¹³ Since *Daubert* involved scientific evidence, the Court in the following terms discerned the reliability requirement in 702's language sanctioning expert testimony about scientific knowledge:

The subject of an expert's testimony must be "scientific . . . knowledge." The adjective "scientific" implies a grounding in the

⁶ See 13 THE OXFORD ENGLISH DICTIONARY 562 (2d ed. 1989) (defining "reliable" as "1.a. That may be relied upon; in which reliance or confidence may be put; trustworthy, safe, sure").

It also defines "trust" as "[c]onfidence in or reliance on some quality or attribute of a person or thing, or the truth of a statement . . ." 18 *Id.* at 623.

⁷ See *In re Paoli R.R. Yard PCB Litig. (Paoli II)*, 35 F.3d 717 (3d Cir. 1994) (stating that credibility may have a bearing on the reliability of expert testimony); ALLEN ET AL., *supra* note 1, at 108-09, 389-90.

⁸ 509 U.S. 579 (1993).

⁹ 522 U.S. 136 (1997).

¹⁰ 526 U.S. 137 (1999).

¹¹ But see Paul C. Giannelli, *Daubert: Interpreting The Federal Rules of Evidence*, 15 CARDOZO L. REV. 1999, 2015 (1994) (noting that the reliability approach developed five to seven years after the enactment of the rules and commenting that "there is an element of magic in the [*Daubert*] Court's 'discovery' of [the reliability] analysis in Rule 702's phrase 'scientific knowledge'").

¹² 509 U.S. at 589.

¹³ *Id.*

methods and procedures of science. Similarly, the word “knowledge” connotes more than subjective belief or unsupported speculation. The term “applies to any body of known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds.”¹⁴

That the “good grounds” (reliability) supporting scientific knowledge are derived from the scientific method is expressed as follows:

[I]n order to qualify as “scientific knowledge,” an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by an appropriate validation—i.e., “good grounds,” based on what is known. In short, the requirement that an expert’s testimony pertain to “scientific knowledge” establishes a standard of evidentiary reliability.¹⁵

In *Daubert*, the Court is careful to make it clear that its discussion centered upon scientific rather than “technical or other specialized knowledge.”¹⁶ However, a change in the adjective modifying knowledge does not change the central concern of reliability—that such knowledge (scientific, technical or specialized other) be capable of acceptance as truth on good grounds. It also seems clear that the *Daubert* factors showing scientific reliability—testability, peer review, error rate, controlling standards, and general acceptance—might prove useful individually, collectively, or in some combination with other factors in determining the reliability of technical or other specialized knowledge.¹⁷ Always, the point of the inquiry is whether the knowledge in question is capable of acceptance as truth on good grounds.

The district court in *Joiner* found that the opinions of plaintiff’s experts were not based on reliable scientific knowledge.¹⁸ In the court’s view the animal and epidemiological studies relied upon by these experts to show a link between the plaintiff’s exposure to PCBs and his development of small cell lung cancer did not sufficiently support the opinions.¹⁹ Under an “abuse of discretion” standard the Supreme Court upheld the district court’s ruling of inadmissibility. The Court affirmed the district court’s discretion to find an expert’s opinion to be an unsupported assertion, where “too great an analytical gap [exists] between the data and the opinion proffered,”

¹⁴ *Id.* at 589-90 (footnote and citations omitted).

¹⁵ *Id.* at 590 (footnote omitted).

¹⁶ *Id.* n.8.

¹⁷ *See id.* at 593-94. This point seems adumbrated in the Court’s emphasis on the flexibility of the reliability inquiry under Rule 702. *Id.* at 594.

¹⁸ 522 U.S. 136, 144 (1997).

¹⁹ *Id.* at 144-45.

and said that “abuse of discretion” is the proper standard whether the trial court’s decision is to admit or exclude scientific evidence.²⁰

In *Kumho*, the adjective modifying “knowledge” changed.²¹ Unlike *Daubert* and *Joiner* where the issues involved the reliability of scientific evidence, *Kumho* considered the reliability of non-scientific expert testimony.²² For reasons grounded in the language of Rule 702,²³ the rationale of the reliability element of the rule,²⁴ and the difficulty of distinguishing between “scientific” on the one hand and “technical or other specialized knowledge” on the other, the Court held that the judge’s gatekeeping function applied to all expert testimony. To reinforce this conclusion, the Court noted the common approach among all experts to “tie observations to conclusions through . . . ‘general truths derived from . . . specialized experience’” and the role of the judge’s reliability determination in helping the jury evaluate that “foreign experience.”²⁵ Elaborating upon the requirements of Rule 702 the Court said:

The Rule, in respect to all such matters, “establishes a standard of evidentiary reliability.” It “requires a valid . . . connection to the pertinent inquiry as a precondition to admissibility.” And where such testimony’s factual basis, data, principles, methods, or their application are called sufficiently into question, the judge must determine whether the testimony has “a reliable basis in the knowledge and experience of [the relevant] discipline.”²⁶

The Court also reaffirmed the flexibility of the reliability determination, the trial court’s discretion to consider one or more of the *Daubert* factors or others in reaching the reliability determination, and the applicability of *Joiner*’s abuse of discretion standard to trial

²⁰ *Id.* at 146.

²¹ 526 U.S. 137 (1999).

²² *Id.* at 141.

²³ In capturing the essence of the *Daubert* analysis of reliability, the Court focused as follows on the noun, *knowledge*, rather than the adjective, *scientific*, in Rule 702:

In *Daubert*, the Court specified that it is the Rule’s word “knowledge,” not the words (like “scientific”) that modify that word, that “establishes a standard of evidentiary reliability.” . . . Hence, as a matter of language, the Rule applies its reliability standard to all “scientific,” “technical,” or “other specialized” matters within its scope.

Id. at 147.

²⁴ On this point the Court reiterated the expert witness’s “testimonial latitude” that is “unavailable to other witnesses on the ‘assumption that the expert’s opinion will have a reliable basis in the knowledge and experience of his discipline.’” *Id.* at 138 (citations omitted).

²⁵ *Id.* at 148-49 (quoting from Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40, 54 (1901)).

²⁶ *Id.* at 149 (internal citations omitted).

court reliability rulings.²⁷

Whether the engineering testimony of *Kumho* was reliable and capable of acceptance as truth on good grounds depended on the nature of the opinion and its basis. The expert in *Kumho* had concluded that a defect rather than misuse in the right rear tire of the minivan that plaintiff had been driving caused it to blow out, killing one passenger and severely injuring others. The issue boiled down to whether overdeflection (misuse) or a defect had caused the tread of the tire to separate from the inner steel-belted carcass, causing the blow-out. The expert asserted the latter cause. In support of this opinion the expert set forth both a general theory and a specific theory. The general theory held that in the absence of abuse B determined by the expert's visual and tactile inspection of the tire B such a separation is caused by a defect. Under the specific theory, unless two of four possible signs of misuse (overdeflection) could be shown, a conclusion that a defect caused the blowout must follow.²⁸

In deciding to affirm the district court's finding that the expert's testimony was not reliable, the Court focused its critical analysis on the grounds of the expert's conclusion. It cited evidence in the record that raised doubts about the reliability of the expert's use of the two-factor test and visual/tactile inspection.²⁹ Some aspects of this evidence bore upon the credibility of the witness.³⁰ Others focused on a critique of the expert's stated methodology.³¹ After considering

²⁷ *Kumho*, 526 U.S. at 141-42.

²⁸ *Id.* at 143, 154. The expert cited as the symptoms of overdeflection: (a) tread wear on the tire's shoulder that is greater than the tread wear along the tire's center . . . (b) signs of a "bead groove," where the beads have been pushed too hard against the bead seat on the inside of the tire's rim, . . . (c) sidewalls of the tire with physical signs of deterioration, such a discoloration, . . . (d) marks on the tire's rim flange.

Id. at 144.

²⁹ *Id.* at 154-57.

