

Reliability and the Admissibility of Experts

*Dale A. Nance**

- I. INTRODUCTION
- II. THE “RELIABILITY” REQUIREMENT: A CONCEPTUAL CRITIQUE
 - The Core Idea of Reliability
 - Alternative Forms of the Reliability Requirement
 - Reliability as Non-Binary
 - Dichotomy by Deference
 - Dichotomy without Deference
 - The Downside of Dichotomy
 - Reliability and Related Gradational Attributes
- III. TOWARD A POLICY-BASED BALANCING FRAMEWORK
 - Avoiding Blind Alleys
 - Connecting Reliability to Counterweights
 - Offsetting Jury Credulity
 - Monitoring Adversarial Presentations
 - Marshaling Cognitive Resources
 - Securing Better Evidence
- IV. CONCLUSION

I. INTRODUCTION

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹ and *Kumho Tire Co. v. Carmichael*,² the United States Supreme Court set the law of expert testimony on a quest for “reliability.” These decisions, as well as the subsequent amendment of Federal Rule of Evidence 702, make it clear that trial judges are to perform a “gatekeeping” function, filtering out offered testimony when the expertise on which it is

* Professor of Law, Case Western Reserve University. For comments on the ideas presented here, I wish to thank all those who participated in a faculty workshop at Case Western Reserve University Law School as well as those who participated in the conference for which this paper was prepared, particularly David Faigman, Edward Imwinkelried, and Calvin Sharpe. For further helpful conversations, I am indebted to Jessica Berg, Paul Giannelli, Susan Haack, Sharona Hoffman, Max Mehlman, and Kevin McMunigal.

¹ 509 U.S. 579 (1993).

² 526 U.S. 137 (1999).

based, whether scientific or otherwise, is not reliable. Rule 702 now provides:

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.³

Conditions (1)-(3) were added in 2000 in response to the teachings of *Daubert* and *Kumho Tire*.

These developments have spawned a substantial literature.⁴ Little attention has been given, however, to providing a careful analysis of the contours and functions of the reliability concept itself and its relationship to the purposes of admissibility rules.⁵ *Daubert* and *Kumho Tire* tell us what kinds of factors to examine in order to determine whether or not proffered expertise is reliable—factors such as whether or not the theory has been tested empirically, and whether or not it has been subjected to the rigors of peer review and publication⁶—and the 2000 amendment requires trial courts to examine three different respects as to which the question of reliability can be assessed. But these authorities say very little about what is meant by reliability or exactly why an interest in reliability should be manifested in the decision to exclude proffered evidence. In *Daubert*, for example, the majority opinion states only that evidentiary reliability is equivalent to “trustworthiness,” something presumably to be demanded, and states this only in a footnote.⁷

Without a formula for the synthesis and use of the identified factors, which the Supreme Court quite understandably eschews, one needs to reflect on how the law of admissibility structures trials in order to synthesize all the pertinent considerations into an overall reliability assessment that serves the law’s goals. This Essay

³ FED. R. EVID. 702.

⁴ See generally DAVID L. FAIGMAN ET AL., SCIENCE IN THE LAW: STANDARDS, STATISTICS AND RESEARCH ISSUES 1-65 (2002).

⁵ Some sensible but brief comments are made in this regard by the current editors of the McCormick treatise. See 1 MCCORMICK ON EVIDENCE § 203 (John W. Strong ed., 5th ed. 1999). An unusually insightful analysis is given in the context of a recent proposal that English law adopt a reliability-based exclusionary rule for criminal trials. See MIKE REDMAYNE, EXPERT EVIDENCE AND CRIMINAL JUSTICE 116-39 (2001).

⁶ *Daubert*, 509 U.S. at 592-94; *Kumho Tire*, 526 U.S. at 149-52.

⁷ *Daubert*, 509 U.S. at 590 n.9.

undertakes such a conceptual examination and explores at least some of its implications for assessing the admissibility of expert testimony.

In Part II, I examine the most plausible understandings of the reliability requirement, concluding that the best understanding is rather different from the way it is commonly and even authoritatively expressed.⁸ I argue that courts and commentators should disavow a binary, all-or-nothing concept of reliability—that evidence is either reliable or unreliable—in favor of a gradational concept—that the reliability of evidence is a matter of degree. The apparently binary mode of expression appearing in Rule 702 is misleadingly conclusory; it provides no analytical framework for deciding how reliable is reliable enough.

In Part III, I develop some of the implications of that conclusion, suggesting a framework for the courts to use in assessing reliability for the purpose of an admissibility ruling. Not surprisingly, this involves weighing the degree of reliability of the proffered expertise against the dangers that might arise from admitting the expertise. But once this framework is identified, important conclusions emerge. Primarily, I argue that concerns about jury misuse of expertise are less important than concerns about controlling advocates so as to provide the trier of fact, judge or jury, with expertise that is as reliable as the circumstances of the litigation permit. I suggest one set of ways to take this into account in a workable interpretation of the reliability requirement in Rule 702.

To situate the present discussion, I should make an obvious but important disclaimer: I do not assert that achieving a satisfactory or optimal interpretation of the reliability requirement of Rule 702 is the most pressing issue affecting the use of expert testimony. Various reform measures—including greater use of court-appointed experts—may well be more important to the administration of justice than the interpretive issue discussed here.⁹ Implementation of such reforms might in turn affect the analysis of Rule 702, but no attempt is made here to account for such complications.

II. THE “RELIABILITY” REQUIREMENT: A CONCEPTUAL CRITIQUE

The Core Idea of Reliability. As suggested by the Supreme Court’s

⁸ In view of the 2000 amendment to Rule 702, I do not address here the question of whether the Supreme Court’s opinion in *Daubert* developed the best interpretation of that rule as it then read.

⁹ See, e.g., JOSEPH SANDERS, *BENEDICTIN ON TRIAL: A STUDY OF MASS TORT LITIGATION* 193-211 (1998) (emphasizing the importance of reforms other than strict application of admissibility standards).

cursory reference to “trustworthiness,” the concept of reliability entails a justifiable willingness to rest decision, in some manner, on the thing considered reliable. Reliability is inherently relative to a particular decision context, and thus relative to the goal or purpose of decision. In the context of the adjudicative legal norms announced in *Daubert*, *Kumho Tire*, and revised Rule 702, the decisions in question are verdicts, and the principal goal in mind is the accuracy of those verdicts, within reasonable constraints of time and cost.¹⁰ The core idea, then, is a concern for the *veritistic* properties of offered expertise and, perforce, the rules that control the admissibility of such expertise. By “veritistic” I mean the *truth producing quality* of the expertise or the regulative rules, truth referring to the factual accuracy of verdicts.¹¹ Evidence, or the rules regulating evidence, may be said to be more or less veritistic, more or less conducive to accurate verdicts.

The requirement of reliability purports to be a veritistic regulation by imposing some kind of supra-relevance veritistic constraint on expertise that is introduced in court. These two ideas are not necessarily conjoined. One could hold the view that the veritistically optimal regulatory scheme is for courts not to concern themselves directly with the issue of reliability at the admissibility decision. Sometimes, the best way to reach a goal is for a particular decision-maker not to focus directly on the goal itself.