³⁰ For example, the Court noted the inconsistency between the expert's claim that visual/tactile inspection permitted him "to ascertain with some certainty the abuse-related significance of minute shoulder/center relative tread wear differences, but insufficiently precise to tell 'with any certainty' from the tread wear whether a tire had traveled less than 10,000 or more than 50,000 miles." *Id.* at 155. It also noted the insufficiency of the expert's opportunity to observe the tire, having inspected it for the first time and for only a few hours on the morning of the deposition and the inconsistencies between the expert's signed report and his deposition and the contradiction between the expert's report on tread depth and an opposing expert's undisputed measurements. *Id.*

³¹ The Court noted the "'subjective[ness]'" of his mode of analysis in response to questions seeking specific information regarding how he could differentiate between

whether the expert's methodology could withstand scrutiny on its own terms, the Court also approved of the trial court's next steps of determining whether any of the *Daubert* factors pointed to reliability, whether the trial court's own analysis "revealed . . . countervailing factors operating in favor of admissibility which could outweigh those identified in *Daubert*,"³² and whether the parties cited any factors favoring admissibility.³³ In its independent assessment of whether the trial court had abused its discretion by finding the expert's testimony unreliable, the Court compared the expert's methodology to that used by other experts in the industry. It also considered any validation of the expert's approach in the literature of the field, and whether the expert's methodology reflected "the same level of intellectual rigor that characterizes the practice of an expert in the relevant field."³⁴ In *Kumho*, the Court sanctioned an approach to determining the reliability of non-scientific expert testimony that tested whether the opinion was capable of acceptance as truth on good grounds. The grounds included the expert's credibility, the soundness and application of his methodology, the *Daubert* factors, and any other relevant factors.

After the Supreme Court's decisions in *Daubert*, *Joiner*, and *Kumho*, Congress amended Rule 702, effective 2000, as follows to incorporate the teachings of *Daubert* and *Kumho*:

If scientific, technical, 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.*³⁵

The legislative history contained in the Advisory Committee Note ("CAN") makes it clear that the amendment was intended to affirm "the trial court's role as gatekeeper and [provide] some general standards that the trial court must use to assess the reliability and

a tire that actually had been overdeflected and a tire that merely looked as though it had been" his failure to examine many similar tires to determine the significance of the bead groove even though he testified that such an examination would have been appropriate. *Id.* at 155.

³² *Id.* at 156.

³³ *Kumho*, 526 U.S. at 156.

³⁴ *Id.* at 152.

³⁵ FED. R. EVID. 702 (emphasis added to highlight amended provisions).

helpfulness of proffered expert testimony.³⁶ The ACN also reiterates *Kumho's* holding that the gatekeeper function is applicable to scientific as well as non-scientific expert testimony. However, the drafters of the amendment recognized that the reliability determination would vary depending upon the nature of the proffered expertise, saying the following:

Some types of expert testimony will be more objectively verifiable, and subject to the expectations of falsifiability, peer review, and publication, than others. Some types of expert testimony will not rely on anything like a scientific method, and so will have to be evaluated by reference to other standard principles attendant to the particular area of expertise. The trial judge in all cases of proffered expert testimony must find that it is properly grounded, well-reasoned, and not speculative before it can be admitted.³⁷

In addition, the ACN endorses not only the *Daubert* factors for assessing reliability, but also cites other factors used by courts and other uncited factors as relevant to the determination.³⁸

The Supreme Court cases and the Rule 702 amendment and legislative history devote much attention to the judge's performance of the Rule 104(a) gatekeeping function of determining the reliability of expert testimony. However, they only hint at how to recognize reliability for purposes of fulfilling the function. Noting that the proponent's burden is demonstrating "by a preponderance of evidence that [experts'] opinions are reliable rather than 'correct,'" the ACN of the Rule 702 amendments quotes with approval a judicial statement that the "evidentiary requirement of reliability is lower than the merits standard of correctness." The ACN also explains that "the rejection of expert testimony should be the exception rather than the rule," citing *Daubert* language suggesting that "shaky" expert testimony should be admitted and subjected to attack by the adversarial elements of cross-examination, conflicting evidence, and the burden of proof. The reliability determination is not to be confused with the sufficiency determination that the trial judge bases on the record as a whole.³⁹ The dual obligations of reliability and

³⁶ FED. R. EVID. 702 advisory committee's note to amended Rule 702.

³⁷ *Id.*

³⁸ Some of these factors are whether the testimony grows out of research conducted independent of litigation, whether the expert's extrapolation from an accepted premise is unjustifiably broad, whether the expert has accounted for obvious alternative explanations, whether the expert's degree of care matches that of his professional practice, whether the expertise is recognized as giving reliable opinions such as those offered by the expert. *Id.*

³⁹ See *Daubert*, 509 U.S. at 596; see also Dale A. Nance, *Reliability and the Admissibility of Experts*, 34 SETON HALL L. REV. 189 (2003).

relevancy determinations under Rule 702 make it clear that reliability is different from relevancy. Hence, the low standard of relevancy will not suffice to establish reliability.⁴⁰

Like Rule 403, the reliability determination seems to require trial judges to demand more of expert testimony than simple relevancy. Yet the Court in *Daubert* does not explicitly tie the reliability determination to 403 analysis—it simply instructs trial judges to be “mindful of other applicable rules.” However, the Court specifically cites Rule 403 and endorses Judge Weinstein’s observation that Rule 403 gives judges more control over expert witnesses than laywitnesses because of the potential prejudice associated with the former.⁴¹

Before the Court decided *Daubert*, Judge Becker had discussed the relationship between 702 and 403 in *United States v. Downing*,⁴² a case relied upon by the *Daubert* Court. In *Downing*, Judge Becker said the following:

After assessing the reliability of the evidence, the court must also weigh any danger that the evidence might confuse or mislead the jury. It may seem paradoxical to suggest that scientific evidence based on principles bearing substantial indicia of reliability could confuse rather than assist the jury, but we do not doubt that this may be so, in some cases. One example might involve a technique which has “assume[d] a posture of mythic infallibility” among lay persons, or at least one whose shortcomings are, for some reason, unlikely to be effectively communicated to the jury. The degree to which an unwarranted “aura of reliability” attaches to scientific evidence will naturally vary with the type of evidence.⁴³

Later in the *Downing* decision, the court acknowledged that its 702 analysis incorporated “to some extent a consideration of the dangers, particularly the danger of unfair prejudice, enumerated in Fed. R. Evid. 403.”⁴⁴ In a post-*Daubert* case, *In re Paoli Railroad Yard PCB Litigation (Paoli II)*,⁴⁵ Judge Becker recalled his pre-*Daubert* discussion in *Downing* of the relationship between 702 and 403, saying that the point in *Downing* was that 702 “partly incorporates Rule 403 analysis

⁴⁰ Federal Rule of Evidence 401 reads, “‘Relevant evidence’ means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.” FED. R. EVID. 401.

⁴¹ 509 U.S. at 595.

⁴² 753 F.2d 1224 (3d Cir. 1985).

⁴³ *Id.* at 1239 (internal citations omitted).

⁴⁴ *Id.* at 1242.

⁴⁵ 35 F.3d 717 (3d Cir. 1994).

but leaves some room for Rule 403 to operate independently.”⁴⁶ He went on to describe how the *Daubert* Court treated the *Downing* view of this relationship:

In *Daubert*, the Supreme Court seems to have inverted our view that much of Rule 403 analysis conflates into Rule 702; rather the Court seems to have conflated the confusion/overwhelming impact prong of our Rule 702 analysis into its Rule 403 analysis. The *Daubert* Court did not mention the confusion/overwhelming prong when discussing Rule 702 but did provide support for application of essentially similar analysis under the rubric of Rule 403.⁴⁷

Since degree of reliability has a bearing upon the probative value of expert testimony, and the risk of misleading the jury and other 403 factors might lead to exclusion under Rule 403 as noted by the Court in *Daubert*, the reliability determination may simply fold into the 403 analysis creating a sliding scale of reliability under 702.⁴⁸ Under this approach, if the degree of reliability (probative value) sufficiently offsets the Rule 403 risks associated with the expert testimony, the evidence is reliable enough to be admissible under 403. Even expert testimony of the highest risk would not lead to exclusion if the expert testimony were highly reliable.⁴⁹ This determination would be consistent with the demand for more than simple relevancy in reaching reliability determinations and the need to observe the distinction between reliability and sufficiency. Although the Court in *Daubert* perceives a relationship between 403 and the reliability determination under 702 and moves it, perhaps, one step beyond *Downing*, the recognition of reliability as a factor in determining the probative value of this species of evidence under 403 analysis remains underdeveloped.