For example, if the conditions of Coase’s Theorem hold, then a court’s decision on the content of the substantive law will not affect the efficiency of the rule at issue, so long as the court provides a clear rule with reference to which parties are allowed to bargain; if the court did not strike upon it, the parties will bargain to the efficient solution.¹² Under these conditions, assuming that wealth

¹⁰ I consider later the possibility that accuracy of verdicts is not the goal relative to which reliability is to be assessed. See *infra* notes 54-56 and accompanying text.

¹¹ The indicated conduciveness may be understood in the sense of justification or it may be understood in causal terms. If certain evidence justifies an inference as accurate, then the use of that evidence is causally conducive to accurate inference, unless the decision making is irrational. Since the concern for irrational decision making (in the form of jury error) is a recurring one in the law of evidence, the causal focus is more general and will be employed here. In this I follow the work of epistemologist Alvin Goldman. See ALVIN I. GOLDMAN, *KNOWLEDGE IN A SOCIAL WORLD* (1999). Although the implications of such a veritistic social epistemology are controversial, the general approach is compatible with traditional academic conceptions of the goals of evidence law. Compare Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491 (2001), with Dale A. Nance, *Naturalized Epistemology and the Critique of Evidence Theory*, 87 VA. L. REV. 1551 (2001).

¹² Assuming that parties attempt to maximize wealth, the principal condition that

maximization is the principal goal that the court wants to achieve, the court should focus not on choosing the allocation of entitlements that is more or most efficient, but rather on choosing an allocation that is clear and, *ceteris paribus*, one that is just or fair on non-efficiency grounds. A court's attempt to go after efficiency directly could well backfire if its ruling, trying to capture the elusive and changing facts that conduce to efficiency, succeeds only in creating a rule that is unclear, thus complicating the bargaining process.

Similarly, a justificatory premise of the adversary system is that the clash of opposing, relevant evidence will yield accurate results, at least frequently enough to render that system superior to the alternatives.¹³ To the extent that the system is veritistically successful, it is due to confluence of two factors. First, the system encourages parties to present all relevant evidence that is reasonably available and not too weak to be of practical use, each side in the dispute having an incentive to present that which is significant and favorable.¹⁴ Second, the system provides a trier of fact capable of shouldering the responsibility of determining what inferences from the evidence are warranted. Accepting these premises, courts need not ordinarily concern themselves with the details of reliability or trustworthiness in deciding whether or not to admit evidence; they need only exclude that which is irrelevant.¹⁵

Of course, many admissibility rules are based on perceived exceptions to this general principle of adversarial veritism. In particular, both the Supreme Court and Congress have decided that the proffer of expertise should activate a decision process that looks beyond mere relevance, and even beyond expert qualifications, to examine the reliability of the expert testimony itself. But what exactly

must be satisfied is that the transaction costs of rearranging entitlements must not exceed the gains from trade. See R.H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 10 (1960).

¹³ The *locus classicus* for this argument is Lon L. Fuller, *The Adversary System*, in TALKS ON AMERICAN LAW 30 (Harold J. Berman ed., 1961). The indicated premise need not be that adversarial presentation of evidence is *more* accurate than alternatives, only that it is not so much *less* accurate as to make it an inferior system all things considered.

¹⁴ The importance of having all relevant evidence is that it constitutes a well-known condition on the validity of inductive inference, often called the principle of total evidence. See, e.g., GOLDMAN, *supra* note 11, describing his non-categorical version as the "truth-in-evidence principle," articulated as follows: "A larger body of evidence is generally a better indicator of the truth-value of a hypothesis than a smaller, contained body of evidence, as long as all the evidence propositions are true and what they indicate is correctly interpreted." *Id.* at 145.

¹⁵ Accordingly, the default rules are that irrelevant evidence is inadmissible and that relevant evidence is admissible. FED. R. EVID. 402.

does that mean? And how can this be implemented intelligently within an adversarial framework?

Alternative Forms of the Reliability Requirement. Authoritative expressions of the reliability requirement, both judicial and statutory, generally express the requirement in a dichotomous or binary fashion. Proffered expertise is considered either reliable or unreliable; reliability, it is implicitly assumed, does not come in degrees.¹⁶ In *Daubert* the Court stated, “[U]nder the [Federal] Rules [of Evidence] the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.”¹⁷ Extending this basic gatekeeping function to all specialized knowledge within the scope of Rule 702, the majority in *Kumho Tire* opined, “The trial judge’s effort to assure that the specialized testimony is reliable and relevant can help the jury evaluate [the expert’s] foreign experience, whether the testimony reflects scientific, technical, or other specialized knowledge.”¹⁸ Not coincidentally, in each opinion the reliability requirement is specifically coupled with the requirement of relevance, which—under prevailing legal conceptions—is clearly dichotomous: evidence is either relevant or irrelevant; it is not somewhat relevant, very relevant, etc.¹⁹ Similarly, amended Rule 702 states that, to be admitted, proffered expert testimony must be “the product of reliable principles and methods,” applied “reliably to the facts of the case.” We are thus encouraged to believe that testimony either is or is not the product of reliable methods reliably applied.

The alternative way of expressing the requirement would be to say that there must be a showing of “sufficient reliability” for admission. This locution implies a concept of reliability that is gradational, a matter of degree. It is suggested by references in *Daubert* to “the common law insistence upon ‘the *most reliable* sources

¹⁶ See 1 MCCORMICK ON EVIDENCE, *supra* note 5, § 203, at 727.

¹⁷ 509 U.S. at 589.

¹⁸ 526 U.S. at 147-49 (quoted language appearing at 149).

¹⁹ This follows from the definition of relevance that is now widely accepted and incorporated into the federal rules: “‘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. Under the linguistic conventions of the federal rules, expressions like “somewhat relevant” and “highly relevant” are simply imprecise ways of expressing something about the probative value of the evidence, a notion that is clearly gradational. Probative value can be low or high, very low or very high, etc. See Richard D. Friedman, *A Close Look at Probative Value*, 66 B.U. L. REV. 733 (1986) (discussing formal measures of probative value).

of information’,²⁰ a locution that presupposes reliability can be “more” or “less” as well as “most.” It may also surface with the reference in Rule 702(1) to “sufficient facts and data.” And it clearly emerges in the drafters comment on the revised rule that numerous factors are relevant “in determining whether expert testimony is *sufficiently reliable* to be considered by the trier of fact.”²¹ I will argue that this conception of reliability is clearly superior, relative to the task we ask courts to perform. This in turn accounts for the tendency of some commentators, in restating the requirement articulated in *Daubert*, *Kumho Tire*, and revised Rule 702, to slip into the “sufficient reliability” formulation.²²

The problem, of course, with expressing the requirement in the latter way, a problem that may explain why that locution is often avoided, is that, without some reasonably determinate algorithm based on appropriate legal norms that would specify what degree of reliability is “sufficient” for this purpose, the requirement of sufficient reliability is vacuous. It tells us almost nothing.²³ Using the dichotomous or binary form of expression might conceal this gap by suggesting either that reliability is a straightforwardly binary factual question, akin to the preliminary question of whether a document offered in court is the original thereof,²⁴ or that there is some criterion of reliability that can be applied without directly invoking such an algorithm.²⁵ If such suggestions are false, then what is needed is a specification of not only the considerations that weigh against admission of (assumedly) relevant expertise, but also some indication of how to determine at what level of reliability these counterweights ought to be controlling. The latter specification can be purposive (laying out the purpose to be achieved and allowing the trial courts to pursue the goal) or formulaic (providing a formula that can be employed without reference to the purpose), or some of

²⁰ 509 U.S. at 591 n.9, 592 (quoting FED. R. EVID. 602 advisory committee’s note) (emphasis added).