THE ARGUMENTS

In *Reliability and the Admissibility of Experts*, Professor Dale Nance’s

⁴⁶ Judge Becker used as an example of this relationship the exclusion of “an expert’s critique of eyewitness testimony even though the critique met the requirements of Rule 702 if there was evidence of defendant’s guilt other than eyewitness testimony which would make efforts to criticize eyewitness testimony a waste of time.” *Id.* at 746.

⁴⁷ *Id.* at 746-47.

⁴⁸ See Calvin William Sharpe, *Two-Step Balancing and the Admissibility of Other Crimes Evidence: A Sliding Scale of Proof*, 59 NOTRE DAME L. REV. 556, 585-89 (1984) (describing proof of other crimes as a sliding scale of proof).

⁴⁹ See Newell H. Blakely, *Article IV: Relevancy and Its Limits*, 30 HOUS. L. REV. 281, 317 (1993) (including a table showing high probative value offsetting high, mid, or low prejudice).

project is to provide “a careful analysis of the contours and functions of the reliability concept . . . and its relationship to the purposes of admissibility rules.”⁵⁰ He argues generally that “courts and commentators should disavow an all-or-nothing concept of reliability” in favor of a concept that recognizes reliability as a non-binary variable characteristic, thus recognizing that the binary mode of expression in Rule 702 indicates the results of an unarticulated analytical framework.⁵¹

Nance argues in the following terms that 702 reliability instead calls for a comparative evaluation of expert testimony:

Much more important in articulating a meaningful content for the reliability requirement of Rule 702, and more likely to result in the exclusion of proffered expertise, is the idea that evidence may be excluded to encourage the presentation of better evidence, evidence that is more probative, or less costly for the tribunal, or otherwise presenting a more favorable balance between the two.⁵²

Specifically, he argues that the reliability element of 702 requires the trial judge to secure the best (most reliable) evidence that is reasonably available to the proponent of expert testimony.⁵³ For Nance, this approach cures many ills including, importantly, the misplaced concern about jury credulity.⁵⁴

Professor Nance recognizes the affinity between his suggested approach and Rule 403. He acknowledges that 403 analysis addresses traditional concerns about reliability,⁵⁵ and he concedes that as a well-understood rule, 403 is a good starting point for 702 analysis.⁵⁶ He also acknowledges McCormick’s entreaty that balancing the probative value and prejudice of scientific evidence “offers a more honest and sensitive basis for making admissibility decisions than the more cramped tests that have characterized this area of the law.”⁵⁷ However, Professor Nance reads the cases and Rule 702 as requiring a content for reliability that is distinctive from 403 analysis.⁵⁸ A

⁵⁰ Nance, *supra* note 39, at 192.

⁵¹ *Id.* at 193.

⁵² *Id.* at 240.

⁵³ *Id.* at 225.

⁵⁴ *Id.* at 227.

⁵⁵ *Id.* at 224 (identifying concerns about jury ability to properly assess probative value, conservation of jury and fact-finder resources, and the availability of better evidence).

⁵⁶ Nance, *supra* note 226.

⁵⁷ *Id.* at 220 (referring to 1 MCCORMICK ON EVIDENCE, *supra* note 4, § 203).

⁵⁸ Referring to the quote from MCCORMICK ON EVIDENCE, Nance says, “How then shall we complete the analysis in such a way as to give due respect to the teachings of

careful search of *Daubert* yields little support for such a view. Indeed, Nance, in dispelling the notion of jury credulity and advocating the use of 403, points to the *Daubert* language affirming the role of 403 in excluding unduly prejudicial, confusing or misleading evidence.⁵⁹ By reminding trial judges “assessing a proffer of expert scientific testimony under Rule 702” to be mindful of other applicable rules, the Court affirms a relationship between 702 and 403 and endorses the following view of Judge Weinstein:

Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it. Because of this risk, the judge in weighing possible prejudice against probative force under Rule 403 of the present rules exercises more control over experts than over lay witnesses. The judge may insist, for example, on strong guarantees that tests relied on by an expert were properly conducted since a careless laboratory is a terrible hazard to justice.⁶⁰

The insistence on the proper conduct of tests that Judge Weinstein described in the example seems to be an effort to establish the reliability (probative value) of the expert’s testimony. The counterweight of possible prejudice that might lead to exclusion under 403 is what makes the admissibility finding essentially the same as the sufficient reliability finding under 702. It is this analysis, endorsed by the *Daubert* Court, that Judge Weinstein used in the *Agent Orange* litigation, where he excluded animal studies with the following reasoning:

There is no evidence that plaintiffs were exposed to the far higher concentrations involved in both the animal and industrial exposure studies. The animal studies are not helpful in the instant cases because they involve different biological species. They are of so little probative force and are so potentially misleading as to be inadmissible.⁶¹

In this pre-*Daubert* case, Judge Weinstein, discussing the trend toward replacing *Frye* analysis with 702 analysis, said that the “general acceptance” standard governing the admissibility of novel scientific evidence had been replaced by “a balancing of the relevance, reliability, and helpfulness of the evidence against the likelihood of

the Supreme Court and the mandate of Congress in amended Rule 702?” *Id.*

⁵⁹ 509 U.S. at 595.

⁶⁰ Jack B. Weinstein, *Rule 702 of the Federal Rules of Evidence is Sound; It Should Not Be Amended*, 138 F.R.D. 631, 632 (1992). The Court’s endorsement of this view also seems to evince a concern about jury credulity that Professor Nance perceived in the Court’s opinion. See Nance, *supra* note 39, at 227-28.

⁶¹ *In re “Agent Orange” Prod. Liab. Litig.*, 611 F. Supp. 1223, 1241 (E.D.N.Y. 1985), *aff’d*, 818 F.2d 145 (2d Cir. 1987) (internal citations omitted).

waste of time, confusion and prejudice.”⁶² Judge Weinstein elaborated, “[w]hen either the expert’s qualification or his testimony lie at the periphery of what the scientific community considers acceptable, special care should be exercised in evaluating the reliability and probative worth of the proffered testimony under Rules 703 and 403.”⁶³ This discussion explicitly links reliability and probative value as part of the 403 analysis. Indeed, Professor Nance points to this connection in making the observation that “expertise that is very unreliable may be excluded under Rule 403 as simply a waste of time.”⁶⁴

Yet the reliability determination under 702 does involve a distinct brand of 403 analysis. Expert testimony raises a special concern under 403, because its terms must apply to specialized knowledge rather than lay knowledge. And reliability is an idiosyncratic measure of expert testimony’s probative value. *Daubert*’s five-factor reliability analysis purports to help judges ask the right questions in assessing the probative value of expert testimony. It is one thing to determine the probative value of an earlier similar theft offered to show the existence of intent to commit a charged theft.⁶⁵ Judges can focus on the similarity of the circumstances and other factors and reach a fairly routine assessment of probative value. It is quite another to determine the probative value of an opinion that Bendectin can cause birth defects based partially on the reanalysis of previously published epidemiological studies.⁶⁶ The *Daubert* and 702 factors help judges to conduct that more complicated assessment of probative value.

Moreover, the potential for unfair prejudice and misleading the jury always lurks with the admission of expert testimony.⁶⁷ As already noted, Professor Nance points out that expert testimony might be excluded under 403 “as simply a waste of time.”⁶⁸ He also observes that expert testimony may require “greater judicial management or monitoring.”⁶⁹ Why this heightened burden? Nance cites the risk

⁶² 611 F. Supp at 1242.

⁶³ *Id.*

⁶⁴ Nance, *supra* note 39, at 231.

⁶⁵ See *United States v. Beechum*, 582 F.2d 898 (5th Cir. 1978).

⁶⁶ See generally *Daubert*, 509 U.S. 579 (involving birth defects caused by Bendectin).

⁶⁷ See *Daubert*, 509 U.S. 579 (1993); see also *In re Paoli R.R. Yard*, 35 F.3d at 717 (3d Cir. 1994); *United States v. Downing*, 753 F.2d 1224 (3d Cir. 1985); *In re “Agent Orange,”* 611 F. Supp. at 1241 (E.D.N.Y. 1985).

⁶⁸ Nance, *supra* note 39, at 231.

⁶⁹ *Id.* at 232. Judge Becker identifies this concern as an independent 403 concern that does not involve an overlap with 702. See *supra* note 46 and

factors of greater manipulability and the “production” of expert testimony, characteristics of expert testimony that are checked in non-expert testimony by the requirement of personal knowledge.⁷⁰ Using the reliability factors as tools, *Daubert* insists upon a balancing analysis that recognizes the special probative components of expert testimony as well as the inherent risks associated with such testimony.

The specialized character of 403 analysis involving expert testimony (the use of *Daubert* factors) is one reason that the 702 reliability overlay is distinctive from the 403 analysis of non-expert testimony.⁷¹ A second reason is the burden of proof. The opponent

accompanying text.

⁷⁰ Responding to his own question concerning the need for heightened scrutiny of expert testimony, Nance reasons:

I am inclined to believe that the answer lies in the greater manipulability of such evidence, as compared to most non-expert testimony. The supply of non-expert testimony, limited as it is by the requirement of first-hand or “personal” knowledge, tends to be fixed by the litigated events. Typically, only a small number of persons will have witnessed the events being litigated. For practical purposes, expert testimony knows no such limitation. When a matter is thought by counsel to be amenable to expert assistance, there are often numerous specializations and hundreds or thousands of practitioners thereof who might be called to testify. Data can often be gathered and experiments can sometimes be conducted in anticipation of trial. Put simply, expert testimony is *produced* in a way that most non-expert testimony is not, coaching of lay witnesses notwithstanding. And there are obvious and powerful distorting and biasing forces at work in this production process, much of which occurs after the events being litigated and with an eye toward trial.

Nance, *supra* note 39, at 232 (footnotes omitted).

⁷¹ In advocating a standard not simply redundant of Rule 403 Professor Nance cites a familiar principle of statutory construction and expresses skepticism that the Supreme Court in *Daubert* simply intended “to deliver the message that trial courts have been improperly applying Rule 403 in the context of expert testimony.” *Id.* at 226. However, 702 viewed as a specialized application of 403 is not simply redundant of the typical 403 analysis. The difference lies in the nature of the evidence rather than the rule. Nance acknowledges the propriety of a heightened burden for the admissibility of expert testimony because of its “greater manipulability” when compared to non-expert testimony. *Id.* at 232. The Court in *Daubert* evinces a similar concern about the vulnerability of the jury in the face of expertise, when it enunciated criteria that would insulate the jury from this undue influence by safeguarding reliability. The Court generated the *Daubert* factors, even though it dismissed concerns about the abandonment of the *Frye* test as growing out of an “overly pessimistic [view] about the capabilities of the jury.” *Daubert*, 509 U.S. at 595-96; see Nance, *supra* note 39, at 228. Perhaps most importantly, the Court adopted Judge Weinstein’s observation about the powerful and misleading quality of expert evidence and the greater control that Rule 403 gives judges over experts than non-experts. *Daubert*, 509 U.S. at 595. Moreover, as Judge Becker suggested in *Downing* “[t]he degree to which an unwarranted ‘aura of reliability’ attaches to scientific evidence will naturally vary with the type of evidence.” 753 F.2d at 1239. While jury credulity may not be a significant problem with scientific trace evidence, see Dale A.

of admissibility has the burden of persuasion in cases dealing with non-expert testimony.⁷² This is a heavy burden, since the rule favors admissibility and the opponent bears the risk of non-persuasion.⁷³ However, the proponent of the expert testimony has the burden of proving sufficient reliability to offset 403 risks.⁷⁴ The balancing standard under 403 does not change.⁷⁵ But, if the judge is at or below equipoise in determining whether the testimony sufficiently offsets those risks to warrant admitting the testimony, the proponent loses and the evidence is excluded. This heightened burden on the proponent of expert testimony under 702 eases the exclusion burden of the opponent under 403.⁷⁶

It is also easier to see under a 403 analysis that the 403 dangers set a minimum threshold of reliability, one that does not fit Professor Nance's description of "an invariant threshold that applies across disciplines and across litigation contexts."⁷⁷ Rather, it varies with the

Nance & Scott B. Morris, *An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Large and Quantifiable Random Match Probability*, 42 JURIMETRICS J. 403 (2002), it may well be a substantial problem with polygraph evidence. See discussion of the polygraph cases, *infra* notes 125-43 and accompanying text.

⁷² See 1 STEPHEN A. SALTZBURG ET AL., FEDERAL RULES OF EVIDENCE MANUAL § 403.02[1] (8th ed. 2002) (describing rule 403 as creating a presumption); see also ALLEN, *supra* note 1, at 873 (explaining the operation of presumptions that shift the burden of proof to the opponent of the evidence).

⁷³ 2 MCCORMICK ON EVIDENCE, *supra* note 4, § 336.

⁷⁴ See DAVID L. FAIGMAN ET AL., SCIENCE IN THE LAW: STANDARDS STATISTICS AND RESEARCH ISSUES § 1-3.1.2 (2002).

⁷⁵ Federal Rule of Evidence 403 states, "Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence." FED. R. EVID. 403.

⁷⁶ Nance applauds this function of the burden of proof under 702 in placing the "stricter demands of *Daubert* [and] *Kumho Tire*" on the proponent of expert testimony. Nance, *supra* note 39, at 235. However, for Nance this burden would only seem to be triggered where the proponent fails to offer more reliable expertise. Noting that the proponent has the burden to establish reliability, Nance says:

Applying this idea here, if the trial judge concludes that the challenged expertise is discernibly less reliable than other expertise offered by the proponent, then the burden would rest on the proponent to convince the judge that it would be a mistake to ignore the challenged expertise because its consideration will materially assist the trier of fact to render an appropriate verdict within the constraints imposed by the process of trial. Doubt on the matter would be resolved in favor of exclusion, reversing the burden as compared to Rule 403. Distinguishing this analysis from that of Rule 403 would help to maintain the representational viewpoint described earlier.

Id.

⁷⁷ *Id.* at 221.

characteristics of each proffer.⁷⁸

The idea of a specialized 403 analysis is not inconsistent with Professor Nance's better evidence principle. While carefully crafting his argument for a distinctive analysis of reliability under 702, Professor Nance rejects a standard that is "simply redundant of Rule 403."⁷⁹ He dismisses the notion that the Supreme Court in *Daubert* simply "intend[ed] to deliver the message that trial courts have been improperly applying Rule 403 in the context of expert testimony" and notes a "general consensus that *Daubert* did not adopt the position advanced most conspicuously by Dean McCormick."⁸⁰ For Nance what gives meaningful content to the reliability requirement of 702 "is the idea that evidence may be excluded to encourage the presentation of better evidence, evidence that is more probative or less costly for the tribunal, or otherwise presenting a more favorable balance between the two."⁸¹ He notes that "[t]his idea is present in a wide variety of rules, including Rule 403."⁸² Indeed, the ACN for 403 concludes with the following sentence: "The availability of other means of proof may also be an appropriate factor [in reaching a decision whether to exclude evidence under 403]."⁸³ Professor Nance makes a persuasive case for giving prominence to the existence of better evidence in reaching the reliability determination.⁸⁴ However, the analysis is likely to be less confusing to judges and more faithful to the spirit of *Daubert*, *Kumho*, and the 702 amendments if its 403

⁷⁸ This observation also has implications for Nance's view that in the absence of available alternatives, the court should admit the proffer. See David L. Faigman, *Expert Evidence in Flatland: The Geometry of a World Without Scientific Culture*, 34 SETON HALL L. REV. 253 (2003).