²¹ FED. R. EVID. 702 advisory committee’s note to 2000 amendment (emphasis added).

²² See, e.g., Robert Epstein, *Fingerprints Meet Daubert: The Myth of Fingerprint “Science” is Revealed*, 75 S. CAL. L. REV. 605, 617, 621 (2002).

²³ Of course, the mere fact that a requirement of “sufficient reliability” is posited by the Court and Congress tells the trial courts that some decision about reliability should be made, as opposed to none. This message, however, gives no guidance about how to make the reliability determination.

²⁴ See FED. R. EVID. 1001, 1002.

²⁵ By using the term “algorithm,” I do not intend to connote quantification or computational precision, merely an intelligibly complete specification of an analytical framework or procedure, something more informative than simply an injunction to “solve the problem.”

each. Despite the principal terminology used in *Daubert*, *Kumho Tire*, and Rule 702, the new standard can be interpreted as attempting, however confusingly and incompletely, such a hybrid instruction.

In fact, we often acknowledge that dichotomous antecedents in legal rules really reflect such weighing algorithms. Take the cause of action for negligence. Expressed in dichotomous fashion, we have: if defendant was negligent (and the other elements of the cause of action are present), compensation should be paid; if defendant was not negligent, compensation should not be paid (under a negligence theory). We all understand, do we not, that “negligence” is a *conclusion* derived from some sort of normatively informed risk algorithm, not an observable or discernible fact.²⁶ In the rule, “If defendant was negligent, compensation should be paid,” the term “negligent” is simply a place-holder for such an algorithm; it carries no independent force, except to alert us to the fact that the referenced algorithm relates to the question of what degree of care is socially demanded.

The dichotomous form of the reliability requirement could be meant, or at least it could operate, in this way. If so, it is simply a less transparently vacuous form of the “sufficient reliability” requirement.²⁷ We accept that reliability comes in degrees, and are told implicitly that there are reasons to exclude evidence even though it has more than *de minimis* reliability. The requirement of reliability then says little more than this: identify those reasons and decide whether they outweigh the law’s default principle of admitting any evidence possessing probative value. In such a scheme, the reliability requirement of Rule 702 is either redundant of the balancing test of Rule 403,²⁸ or it rather cryptically imposes restrictions at least in some ways more demanding than Rule 403, restrictions that need to be elucidated.

Reliability as Non-Binary. The Court in *Daubert* identified reliability, in the context of scientific evidence, with scientific

²⁶ The algorithm might be a cost-benefit calculation *a la* Learned Hand, an elaboration of a categorical imperative *a la* Kant, or yet some other idea.

²⁷ Even before *Daubert* was decided, Professor Moenssens asserted the importance of reliability assessments, using dichotomous reliability terminology, but then felt impelled to recognize that such a concept of reliability is a conclusion, a “value judgment,” derived from the weighing of legal norms. See Andre Moenssens, *Admissibility of Scientific Evidence—An Alternative to the Frye Rule*, 25 WM. & MARY L. REV. 545, 564-67 (1984).

²⁸ FED. R. EVID. 403: “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.”

validity.²⁹ The nature of the problem is thus illuminated by a digression relating to scientific validity. A quick review of some of the *Daubert* factors demonstrates why scientific validity is not easily understood in binary terms, except once again as the expression of a conclusion relative to some particular task.

Consider the first “factor” adduced for the assessment of the reliability of a purportedly scientific theory or technique: “whether it can be (and has been) tested.”³⁰ How much testing does it take before a principle or technique can be said to have been “tested” for this purpose? Good scientists know that replication of results is an important part of the scientific method. Can, then, a principle or technique be considered “tested” if only one test has ever been conducted? Two? Three? Does it not depend on the quality of the tests as well as their quantity? If so, this points to a concept of testing that is a matter of degree, more testing being better than less, *ceteris paribus*, and testing that is more closely related to the proposition of interest being better than testing that is more distant.³¹

Similarly, if subjection to the demands of peer review and publication is an important factor, as the Court opines,³² it should matter how rigorous the peer review of a particular publication is, and that is surely a matter of degree as to which not all publications are alike.³³ Relatedly, although the degree of acceptance of a principle or technique within a particular community of experts is certainly pertinent to an assessment of the degree of its reliability, little should turn on the artificial question of whether that degree of acceptance has passed from just below to just over 50% of the relevant community. Better is for courts to appreciate the *degree* of the support in the relevant community in deciding what to do with the proffer under some standard that does not make general

²⁹ 509 U.S. at 590 n.9 (“In a case involving scientific evidence, *evidentiary reliability* will be based upon *scientific validity*”); *id.* at 592-93 (stating that the trial judge’s reliability determination “entails a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue”).

³⁰ *Id.* at 593.

³¹ See FAIGMAN ET AL., *supra* note 4, § 1-3.4.2[2], at 35 (“Only through replications, using various designs and methods, do scientists gain confidence that a hypothesis has been sufficiently corroborated. No magic number or moment determines this point, however; like many areas of the law, science presents a broad spectrum of grays that over time sharpens into black and white.”).

³² 509 U.S. at 593-94.

³³ See FAIGMAN ET AL., *supra* note 4, § 1-3.4.3 (noting variance in the status of journals and the quality of peer reviews).

acceptance a touchstone.³⁴

Then there is the matter of “error rates.”³⁵ How precisely must we know the error rate of a particular technique? And what does one make of a technique with a known, very high error rate? The rate may be known with some precision, but that does not mean the technique gets a “yes” answer to this component of the reliability inquiry. Rather, the question must be transformed into one that inquires how large the known error rate can be before the technique must be rejected by the courts as unhelpful. That is a “How high is too high?” question, not a “Yes or no?” question.³⁶ The same is true for the maintenance of standards controlling a techniques operation, considered as an indicium of reliability.³⁷

These quick observations could be multiplied easily. They point to a fact that scientists, at least sophisticated ones, surely understand: scientific validity is not an all-or-nothing characteristic; rather, it is a matter of degree.³⁸ This proposition poses the question of how one gets from a consideration of the degree of scientific validity to the undeniably binary decision of whether or not to admit proffered testimony. Of course, the same question is posed by the notion of validity as a synonym or proxy for the reliability of non-scientific expertise. Thus, *the fundamental problem is how to map from a*

³⁴ Indeed, the *Daubert* opinion does not state starkly that majority approval is significant in and of itself, even as a “factor”; rather, quoting an earlier Third Circuit opinion, it states,

A “reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.” Widespread acceptance can be an important factor in ruling particular evidence admissible, and “a known technique which has been able to attract only minimal support within the community,” may properly be viewed with skepticism.

509 U.S. at 594 (internal citations omitted).

³⁵ “[I]n the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error” as a pertinent factor. *Id.*

³⁶ See FAIGMAN ET AL., *supra* note 4, § 1-3.4.2[1], at 35 (noting that “any science that is even slightly better than flipping a coin” is likely to meet the test of “assisting the trier of fact,” so that Rule 702 “must embody, at least in part, a Rule 403 analysis”).

³⁷ *Daubert*, 509 U.S. at 594.

³⁸ FAIGMAN ET AL., *supra* note 4, § 1-3.4, at 25 (“[A]s every scientist knows, validity is not a categorical conclusion.”); KENNETH R. FOSTER & PETER W. HUBER, *JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS* 17 (1999) (“‘Validity’ in science is not a binary attribute, like pregnancy.”); Erica Beecher-Monas, *A Ray of Light for Judges Blinded by Science: Triers of Science and Intellectual Due Process*, 33 GA. L. REV. 1047, 1062 (1999) (“The Supreme Court . . . failed to acknowledge that . . . scientific reliability is not an all-or-nothing proposition, but rather depends on the application of the evidence and the acceptable risk of error.”).

gradational epistemic conception of reliability to a dichotomous legal choice on admissibility. The following sections consider the options available for solving this problem.

Dichotomy by Deference. Perhaps the Supreme Court intended a dichotomous reliability requirement that can be determined by reference to more or less factual propositions about the standards of non-legal institutions or their proper application, without the aid of a legal-policy informed algorithm of the type described above. Indeed, there is much language in *Daubert* that suggests this. As already noted, the Court's opinion identifies reliability, in the context of scientific evidence, with scientific validity. That would appear to be a more or less factual (at least, non-legal) inquiry into the application of standards of scientific inquiry specified by the scientific community.³⁹ The well-known factors identified in the Court's opinion seem to have a similarly factual and deferential character: Has the theory or principle been tested? Does it have a known error rate? Has it been published and subjected to peer review? Has the principle been applied according to standards developed by the scientific community? Does it enjoy general acceptance among the relevant scientific community?⁴⁰

To be sure, the Court made a point of refusing to provide a means of determining how many "yes" answers to the five questions posed above are required before reliability is established. And if, as I have argued above, it makes little sense to talk in terms of looking for "yes" or "no" answers to the questions posed by the Court's enumerated factors, the problem of synthesis is all the more difficult. This, however, is actually consistent with the idea that the Court is deferring to the norms of the scientific community: The Court has no particular skill for answering the synthetic question *ex ante*; rather, it must be answered by investigation, more or less the way a court in state X decides a question of law governed by the law of state Y. In this case, however, state Y is the State of Science, a state with no recognized supreme court. Moreover, because the Court acknowledged that yet other questions might need to be answered, that other factors might need to be considered, no fixed formula can be specified. The Court understandably concedes the imperfections in its knowledge of, and thus its ability to articulate *ex ante*, the validity

³⁹ See Michael J. Saks, *The Aftermath of Daubert: An Evolving Jurisprudence of Expert Evidence*, 40 JURIMETRICS J. 229, 231 (2000) ("The move from *Frye* to *Daubert* increases judges' gatekeeping duty by requiring them to evaluate claims of scientific expertise much as scientists would.").

⁴⁰ See 509 U.S. at 593-94.

laws of the State of Science.

Those familiar with admissibility standards in the mold of *Frye v. United States*⁴¹ may quibble with my use of the term “deferential” to describe this interpretation of *Daubert*’s reliability inquiry. *Daubert*, after all, rejected *Frye* as a matter of statutory interpretation, and *Frye* represented deference to the scientific community.⁴² From a broader perspective, however, *Daubert* might be understood as endorsing simply a different kind of deference. *Frye* mandated deference to the collective opinion of scientists in the relevant field concerning the validity of the proffered expertise. *Daubert* might be understood as mandating instead deference to the norms of science, norms that must be applied directly by trial and appellate judges, without accepting the collective opinion of scientists as a conclusive proxy for validity.⁴³ This interpretation of *Daubert* cannot be rejected simply because *Daubert* rejected *Frye*-type deference.

Still, it is hard to believe that this non-*Frye* form of deference is what the Court really intended, or would intend if it had clearly thought the matter through. The obvious problem, one frequently noted, is that judges must become junior scientists in order to try to apply the normative standards of the scientific community. This is a task that some judges are probably incapable of performing adequately, and one certainly that some judges *believe* they are incapable of performing adequately.⁴⁴ It is complicated by disagreement among scientists and philosophers of science regarding the norms of scientific disciplines, disagreement of which there is barely a hint in the Court’s opinions.⁴⁵ This difficulty, however real in some cases, may be somewhat exaggerated. In many contexts of ordinary science, it surely is possible for judges, with some assistance, to make reasonable calls that would parallel those of good scientists who have not lost their objectivity.⁴⁶

⁴¹ 293 F. 1013 (D.C. Cir. 1923) (applying a “general acceptance in the particular field” test for novel scientific evidence).

⁴² See *Daubert*, 509 U.S. at 585-89.

⁴³ See FAIGMAN ET AL., *supra* note 4, § 1-3.0, at 13 (comparing *Frye* and *Daubert*).

⁴⁴ See, e.g., *Daubert*, 509 U.S. 579, at 598-601 (Rehnquist, C.J., dissenting in part) (casting doubt on the ability of judges to perform the contemplated task); *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1315-16 (9th Cir. 1995) (Kozinski, J., on remand, bemoaning the Supreme Court’s “Brave New World” in which judges must decide whether a qualified scientist’s testimony is “good science” “derived by the scientific method.”).

⁴⁵ See generally David S. Caudill & Richard E. Redding, *Junk Philosophy of Science?: The Paradox of Expertise and Interdisciplinarity in Federal Courts*, 57 WASH. & LEE L. REV. 685 (2000).

⁴⁶ See DAVID L. FAIGMAN, LEGAL ALCHEMY: THE USE AND MISUSE OF SCIENCE IN THE

The more fundamental question is this: Why should the final determination, reliable or not *in court*, be determined by the norms of the scientific community instead of those of the legal community (other than the legal norm of deference itself)? If the scientific community recognizes scientific validity as a gradational, non-binary concept, as I think it does and must, then that community must employ one or more normatively informed, more or less determinate algorithms of its own for making any categorical determinations of validity that it is called upon to make. These norms will have developed to serve the interests of science or, more precisely, the scientific community.⁴⁷ And if the scientific community sometimes thinks of scientific validity as an all-or-nothing concept, it is only because that community has developed dichotomous rules of thumb that, while over- and under-inclusive in some cases, roughly serve to further those interests. Why then should the scientific community's balancing of those interests, whether consciously and directly or indirectly by the application of rules of thumb, be taken as controlling the conduct of litigation?

Three initially plausible reasons can be identified, though each must be rejected in the end.⁴⁸ First, it might be thought that the purposes and goals of the scientific community are the same as those of the law of adjudication, or at least so nearly the same that the difference is not worth much concern. This supposed reason, however, is not likely to hold true in the scientific context. To be sure, both disciplines place a high priority on the ascertainment of truth by the application of relatively formal procedures. But the parameters that give shape to this quest differ noticeably: those of good adjudication include, for example, an emphasis on promptness and finality of decision that is incompatible with the goals of science, as Justice Blackmun himself observed pointedly in *Daubert*.⁴⁹

LAW 64 (1999) (“[M]ost judges are intelligent and well educated and there is no reason why they cannot with diligence master the basics of the scientific method and statistical theory.”).

⁴⁷ I do not mean to imply that these interests are narrow or selfish, just that they are accepted within the scientific community. On the other hand, that does not preclude the possibility of significant rent-seeking, including the shaping of norms so as to augment public funding of scientific research.

⁴⁸ A fourth possible reason actually has considerable merit. It is based on the *assertion* by the proponent or its expert that the expertise is scientifically sound. This certainly provides a reason to defer to scientific standards, at least as part of the analysis. This reason, however, which depends on the precise claim being made by the proponent, is readily accommodated within a framework that employs a gradational reliability concept and a balancing test. See *infra* notes 185-93 and accompanying text.