⁷⁹ Nance, *supra* note 39, at 226.

⁸⁰ *Id.*

⁸¹ *Id.* at 240.

⁸² In noting support for this point, Nance cites *United States v. Old Chief*, 519 U.S. 172 (1997), which applied this understanding of 403 to hold that the trial court abused its discretion by not accepting a less prejudicial stipulation rather than the more prejudicial conviction record of the defendant. It should be noted that the existence of a better evidence alternative in *Old Chief* reduced the probative value of the prosecutor's submission, making exclusion more likely under 403, much like the absence of *Daubert* factors in a 702 reliability ruling might operate.

⁸³ See *Downing*, 753 F.2d at 1243 ("The availability of other methods that would serve the purposes for which the appellant seeks to introduce expert testimony may also serve to justify exclusion under Rule 403 . . .").

⁸⁴ Professor Nance bases this reliability requirement on the courts' "need to be demanding consumers of expertise, especially those kinds of expertise that find their reason for existence in the demand for expert testimony." Nance, *supra* note 39, at 240. He adds that the better evidence consideration "represents the primary, if not exclusive argument structure appropriate for the reliability inquiry of Rule 702." *Id.* at 241.

provenance is clear.

Professors David L. Faigman, David H. Kaye, Michael J. Saks, and Joseph Sanders have adopted a similar better evidence principle:

[W]e do endorse a *better evidence principle* in our analyses of *certain* issues in the law of expert testimony. That is to say, we believe that there are circumstances in which a court properly may exclude proffered evidence when other evidence of greater probative value is or should be available.⁸⁵

Like Nance, these professors believe that a better evidence principle will produce the beneficial effects of increased accuracy in factfinding as well as enhanced research and expert knowledge.⁸⁶

While they describe *Daubert* as raising the bar to admissibility by placing a heavier cognitive burden on judges,⁸⁷ they quote Judge Weinstein's analysis in the *Agent Orange* litigation as an example of courts requiring better evidence.⁸⁸ As already pointed out, Judge Weinstein in that case used a 403 analysis in ruling the animal studies inadmissible.⁸⁹

Professor Edward J. Imwinkelried in *A Final Comment—The Importance of the Procedural Framework*,⁹⁰ responds to the better evidence principle proposed by Professors Faigman, Kaye, Saks and Sanders, taking delight in their rejection of a “best evidence test” and agreeing with their analytical outcomes in exemplary cases where they applied the better evidence principle.⁹¹ However, Professor Imwinkelried argues that such a better evidence principle is unnecessary to the analysis and that the announcement of a better evidence principle “further complicate[s] the analysis of the admissibility of scientific testimony.”⁹² In nicely setting forth the procedural elements of judicial reliability factfinding under 104(a), Imwinkelried acknowledges the significance to this inquiry of the presence of better evidence.⁹³ He also indicates that the judge, in accordance with 104(a), evaluates the proponent's foundation, laid “as a means

⁸⁵ David L. Faigman et al., *How Good is Good Enough?: Expert Evidence Under Daubert and Kumho*, 50 CASE W. RES. L. REV. 645, 654 (2000) (emphasis in the original).

⁸⁶ *Id.* at 667.

⁸⁷ *Id.* at 656.

⁸⁸ *Id.* at 659.

⁸⁹ *See In re “Agent Orange,”* 611 F. Supp. at 1241.

⁹⁰ Edward J. Imwinkelried, *A Final Comment—The Importance of the Procedural Framework*, 50 CASE W. RES. L. REV. 669 (2000).

⁹¹ *Id.* at 669.

⁹² *Id.* at 670.

⁹³ *Id.* at 678-79

to the end of ensuring reliability,”⁹⁴ for sufficient probative value.⁹⁵ Though Professor Imwinkelried does not mention 403 in this piece, it is implicit in his discussion.⁹⁶

Professor Michael H. Graham in *The Expert Witness Predicament: Determining ‘Reliable’ Under the Gatekeeping Test of Daubert, Kumho, and Proposed Amended Rule 702 of the Federal Rules of Evidence*,⁹⁷ identifies two meanings of the term “reliable” as used in *Daubert* and *Kumho*: (1) “‘reliable’ . . . taken to mean that the explanative theory [theory, technique, reasoning, methodology, etc.] actually works, i.e., produces a correct, accurate, truthful, or valid conclusion,”⁹⁸ and (2) “‘reliable’ refer[ring] to meriting confidence worthy of dependence or reliance, i.e., possesses sufficient assurance of correctness to warrant acceptance by the trier of fact.”⁹⁹ He notes that the two definitions call for separate analyses and criticizes both *Daubert* and *Kumho* for creating confusion by using both meanings.¹⁰⁰ While approving of the trial and appellate courts’ apparent favoring of the second definition, Graham never identifies that definition as part of the assessment of probative value in the 403 analysis.¹⁰¹

⁹⁴ *Id.* at 671.

⁹⁵ *Id.* at 675.

⁹⁶ In an earlier article to which Faigman, *supra* note 84, responds, Professor Imwinkelried argued that it is appropriate for the opponent to point out the absence of the best evidence to attack the legal sufficiency or weight of scientific testimony, but not its admissibility—“either an order of preference among types of proffered scientific evidence or a regulation of the sufficiency of the foundation for proffered scientific testimony.” Edward J. Imwinkelried, *Should The Courts Incorporate A Best Evidence Rule Into The Standard Determining The Admissibility of Scientific Testimony?: Enough Is Enough Even When It Is Not The Best*, 50 CASE W. RES. L. REV. 19, 49 (1999). In that article Professor Imwinkelried expressed the link between the reliability foundational requirement and probative value as follows:

Faced with a foundational objection, the trial judge must determine whether the proponent’s predicate has enough probative worth to justify the proffered opinion. The judge can make that determination if he or she decides whether the expert has properly applied the scientific methodologies and, if so, how significant the expert’s findings are.

Id. at 47.

⁹⁷ Michael H. Graham, *The Expert Witness Predicament: Determining ‘Reliable’ Under the Gatekeeping Test of Daubert, Kumho, and Proposed Amended Rule 702 of the Federal Rules of Evidence*, 54 U. MIAMI L. REV. 317 (2000).

⁹⁸ *Id.* at 319, 336.

⁹⁹ *Id.* at 336.

¹⁰⁰ *Id.* at 336-37.