⁴⁹ 509 U.S. at 596-97.

For example, the law's concession to unavoidable doubt often entails an aversion to false negatives (e.g., refusing to impose civil liability on defendant for want of proof of causation when its product did in fact cause plaintiff's injury) that is almost as strong as its aversion to false positives (e.g., imposing liability when defendant's product did not cause the plaintiff's injury).⁵⁰ This trade-off is not reflected in the typical scientific standards of hypothesis testing, themselves important norms of scientific validity.⁵¹ Thus, the conservative convention in social science that (generally) insists on results that possess a statistical "p value" of less than 0.05 before a difference among one or more comparison groups will be deemed "statistically significant" is an inferential norm that gives priority to the avoidance of false positive conclusions. The idea is that, with a risk of false positive conclusions greater than 5% were the "null hypothesis" of "no difference" true, that hypothesis cannot safely be rejected. In that case, scientists are prepared to leave the matter in limbo pending further study. The law of trials, however, cannot so indefinitely postpone making a judgment, but must rely on whatever evidence is reasonably available at the time of decision.⁵² Even in criminal cases, this difference is significant, most conspicuously in assessing evidence offered by the accused.

These observations reflect the fact that the argument from identity of goals is profoundly incomplete without an explanation of how legal *admissibility* norms, which concern the *helpfulness* of information in reaching an accurate and final verdict on *particular* events in accordance with the applicable burden of persuasion in an adversarial system, map to scientific *validity* norms, which relate to the permissibility of scientists deriving revisable *conclusions* about the

⁵⁰ Almost, but not quite. See Dale A. Nance, *Civility and the Burden of Proof*, 17 HARV. J.L. & PUB. POL'Y 647, 659-72 (1994) (arguing that a false positive verdict is more serious than a false negative one whenever the former entails an imputation of serious wrongdoing by the defendant but the latter does not entail an imputation of serious wrongdoing by the plaintiff).

⁵¹ See, e.g., Ronald J. Allen, *Expertise and the Daubert Decision*, 84 J. CRIM. L. & CRIMINOLOGY 1157, 1164-67 (1994) (discussing this difference in the context of the epidemiological evidence in *Daubert*); Margaret A. Berger, *Upsetting the Balance Between Adverse Interests: The Impact of the Supreme Court's Trilogy on Expert Testimony in Toxic Tort Litigation*, 64 LAW & CONTEMP. PROBS. 289, 300-02 (2001) (emphasizing differences between law's assessment of causation and scientific assessments of causation); Beecher-Monas, *supra* note 38, at 1099-1102 (same).

⁵² The point is carefully examined in Neil B. Cohen, *The Gatekeeping Role in Civil Litigation and the Abdication of Legal Values in Favor of Scientific Values*, 33 SETON HALL L. REV. 943 (2003); see also David H. Kaye, *Statistical Significance and the Burden of Persuasion*, 46 LAW & CONTEMP. PROBS. 13 (1983) (explaining the complex relationship between statistical significance and the burden of persuasion).

general patterns of causation drawn from data or information. While it is fairly safe to assume that such scientifically well-grounded conclusions are going to be helpful in lawsuits to which they are relevant, the converse cannot be safely assumed. Scientific validity, as understood by scientists, should not be considered necessary in all cases for adjudicative helpfulness.⁵³

The second possible reason to defer in this way is more subtle. Even if the criterion of sufficient reliability for conclusions in science is seriously different from the theoretically appropriate criterion of sufficient reliability for admissibility in adjudication, the law might choose to defer to the scientific criterion simply to avoid controversy, to use a more determinate scientific criterion in place of a more debatable legal criterion. This argumentative candidate has at least a chance of success in the context of scientific evidence, because portions of the scientific community have standards of reliability that are reasonably determinate in some contexts.

Ultimately, however, it is unconvincing. The voluminous post-*Daubert* litigation has shown that the new standards have hardly avoided controversy. Relatedly, because of the significant difference between the goals and purposes of the law and those of science, an inevitable tension will arise in attempting to maintain this kind of deference. Courts will be pulled to develop more pertinent legal standards, even in the face of potential controversy about the content thereof. In any event, the failure of such deference, to the extent based on this consideration, would be well deserved. Avoiding controversy does not really concern the kind of costs that properly should be balanced against our interest in accuracy. Rather, it involves trading off (rather unsuccessfully at that) the goals of the law for the convenience and comfort of judges and the rest of the political establishment.⁵⁴

Perhaps, however, the avoidance of controversy is not really the point. We can recast this into a third argument for deference if we are willing to relax more candidly the assumption that accuracy of verdicts is the primary consideration with respect to which reliability

⁵³ See generally Peter Donnelly & Richard D. Friedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 MICH. L. REV. 931, 971-78 (1999) (noting that the majority opinion in *Daubert* failed to realize the implications of its acknowledgment of the differences between law and science).

⁵⁴ Some kinds of costs obviously must constitute counterweights to the admission of relevant evidence and indeed to the implementation of all rules that are veritistic in design. There are limits to the amount of social resources that should be allocated to improving accuracy of decision. The experience of controversy may be unsettling, but it is a cost we expect judges to endure in the course of adjudicating disputes, at least if it cannot be reduced without sacrificing accuracy.

is required. Suppose instead that *public confidence* in verdicts is the primary consideration.⁵⁵ If so, then a desire to create or maintain an undeserved public confidence that verdicts reflect the certain truth might cause the Court and Congress to insist that seemingly scientific experts testify only when their underlying expertise meets some demanding standards employed in scientific investigation. In a civil case involving a plaintiff's offer of ostensible expertise that does not meet the standards of science, the tortured thinking might run like this: If the jury decides in favor of the plaintiff, and the public becomes aware (as indeed the media will make sure they do) that some of the evidence in the case was not up to scientific standards, then the public might believe that the jury was duped into relying excessively upon that particular piece of expertise (as indeed the defense will publicly assert). Thus, the public will come to doubt the veracity of the jury's verdict, even if that verdict is entirely proper.

This, too, is ultimately unsatisfactory as a justification of deference, even if the embedded behavioral assumptions are true. Within a broad range of normal applications, public confidence in the law, though certainly important, is a value that should remain subordinate (indeed, lexicographically inferior) to accuracy of inference and decision in accordance with the established burden of persuasion. Moreover, in the long run, to sacrifice these goals for the sake of public confidence is likely to endanger the very public confidence that is desired. Playing the public confidence game can backfire once it becomes public that this is the game being played.⁵⁶

Is the situation different for non-scientific expertise, what Rule 702 refers to as "technical or other specialized knowledge"? There are serious, practical disciplines, such as clinical medicine, for which it can be said that the parameters of decision, such as the priority given to accuracy, promptness, and even finality, are more akin to those of adjudication than are the pure sciences. We recognize this in myriad ways, such as the important deference represented by the exception to the hearsay rule contained within Rule 703, allowing expert opinions to be based on facts or data that are not necessarily admissible if they are "of a type reasonably relied upon by experts in

⁵⁵ Cf. Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357 (1985) (arguing that many rules of admissibility and sufficiency are attributable to an explicit or implicit policy of generating verdicts acceptable to the public, whether accurate or not).