¹⁰¹ Professor Nance endorses Professor Graham’s approval of the second definition as a variable rather than dichotomous approach to reliability. Nance, *supra* note 39, at 222. However, Nance ultimately finds Graham’s “sufficient assurances” definition unsatisfactory due to the risk of confusion with “sufficiency,” its affinity with the *Frye* test and its problems, and its failure to account for the better

In *Daubert: Interpreting The Federal Rules of Evidence*,¹⁰² Professor Paul C. Giannelli argued that a relevancy approach that would treat reliability as an aspect of probative value “differs significantly from the reliability approach adopted in *Daubert* and it does not depend on Rule 702.”¹⁰³ Professor Giannelli’s rejection of the relevancy approach appears to be based in part on Professor Strong’s articulation of it as follows:

[S]cientific evidence, like other evidence, requires the striking of a balance between the probative worth of the evidence and its capacity to confuse or prejudice the jury [I]n the case of scientific evidence the court will generally be *forced* to accept the probative value of the evidence as what a qualified expert testifies it to be.¹⁰⁴

Referring to Strong’s belief that the expert qualification requirement “was a substantial barrier to junk science”—“the qualification of the expert presumptively qualifies the technique,” Giannelli argued that “[t]his formulation of the relevancy approach makes the trial court too dependent on the testifying expert.”¹⁰⁵ Giannelli’s rejection of the relevancy approach is also based on three other factors. First, he finds “‘weighing’ probative value against factors such as misleading the jury . . . frequently illusory,” noting that a judge believing a scientific technique to be “reliable (when it is not) . . . will not appreciate its misleading character.”¹⁰⁶ Second, Professor Giannelli makes the point that 403’s bias toward admissibility “further erode[s] the barriers to admissibility.”¹⁰⁷ Third, Giannelli finds troubling the “abuse of discretion” standard of review of trial court rulings saying that “[u]nder this approach, it would not be wrong for one trial judge to admit polygraph evidence while another judge excluded it. . . . This approach is inconsistent with *Daubert*, which I believe is more demanding.”¹⁰⁸ Professor Giannelli goes on to note *Daubert*’s imposition of an independent assessment obligation upon the trial judge despite the expert’s claims of reliability and the Court’s willingness “to pay the price for a demanding standard, by noting that ‘inevitably on occasion [its approach] will prevent the jury from

evidence principle. *Id.* at 223.

¹⁰² Giannelli, *supra* note 11.

¹⁰³ *Id.* at 2009-10.

¹⁰⁴ *Id.* at 2010 (quoting John W. Strong, *Questions Affecting the Admissibility of Scientific Evidence*, 1970 U. ILL. L.F. 1, 22) (emphasis added).

¹⁰⁵ *Id.* at 2010-11.

¹⁰⁶ *Id.* at 2011.

¹⁰⁷ *Id.*

¹⁰⁸ Giannelli, *supra* note 11, at 2011.

learning of authentic insights and innovations.’”¹⁰⁹

Daubert and its progeny show that Professor Giannelli’s concerns do not discredit approaching reliability under 702 as an aspect of probative value. First, *Daubert* rejects the notion that expertise alone establishes the reliability of expert testimony. In fact, even though the Court acknowledged the impressive credentials of Petitioners’ eight experts, it remanded the case for a determination of reliability based on the *Daubert* factors.¹¹⁰ Second, if a mistaken reliability finding prevents the judge from appreciating whether the jury is misled, it would seem to make little difference whether such a finding is deemed to be an aspect of probative value under a 403 analysis or an independent requirement under 702. In either case the reliability finding (mistaken though it is) would lead to admission. Third, even though 403 favors admissibility, the *Daubert* Court makes it clear that simple relevancy is different from reliability.¹¹¹ The reliability factors call for sufficient probative value to offset (I argue under the 403 standard) the inherently unfair prejudice and other risk factors (both inherent and case-specific) that characterize expert testimony. Importantly, unlike the 403 analysis of non-expert testimony,¹¹² the proponent of expert testimony has the burden of proving reliability, i.e. sufficient probative value.¹¹³ This makes the specialized application more demanding than the routine 403 balancing rule.¹¹⁴ Finally, having the benefit of 20/20 hindsight, we know that the trial court’s reliability determination, like the 403 determination, is subject to an abuse of discretion standard that insulates a judge’s decision to admit or exclude evidence.¹¹⁵

In their excellent treatise, Professors Faigman, Kaye, Saks, and Sanders noted that the *Daubert* Court “devoted relatively little attention to the balance of probative value and unfair prejudice encapsulated in Rule 403.”¹¹⁶ They suggested that “this Rule might prove to be one of the most important tools lower courts have for

¹⁰⁹ *Id.*

¹¹⁰ *Daubert*, 509 U.S. at 598.

¹¹¹ See FED. R. EVID. 401.

¹¹² See *supra* note 71.

¹¹³ *Daubert*, 509 U.S. at 590-97; see 1 SALTZBURG, *supra* note 71, § 104.02[9].

¹¹⁴ But see Eleanor Swift, *One Hundred Years of Evidence Law Reform: Thayer’s Triumph*, 88 CAL. L. REV. 2437 (2000) (tracing the current excess of trial court discretion under the Federal Rules of Evidence to James Bradley Thayer and specifically criticizing the losses in consistency, predictability, and integrity of our adjudication caused by the *Daubert*, *Joiner*, and *Kumho* decisions).

¹¹⁵ See *Joiner*, 522 U.S. at 146.

¹¹⁶ FAIGMAN ET AL., *supra* note 74, § 1-3.8.

managing scientific evidence.”¹¹⁷ They posit a case where the judge might decide under 702 and 104(a) that the proponent has established the validity of the evidence by a preponderance of the evidence but believe that “it is not valid enough, in light of the dangers associated with its use.”¹¹⁸ This approach exemplifies the role of reliability as an aspect of probative value, whose sufficiency is a function of the dangers associated with the use of expert testimony.¹¹⁹

Kumho makes it clear that there is “no relevant distinction” in reliability analysis between scientific, technical and other specialized knowledge.¹²⁰ It is also clear that reliability assessment requires

¹¹⁷ *Id.*

¹¹⁸ *Id.* Even though the results of a polygraph examination present a prime example of the kind of technology that is regulated through 403 balancing, the authors add that “[v]irtually all other forms of scientific evidence present similar difficulties and opportunities.” *Id.* The authors say the following about polygraphy in the context of 403 analysis:

Although the research supporting the validity of polygraphy remains controversial, significant research has been conducted on the validity and reliability of polygraph tests. Despite the flaws associated with this research, a court could reasonably conclude that some form of polygraphy was more likely than not valid. But few courts, if any, would complete their scrutiny there.

Polygraphy is potentially awesome technique that might displace jurors’ traditional task of evaluating credibility. A large percentage of courts and observers fear the overwhelming impact polygraphy might have, causing jurors to overlook the significant errors associated with even the best application of the technology. The regulation of this technology is largely accomplished through the balancing mechanism provided by 403.

Id. (footnotes omitted).

¹¹⁹ The approach suggested here differs from that of Faigman, since it views 702 gatekeeping as a part of 403 analysis, rather than being in lockstep. Under this view, any required adjustment in the showing of validity is made in light of the 403 dangers. This analysis is similar to Nance’s comparative reliability of better evidence approach; however, the existence of better evidence is only one factor in the 403 balance. *Cf.* FAIGMAN ET AL., *supra* note 74, § 1-3.8 n.230 (“Rule 403 only comes into play if the court finds that the evidence is ‘sufficiently’ reliable under Rule 702. Thus, Rule 403 provides additional power to keep evidence out after Rule 702 gatekeeping is done. Yet, if the Rule 702 threshold is substantially higher than mere relevancy, then the courts’ ability to adjust the showing of validity to the use and context of the evidence is limited.”).

¹²⁰ 526 U.S. at 147. As Faigman says:

In asking whether the expertise is “science” or “non-science,” courts have asked the wrong question. Rather, in all cases where expert testimony is proffered courts should be inquiring into the methods experts are using, whether expert judgments are based on experimental research, clinical evaluation, or other types of experience.

FAIGMAN ET AL., *supra* note 74, § 1-3.5, at 41.

neither proof of the accuracy of expert testimony nor exclusion of “shaky evidence,” leaving the question of how much reliability is enough.¹²¹ The trial court’s analysis should focus on whether the expert testimony is sufficiently reliable to offset the dangers associated with the testimony.¹²² Though this sufficiency approach should consider the availability of better evidence as suggested by Professor Nance, the existence of such evidence should not be the sole determinant of probative value.¹²³ This approach also suggests that smaller degrees of reliability may be tolerable, where associated dangers are lower, posing a lower threat to accuracy in factfinding.¹²⁴

THE POLYGRAPH CASES

Under the 403 approach, expert testimony that is not sufficiently supported by the data or that amounts to unsupported speculation lacks sufficient probative value to justify time and risk of overvaluation that its production might entail.¹²⁵ Hence, the cases may be examined against three tests that reveal the link between reliability and 403 analysis. First, admissibility decisions that analyze reliable testimony as admissible and unreliable testimony as inadmissible under 403 tend to substantiate reliability analysis as a specialized application of 403. Second, the exclusion of reliable testimony under 403 would suggest a distinction between reliability and 403 analysis. Third, the converse is also true—the admission of expert testimony under a 403 analysis despite a finding of unreliability would demonstrate a distinction between reliability and 403 analysis.

The polygraph cases most dramatically demonstrate the demands that 403 places on reliability. The exemplary case showing the connection between reliability and 403 analysis is *United States v. Lea*.¹²⁶ In that case a criminal defendant challenged the trial court’s exclusion of exculpatory polygraph evidence in part because the

¹²¹ See *Daubert*, 509 U.S. at 595.

¹²² Cf. FAIGMAN ET AL., *supra* note 74, § 1-3.5.1 (defining sufficiency as “how much evidence it takes before we believe a certain proposition”).

¹²³ Some combination of the *Daubert* and 702 factors will help determine probative value. See Faigman, *supra* note 78, at 261 (arguing that judges should ask “whether better evidence *should* be available”) (emphasis in original).

¹²⁴ See FAIGMAN ET AL., *supra* note 74, § 1-3.5.1. (noting that courts appear to be expecting better research in criminal cases, “the more likely the jury is to be overwhelmed by the expert opinion”).

¹²⁵ See *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997) (noting that opinion connected with data only by *ipse dixit* or involving “too great an analytical gap between the data and the opinion proffered” may lack reliability).

¹²⁶ 249 F.3d 632, 639-40 (7th Cir. 2001).

judge had performed a reliability rather than a 403 analysis. The defendant had been convicted and sentenced to 36 months imprisonment, a year of supervised release, and restitution in the amount of \$2.2 million for the sabotage of a former business associate's business product that had resulted in losses of \$2.5 million. The defense had been that an employee of the former associate had committed the sabotage and not the defendant. Part of the defendant's proof was the opinion of a polygraph examiner that the employee's answers to the polygraph examination about his involvement in the sabotage indicated deception. However, the examiner had been unable to conduct a post-examination interview with the employee, because he had refused to participate. After questioning the examiner the trial court excluded the opinion on the grounds that the examiner had "failed to establish the reliability of [the examiner's] opinion resulting from his polygraph examination of [the employee]." ¹²⁷ The trial court had found that the examiner "could only speculate as to the accuracy of the polygraph examination he had performed," and was "unaware of whether there were any known statistics on the accuracy rate of the methodology employed in examining [the employee]." ¹²⁸

On appeal, based on an earlier Seventh Circuit decision holding that trial courts must determine the admissibility of polygraph evidence by delicately balancing 403 factors, the defendant in *Lea* argued that exclusion could only be based on Rule 403 factors and that the trial court had inappropriately excluded the evidence because of reliability concerns. ¹²⁹ This argument forced the Seventh Circuit to address the relationship between 702 reliability and 403 analysis. Conceding the defendant's point about the applicability of 403, the Seventh Circuit said the following:

While our recent case law has not explicitly retained the notion that reliability concerns can factor into the admissibility decision, we note that 403 allows for the exclusion of otherwise relevant evidence if the probative value is 'substantially outweighed by the danger of . . . misleading the jury.' ¹³⁰

Citing the concerns about the "aura of infallibility attending polygraph evidence" and misled juries as well as juries giving

¹²⁷ *Id.* at 637.

¹²⁸ *Id.* at 637-38

¹²⁹ *See* United States v. Olson, 978 F.2d 1472, 1480 (7th Cir. 1992) ("When dealing with the admissibility of polygraph evidence, and the accuracy thereof, the trial court must engage in a delicate balancing of many factors including probative value, prejudicial effect, confusion of the issues, misleading the jury, and undue delay.").

¹³⁰ *Lea*, 249 F.3d at 639.

excessive weight to polygrapher opinions, the court continued:¹³¹

Such concerns are undoubtedly heightened when the reliability of the particular examination is called into question. As the reliability of the evidence decreases, the likelihood increases that the probative value may be substantially outweighed by the prejudicial effect. Thus, while reliability is an explicitly referenced concern that is appropriately discussed in a *Daubert* framework, the issue may also become an integral part of a 403 inquiry.¹³²

The court also referenced as consistent with this analysis one of its earlier decisions where “[t]he court had examined the reliability concerns under the *Daubert* framework, and determined that the ‘reliability problems rendered the probative value minimal . . . [while] there was a danger that the jury would consider the polygraph test to be conclusive regarding [the witness’s] veracity.’”¹³³ The *Lea* court noted that the trial court had used *Daubert* as a guide to focus on the known or potential error rate and considered the examiner’s inability to complete the examination. The Seventh Circuit concluded that “[t]hose factors reduced the reliability of [the examiner’s] opinion, tipping the balance under Rule 403 in favor of exclusion.”¹³⁴

Like the Seventh Circuit cases, *United States v. Posado*,¹³⁵ *United States v. Cordoba*,¹³⁶ and *United States v. Waters*¹³⁷ are three post-*Daubert* cases that demonstrate the 403 link to reliability. Each of these cases involved the criminal defendant’s proffer of exculpatory polygraph evidence. In *Posado*, the Fifth Circuit reversed the trial court’s refusal to conduct a *Daubert* hearing on the admissibility of the polygraph examiner’s expert testimony. The court noted the “tremendous advances in polygraph instrumentation and technique in the years since *Frye*” and set forth *Daubert*’s interpretation of the reliability requirement under 702. The court ultimately remanded the case to

¹³¹ The *Lea* court noted that the Supreme Court in *United States v. Scheffer*, 523 U.S. 303 (1998), voiced these concerns.

¹³² *Lea*, 249 F.3d at 639.

¹³³ *Id.* (quoting *United States v. Taylor*, 154 F.3d 675, 683 (7th Cir. 1998)). In *Taylor*, the trial court had specifically pointed to the reliability problem caused by the examiner’s use of “‘stock’ questions in the test rather than questions tailored to the circumstances of [the] case [and] a subjective visual scoring technique in calculating the results rather than the more reliable objective numerical scoring system.” 154 F.3d at 683.

¹³⁴ *Lea*, 249 F.3d at 639.

¹³⁵ 57 F.3d 428 (5th Cir. 1995).

¹³⁶ 194 F.3d 1053 (9th Cir. 1999).

¹³⁷ 194 F.3d 926 (8th Cir. 1999).

the trial court for a determination of the relevance and reliability of the polygraph evidence, but it removed the *per se* barrier. It also cited the evidence in support of the examination's accuracy: (a) the office of technology assessment studies showing accuracy rates of 58-98 percent and higher, and (b) other studies and rates higher than 90%. Based on these figures the court assumed reliability and proceeded to demonstrate what factors should be considered in the 403 analysis, including the mesmerizing prejudicial effect.¹³⁸ It considered such probative factors as the opportunity extended to the prosecutor to participate in the examination, offering the evidence in a pre-trial setting before a judge rather than in a trial before the jury, and the credibility problem giving rise to a need for the evidence. Following this exercise, the court announced that it was leaving it up to the trial court to determine reliability and relevance.

The court's approach in assuming reliability before proceeding to a 403 analysis in *Posado* suggests a distinction between 702 and 403 analysis. Unlike the Seventh Circuit in *Lea*, the court in *Posado* did not specifically incorporate the reliability factor into its assessment of probative value. Rather, it treated various procedural safeguards—the prosecutor's participation in the examination and the offering of the evidence initially in the pre-trial setting—as contributing to probative value. However, if these safeguards are seen as contributing to the accuracy of exam results, as the court suggested, they contribute to probative value by enhancing reliability. The court in *Posado* may not have fully appreciated this relationship. The need for evidence may certainly heighten its value; but it does not obviate balancing the unfair prejudice associated with such evidence. In the case of expert testimony, reliability must still be sufficient to offset policy counterweights under 403, even in the face of heightened need.¹³⁹

In *United States v. Cordoba*, the trial court ruled the polygraph evidence inadmissible under 702, because it contained defects under industry standards. In that case, the duration and substance of the pre-test was not preserved, no tape or video was made of the pre-test interview or the polygraph exam, the examiner did not calibrate the

¹³⁸ The court referred to this effect as the traditional objection to polygraph evidence; the Supreme Court in *Scheffer* described it as “the aura of infallibility attending polygraph evidence [that] can lead jurors to abandon their duty to assess credibility and guilt.” 523 U.S. at 314.