⁵⁶ In the same way, the value of impartiality in the decision-maker should carry a higher priority than the value, important as it is, of projecting an *appearance* of impartiality to the public, and one of the best (though not infallible) means for maintaining the appearance of impartiality is to maintain real impartiality.

the particular field in forming opinions of inferences upon the subject.”⁵⁷ Although this provision facially does not distinguish among different types of expertise, the conspicuous illustration usually given to justify it is the physician making “life-and-death decisions” based on hospital records, X-rays, and statements by the patient’s relatives and other medical personnel.⁵⁸

On the other hand, there are countless forms of “specialized knowledge,” ranging from astrology to plumbing, for which this is not true. And some non-scientific disciplines, even some that have been regularly employed in litigated cases for decades, have little or no extra-disciplinary checks on reliability.⁵⁹ As recent debates have made clear, this may be especially true for disciplines, such as handwriting identification, that have been developed for, and find little application other than, forensic uses.⁶⁰ To defer to the normative standards of reliability in such disciplines would be to abdicate the basic gatekeeping function.⁶¹ As a general strategy, therefore, deference to the reliability norms of non-legal institutions is no more plausible in the context of non-scientific expertise than it is with regard to scientific expertise.

The law of admissibility might attempt to tailor this discipline by discipline, deferring only to those non-legal institutional norm systems that generate reliability standards fairly close to those that law’s normative, policy-based prescriptions would yield directly. But why go to this trouble? It would, after all, require the same non-deferential assessment that, under a more generally non-deferential approach, would have to be done directly, except that it would be

⁵⁷ FED. R. EVID. 703.

⁵⁸ FED. R. EVID. 703 advisory committee’s note (distinguishing as impermissible the case of an accidentologist relying on reports of bystanders who observed an accident).

⁵⁹ Even in a context like conventional medicine, there may be good reasons to doubt the reliability of decision making; a nominal concern for accuracy, promptness, and finality that is similar to the law’s may not in fact translate into a degree of reliability that the law should consider appropriate for the resolution of disputes. See, e.g., John M. Eisenberg, *What Does Evidence Mean? Can the Law and Medicine Be Reconciled?*, 26 J. HEALTH POL., POL’Y & L. 369, 369-70 (2001) (“[T]here is sufficient evidence to suggest that most clinicians’ practices do not reflect the principles of evidence-based medicine but rather are based on tradition, their most recent experience, what they learned years ago in medical school, or what they have heard from their friends.”).

⁶⁰ See Michael J. Saks, *Merlin and Solomon: Lessons from the Law’s Formative Encounters with Forensic Identification Science*, 49 HASTINGS L.J. 1069, 1091 (1998).

⁶¹ This, of course, is not to say that the degree to which a proposed expert follows the normative standards of his or her own discipline is irrelevant to a reliability assessment. Significant divergence is a red flag suggesting bias and calls for explanation.

attempted over whole classes of expertise. Moreover, to use a deferential standard for some categories of expertise and a non-deferential standard for others would require a kind of disciplinary line-drawing that was rejected in *Kumho Tire*. Affirming the applicability of the reliability standard to all kinds of expertise, whether or not commonly regarded as “scientific,” the majority rightly opined that, while the factors that might be useful in assessing reliability should vary with context, no distinction between science and non-science can be doctrinally justified or practicably maintained.⁶² The same constraints prevent us from using a deferential, dichotomous reliability concept for some non-scientific expertise but not others.

Indeed, *Kumho Tire* represents the end of the deferential regime that *Daubert* might seem to have put in place, notwithstanding the emphasis placed on professional standards.⁶³ When *Daubert* tied reliability (for scientific evidence) to the idea of scientific validity and suggested factors that might be taken into account in assessing validity, the opinion left unclear whether the judiciary should use those factors because the judgment of (hypothetical) good scientists is *determinative*, or rather because those factors are generally useful tools for reaching a legal-policy based assessment of validity that may or may not accord with what scientific norms would require. *Kumho Tire* appears to have resolved this question: “[A] trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of the expert testimony.”⁶⁴ What is “reasonable,” in turn, is a matter of legal policy, not scientific policy.⁶⁵ *Kumho Tire* thus frees *Daubert* to allow that there is an important analytical difference, even in the context of scientific evidence, between legal norms of reliability and norms of validity or reliability that inform the inference processes in any non-legal discipline that is invoked in testimony.⁶⁶

⁶² *Kumho Tire*, 526 U.S. at 147-49.

⁶³ *See id.* at 152 (indicating that one purpose of the reliability requirement is to assure that an expert employs “the same level of intellectual rigor that characterizes the practice of an expert in the relevant field”); *see supra* note 61.

⁶⁴ *Kumho Tire*, 526 U.S. at 152.

⁶⁵ *Cf.* FAIGMAN ET AL., *supra* note 4, § 1-3.5.1, at 43 (“Whether the basis for a proffered expert opinion is valid depends on what that opinion is, and what consequences follow from it. This is a matter of policy, not science.”).

⁶⁶ I am not suggesting any extreme post-modernism, any wholly social construction of truth. I present no challenge to the philosophically conventional thrust of the reliability factors identified by the Supreme Court and lower courts or the essentially veritistic aims associated with them. Of course, what is *taken* as true after a trial, the verdict upon which legal action is taken, is certainly a proposition

Dichotomy without Deference. Having rejected deference, we should consider whether it is possible to maintain such a dichotomous reliability concept without invoking *sub silentio* a policy-based balancing formula that renders the supposedly binary determination simply a report of the conclusion obtained by such balancing. Is there a way to map facts, including non-legal institutional facts, to such a dichotomous legal determination of reliability? The answer is certainly, “Yes,” but again it is very hard to find a workable formulation that is compatible with established doctrine and institutional capacities.

What exactly are we looking for here? One possibility is a binary factual determination the satisfaction of which, by the consequent admission of evidence, can be expected to improve verdict accuracy. A direct approach would have courts make assessments of the overall propensity of certain evidence to yield accurate verdicts. That is, proffered expertise would be considered reliable if its admission has the propensity to produce more accurate verdicts, in all cases or at least in cases of this type.⁶⁷ Alternatively, the suggestion might be that judges should assess directly whether admission will achieve the proper balance between the risk of false positive verdicts and the risk of false negative verdicts.⁶⁸ These suggestions are problematic, however, if they assume that the judge, in making admissibility rulings, uses knowledge of the truth about the disputed material facts. The reliability inquiry should not lead to a practice of using

that might diverge from what, on purely epistemic grounds, one ought to believe to be true, if for no other reason than the fact that considerable relevant evidence is privileged and thus removed from consideration. That does not alter the fact that the primary objective of the trial is veritistic, looking to the accurate application of substantive law to probable adjudicative facts. See JOHN RAWLS, *A THEORY OF JUSTICE* 85 (1971) (“The theory of trials examines which procedures and rules of evidence, and the like, are best calculated to advance this purpose consistent with the other ends of the law.”).

⁶⁷ Analogously, speaking about the evaluation of social practices in terms of the accuracy of its user’s beliefs (the beliefs’ “veritistic value” or “V-value”), Professor Goldman writes:

Suppose the range of pertinent applications of the practice has been settled. Theoretically, then, we can take the *average* (mean) performance of the practice across those applications as a measure of its V-value. If on average the practice increases the V-values of the user’s belief states, the practice has *positive* V-value. If on average it decreases the V-values of the user’s belief states, the practice has *negative* V-value. And if on average it leaves the V-values of the beliefs unchanged, the V-value of the practice is neutral.