¹³⁹ See *United States v. King*, 713 F.2d 627, 631 (11th Cir. 1983) (“[W]hile prosecutorial need alone does not mean probative value outweighs prejudice, the more essential the evidence, the greater its probative value, and the less likely that a trial court should order the evidence excluded.”).

machine at the prison test site, and the questions were improper. Moreover, even though the examiner found deception in the defendant's answer, the examiner scored the answer truthful after the defendant's explanation. Finally, the examiner's report was filled with errors and defects and was completed before the test, and numerous other problems existed.

In doing the 403 analysis in *Cordoba*, the court of appeals pointed to the flawed exam as lacking probative value. The factors that rendered the exam unreliable under 702 also made it less probative under 403. The court also noted that the risks associated with such a flawed exam "greatly outweighed the probative value."¹⁴⁰ *Cordoba* shows completely overlapping reliability and probative value inquiries under 702 and 403.

In *Waters* the court bypassed 702 and "independently" excluded the evidence under 403. It is noteworthy that in that case the proponent produced no evidence of the reliability of the polygraph evidence. Thus, the non-existence of probative value facilitated the trial court's 403 decision to exclude. Noting the concern in *Scheffer* about "excessive weight," the trial court had called the evidence collateral and confusing.

In polygraph cases, the "aura of infallibility" that would cause the jury to overvalue the evidence sets a high bar for establishing sufficient reliability. As Professors Giannelli and Imwinkelried point out in their treatise on scientific evidence, "[t]he validity of polygraph testing in criminal investigations remains controversial."¹⁴¹

Yet, polygraph research is ongoing. The director of the Defense Polygraph Institute noted in 1995, "the period between 1986 and the present has been one of unparalleled advances in the psychophysiological detection of deception testing procedures and processes."¹⁴² Validation studies conducted by the now-defunct Office of Technology Assessment of the United States Congress, the Department of Defense, and others show accuracy rates in criminal investigations ranging from 63 to 96 percent, depending upon whether one is confirming truthful answers (higher), or deceptive answers (lower), and whether interpretations were blind or by the original examiner.¹⁴³

These observations suggest that under a specialized 403 inquiry

¹⁴⁰ *Cordoba*, 194 F.3d at 1063.

¹⁴¹ 1 PAUL C. GIANNELLI & EDWARD J. IMWINKELRIED, SCIENTIFIC EVIDENCE, § 8-3 (3d ed. 1999).

¹⁴² 1 *Id.*

¹⁴³ See 1 *id.* § 8-3(C), at 381-82.

based on a reliability showing of probative value, the admissibility of polygraph evidence may well become routine as its accuracy rates continue to rise. With unfair prejudice as high as that associated with the polygraph, reliability may be nothing less than a call for the best evidence.

THAYER REDUX

The 403 approach to determining reliability is certainly not perfect. In fact, Eleanor Swift, harkening back to Thayer, criticized excessive discretion in the area of expert testimony as leading to the loss of consistency and predictability in making reliability decisions.¹⁴⁴ She points, critically, to all the flexibility and potential for inconsistency under *Daubert* and *Kumho*, combined with *Joiner's* “abuse of discretion” standard.¹⁴⁵

She also expresses the concern that judges may inappropriately define substantive law by controlling the use of essential scientific testimony. For example, judges might create a bright-line requirement that without published epidemiological studies which confirm a causal link between the alleged toxic substance and the plaintiff's medical condition, expert testimony that such a link exists will be excluded.¹⁴⁶ One could certainly argue that appellate courts or legislatures, rather than trial judges insulated from judicial review by an “abuse of discretion” standard, should be adding this kind of substantive term. Swift argues that an “abuse of discretion” standard should not hamper efforts to deal with the policy question of how to reconcile the tension between the right to jury trial and the need to use expertise at trials.¹⁴⁷

However, Professor Swift acknowledges the arguments of commentators that in some ways judicial discretion may be more appropriate than bright-line rules in dealing with the dynamics of the trial. Many questions coming up in litigation require “individualized, flexible decision-making” within the context of a particular case.¹⁴⁸ Where cases involve narrow facts that resist generalization, trying to apply rigid rules would promote error and injustice.¹⁴⁹ When it is important to take the trial context into account, trial courts are in a

¹⁴⁴ See Swift, *supra* note 114, at 2467.

¹⁴⁵ *Id.* at 2472-73.

¹⁴⁶ *Id.* at 2473.

¹⁴⁷ *Id.* at 2446

¹⁴⁸ *Id.*

¹⁴⁹ *Id.*

superior position to appellate courts and to legislatures, and rules allowing the trial judge to exercise sensitivity to the complexity and uniqueness of a particular case necessarily promote more truth than mechanical rules.¹⁵⁰

It may be true that standards governing the admissibility of expert testimony may not be appropriate for the exercise of discretion, since these admissibility decisions transcend the individual case. Yet, currently, that is not an open policy question. The reliability standard embodied in 702 as amended and interpreted by the court is a given. Recognizing its provenance in 403 does at least two things. First, it leads to the recognition that even though the standard is variable, there is a threshold of reliability in every case. Second, it reveals that reliability under 403 may be established in some cases without reference to better evidence. Conversely, in some cases as demonstrated by the polygraph cases, it may well mean not just better evidence, but the best evidence.

The beauty of understanding reliability as a specialized application of 403 is that it establishes a connection with basic evidence principles as we move into this area of increasing complexity. It intertwines reliability with probative value and focuses on admissibility as distinctive from sufficiency.

CONCLUSION

Rule 403 is a governor that requires more than minimal relevancy when evidence presents risks to accuracy in factfinding or judicial efficiency. However, 403 does not exclude evidence whose probative value is high enough to offset countervailing risks.¹⁵¹ Since reliability contributes to the probative value of specialized knowledge, even highly risky expert testimony such as polygraph results will be admitted under 403 upon a showing of sufficiently high reliability. Conversely, if expert testimony presents little risk, probative value need not be as high in order to scale the 403 hurdle. Indeed, expert testimony may be “shaky but admissible.”¹⁵²

Because expert testimony carries inherent risks, Rule 702 requires sufficient reliability—substantially more than minimal

¹⁵⁰ Swift, *supra* note 114, at 2444, 2446; *see also* Kevin C. McMunigal & Calvin William Sharpe, *Reforming Extrinsic Impeachment*, 33 CONN. L. REV. 363, 380 (2001) (discussing the preference for a discretionary approach to extrinsic impeachment rather than a bright line rule based partially on relative institutional competence).

¹⁵¹ *See* Blakely, *supra* note 49, at 317.

¹⁵² *See Daubert*, 509 U.S. at 595.

relevancy—for admissibility.¹⁵³ How much more depends upon the countervailing level of risk to be offset. This is why the reliability standard is so often articulated in variable terms—as perhaps a sliding scale without markers.¹⁵⁴ Viewed as a specialized application of 403, the reliability determination under 702 is broad enough to encompass Professor Nance’s better evidence concerns, while accounting for a myriad of risks, some, perhaps, so substantial as to call for the best evidence.

¹⁵³ The “any tendency” standard of Rule 401 defines minimal relevancy. *See* FED. R. EVID. 401.

¹⁵⁴ *See* Nance *supra* note 39, at 221 (referencing to “an ascending scale of reliability with a mark that separates the insufficiently reliable from the sufficiently reliable”).