GOLDMAN, *supra* note 11, at 92.

⁶⁸ See David L. Faigman, *Expert Evidence in Flatland: The Geometry of a World Without Scientific Culture*, 34 SETON HALL L. REV. 266-67 (2003).

admissibility to engineer what the trial judge believes to be the right verdict. On the other hand, such suggestions presumably mean something more demanding than simply the exclusion of expertise that does not assist the trier of fact.⁶⁹ If, then, the indicated suggestions are to occupy the precarious middle ground between these two conceptions of the judicial role, can they work? I think not, for they suffer from even more daunting practical difficulties than the deferential approach.

The difficulty already noted—that judges under *Daubert* must become surrogate experts in the relevant field—is thereby compounded by a different problem. Now the greater problem is that judges in each context would have to become a special kind of expert, predicting the impact of alleged expertise (of whatever species) on the adjudication of disputes. In the vast majority of contexts, moreover, this would have to be done without the benefit of extant social science regarding the behavior of triers of fact. For example, more so than having to become an expert in DNA technology in order to evaluate some new genetic evidence technique, the judge would have to become trained in the analysis of the impact of the new technique on juries, presumably as indicated by the social science of similar or related techniques, if any exist. Without such knowledge, the judiciary would have to proceed based on the collective prejudices of the legal profession as they are passed from generation to generation. Suffice it to say, the perspicacity of the latter has been shown to be seriously wanting.⁷⁰

And there are further problems. In this context, the determination of reliability likely will be dependent on the meaning given to “cases of this type.”⁷¹ Depending on the range of potential applications (admissions) of the contemplated expertise, it might be reliable or unreliable. Nor is this problem avoided by the emphasis in *Kumho Tire* on addressing reliability not in the abstract but rather in regard to the “task at hand.”⁷² Commentators have noted that this

⁶⁹ The idea of assisting the trier of fact to perform its factfinding task is addressed in Part III, *infra*.

⁷⁰ Outside the area of experts, empirical research has shown that professional expectations about jury behavior are often wide of the mark. See, e.g., Shari Diamond & Neal Vidmar, *Jury Room Ruminations on Forbidden Topics*, 87 VA. L. REV. 1857 (2001) (reporting results of observations of jury deliberations in regard to discussion of insurance); Peter Meine et al., *Juror Decision Making and the Evaluation of Hearsay Evidence*, 76 MINN. L. REV. 683 (1992) (reporting results of mock jury studies in regard to hearsay evidence). On professional expectations of jurors in regard to expert testimony, see discussion *infra*, notes 128-46 and accompanying text.

⁷¹ See *supra* text accompanying note 67.

⁷² 526 U.S. at 141 (quoting *Daubert*, 509 U.S. at 597).

appears to require not a global focus but a local one; reliability must be established relative to the particular inference the expert is being called upon to make.⁷³ There is an undeniable element of truth here, but pressed to its logical conclusion, this would make determinations of reliability all but impossible, for the particular task at hand in a lawsuit is never replicated in research. Trials, and the issues that they present, are unique events. In order to make any assessment of reliability of a claimed expertise, one must address the veritistic propensity of such testimony, the evidence for which is inherently statistical, based on experienced frequencies over some class of cases. Of course, the “task at hand” should not be understood too broadly—and certainly some courts have been guilty of such untailored analysis—but even with appropriate sharpening of focus, at the margin there may remain enough play in the joints that veritistic propensities can be manipulated by task selection.

These problems might be avoided if the courts were to employ a reliability test that is less epistemically demanding for judges. Many rules of admissibility that do not concern experts provide models of what this might look like. Consider the rule excluding hearsay. That rule does *not* provide that, when confronted with evidence of a statement not made on the stand in the current trial, the trial judge should consider designated factors, such as the absence of cross-examination of the declarant, and weigh these against the probative value of such evidence in order to decide admissibility. Nor does it say, more elliptically, that the trial judge should simply weigh all competing factors, without specification of what those might be. Instead, it requires categorically the exclusion of such a statement when offered to prove the truth of the matter asserted, if it does not come within one of the (for the most part equally) factually binary exceptions.⁷⁴ In other words, these dichotomous tests encapsulate, or so it is hoped, the goals of the adjudicative law, the most important of which in this context is verdict accuracy, but without requiring, at least in the ordinary case, any advertence to those goals or the balance of competing considerations that relate thereto. Indeed, this in many ways is the essence of rule-governed decision-making.⁷⁵

In this vein, one might select one or more dichotomous proxies for reliability, each of which tests for one particular and important

⁷³ See, e.g., D. Michael Risinger, *Defining the “Task at Hand”: Non-Science Forensic Science After Kumho Tire Co. v. Carmichael*, 57 WASH. & LEE L. REV. 767 (2000).

⁷⁴ See FED. R. EVID. art. VIII.

⁷⁵ See generally FREDERICK SCHAUER, *PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE* (1991).

way in which reliability can be undermined. If the expertise is not undermined in one of these specific ways, the expertise satisfies the reliability requirement. Such a proxy test, if well or fortuitously designed or evolved, might not be too over-inclusive or under-inclusive for practical use. Superficially, this might look like the scheme that the Supreme Court initiated with *Daubert*, its “factors” constituting the proxies. Such a scheme cannot, however, create an intelligible dichotomous standard of reliability, unless either (a) only a single, dichotomous proxy is used, so that the expertise would be deemed “reliable” relative to the present requirement provided the single proxy test is passed, or (b) in the multiple proxy context, one has a determinate way of synthesizing the results of those tests or factor applications into a single, binary judgment without performing an independent assessment of reliability. Pretty clearly, *Daubert* does neither. Its factors cannot yield simple “yes” or “no” answers; its formulation rejects any single factor as determinative; the opinion expressly leaves open the possibility of additional pertinent factors; and *a fortiori* it eschews any reasonably determinate means of synthesizing the various factors to be considered.⁷⁶

Of course, the Court’s reticence does not necessarily mean that no such reasonably determinate synthesis is possible, and at least one eminent evidence scholar has attempted to provide such a synthesis by listing binary conditions for a determination of reliability. Edward Imwinkelried suggests filling out the requirement in the following way, at least for scientific evidence:

[T]here is “enough” validation when the proponent’s foundation establishes that an hypothesis has been verified by an empirical study which: (1) is based on a large, representative database; (2) was conducted under conditions approximating those in the instant case; and (3) yielded a validity rate justifying the definiteness of the tendered opinion.⁷⁷

Like amended Rule 702, this suggestion identifies certain areas of interest on which trial judges should focus. That is certainly helpful, as far as it goes, but it soon throws up difficult questions: How large a database is large enough? How representative must it be to support this opinion? How close is close enough for the conditions to approximate those of the litigated case? How high a “validity rate” is high enough to justify an opinion of a given definiteness? These

⁷⁶ See *Daubert*, 509 U.S. at 593-95; see also *Kumho Tire*, 526 U.S. at 149-52 (also emphasizing *ex ante* flexibility).

⁷⁷ See 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, 50 (1999).

questions cannot be understood as binary questions of the type we are seeking here. They can only be answered by assessing what is at risk given the degree to which these parameters are not satisfied in the proponent's evidence. And this means that the criteria stated reflect, at best, conclusions of an analysis that implicitly invokes an unspecified weighing algorithm.⁷⁸

What else might we try? We might be more modest in the goals that a dichotomous reliability concept is expected to achieve. Perhaps all that is necessary in the context of an admissibility decision is the elimination of expertise that is unreliable in certain specific ways that are amenable to dichotomous tests. For example, drawing on the first articulated *Daubert* factor, suppose the gatekeeping role consisted solely of determining whether the basis for an opinion is testable. An expert in prayer is offered to testify that God revealed to him during prayer that the defendant is guilty. God also revealed during the same prayer that God will not reveal anything else to the expert, certainly not anything that can be independently confirmed. In such a context, the expert may be excluded on the ground that the basis for his testimony is untestable. Use of the other *Daubert* factors, suitably developed, could similarly eliminate supposed expertise that is defective in particular ways.⁷⁹

This might be developed into a coherent admissibility scheme,⁸⁰

⁷⁸ The idea might be that some kind of invisible hand mechanism, say the accumulating experience of judges, will allow the intuitive balancing of competing considerations without an explicit process of identifying the counterweights. This seems to be the underlying, perhaps unintended, idea in an earlier article by Professor Imwinkelried, from which he drew the test here quoted. See Edward J. Imwinkelried, *The Judge as Daubert Gatekeeper: Adapting Old Maps to the Unfamiliar Terrain of the "Brave New World,"* PROCEEDINGS OF THE 1999 NATIONAL CONFERENCE ON SCIENCE AND THE LAW 46 (1999), at <http://www.ncjrs.org/pdffiles1/nij/179630.pdf> (categorizing admissibility rulings according to an analogy with the burden of production on the merits). But it is hard to discern the mechanism that would produce a *desirable* convergence of rulings, especially in the rapidly changing arena of expert testimony, without explicit attention to the counterweights that determine, once the degree of reliability is assessed, how reliable is reliable enough to warrant jury consideration in the context of the case.

⁷⁹ Cf. 1 MCCORMICK ON EVIDENCE, *supra* note 5, at 736 n.65 (arguing that the *Daubert* reliability requirement might be read to exclude only "subjective belief or unsupported speculation" and "inference[s] or assertion[s] [not] derived by the scientific method," leaving any further oversight to be carried on pursuant to a general balancing of probative value and counterweights, presumably via Rule 403) (citation omitted).

⁸⁰ Such an approach would, in fact, bear some similarity to the work of philosopher Karl Popper, upon whom the *Daubert* Court confusingly relied. See Susan Haack, *An Epistemologist in the Bramble-Bush: At the Supreme Court with Mr. Joiner*, 26 J. HEALTH POL., POL'Y & L. 217, 230-32 (2001) (criticizing *Daubert's* reliance on Popper, noting that under Popper's philosophy of science, a claim or theory cannot

but it is definitely not what *Daubert* mandated. For *Daubert* did not simply require the exclusion of expertise that is defective in one or more of several specified ways. It mandated that only reliable expertise be admitted. Admission under *Daubert*, in its dichotomous reliability interpretation, thus involves a judicial determination that the expertise is reliable, not one that the expertise has not been found to be unreliable in designated respects. Expertise that is unreliable (or very unreliable, to put it in gradational terms), but not in one of the articulated ways, should not come in under *Daubert*, but would come in under the contemplated scheme. The two frameworks are not compatible.

The Downside of Dichotomy. Faced with the very difficult task of coming up with a coherent scheme of dichotomous reliability that is faithful to the mandates of the Supreme Court and Congress, there is some tendency for trial judges to try instead to answer a different, more readily answerable, but clearly dichotomous question. Courts might, for example, treat the admissibility decision as a sufficiency decision on the merits, excluding the proffered expertise unless it possesses sufficient reliability to meet the burden of production, at least to the extent the proponent bears such a burden. With exclusion of such evidence often resulting in summary judgment or directed verdict, it is tempting to treat the admissibility question as determined by the answer to the sufficiency question. Indeed, there is good reason to think this sort of thing is occurring.⁸¹

Treating reliability as a dichotomous characteristic of expert testimony thus reinforces a pre-*Daubert* trend: What some courts really have wanted to do in certain toxic tort litigation is to rule on the merits against the plaintiff; for want of suitable doctrinal means of so doing directly, they struck upon the idea of excluding the plaintiff's critical evidence and then granting the defense summary

be known to be true or even probable, but can only be "corroborated" by failures to falsify it).

⁸¹ Judge Kozinski's ruling on remand in *Daubert* illustrates it nicely. 43 F.3d 1311, at 1320-22 (applying the reliability test in light of California law on sufficiency of the evidence); see also Beecher-Monas, *supra* note 38, at 1073-74 (arguing that *Daubert* has contributed to the confusion in this regard); Berger, *supra* note 51, at 304-06 (discussing rulings excluding expert testimony in toxic tort cases when based on studies showing a relative risk of less than 2); *id.* at 322-25 (discussing doctrinal aspects of a conflation of admissibility with sufficiency). Certain statements in *Daubert* and *Kumho Tire* do encourage this. See *Kumho Tire*, 526 U.S. at 153 (apparently endorsing the trial court's assessment that the challenged expert testimony "fell outside the range where experts might reasonably differ, and where the jury must decide among the conflicting views of different experts, even though the evidence is 'shaky.' *Daubert*, 509 U.S. at 596.").

judgment.⁸² As Sam Gross has emphasized, it may well be important to facilitate explicit preemptive determinations on the merits by judges in cases for which the scientific evidence is overwhelmingly one-sided, but the entrenchment of a reliability-based admissibility regime is not the right way to achieve this goal.⁸³ Here, as elsewhere, the distinction between admissibility and sufficiency should be kept clear in order to avoid confusion, mistaken rulings, and the improper use of precedents.⁸⁴

Deciding sufficiency in the guise of admissibility might or might not yield the same results that would occur if the admissibility and sufficiency determinations were kept properly distinct. For example, in deciding whether to admit one expert opinion or set of opinions under such a sufficiency standard, if no other evidence favoring the proponent on the disputed fact is considered, such a procedure places an artificially high burden on the proffered expertise. It is not clear what grounds can be offered for requiring that the expert testimony be enough *by itself* to sustain the burden. Even if all other favorable expert evidence on that disputed fact is considered in making the sufficiency-imbued admissibility determination, the proponent may still be disadvantaged by not having the benefit of all non-expert evidence relating thereto.⁸⁵

Moreover, even if the trial court considers all evidence in the case on the material issue in question (including that which the

⁸² See Samuel R. Gross, *Substance & Form in Scientific Evidence: What Daubert Didn't Do*, in REFORMING THE CIVIL JUSTICE SYSTEM 234, 246-48 (Larry Kramer ed., 1996). "As the Bendectin cases illustrate, courts do not like to admit evidence and say it's insufficient. Instead, they will go to unfortunate lengths to find that essential parts are inadmissible, and then say that there's not enough left to go to the jury." *Id.* at 252.

⁸³ *Id.* at 255-61. Gross identifies part of the problem as the conventional but poorly reasoned principle of the law of sufficiency that the opinion of any qualified expert is sufficient to sustain a verdict coinciding with the opinion. This principle erroneously treats expert opinion as if it were just another form of direct evidence, whereas in fact it is a form of expert-assisted circumstantial inference. *Id.* at 252-53.

⁸⁴ See, e.g., Dale A. Nance, *Conditional Relevance Reinterpreted*, 70 B.U. L. REV. 447, 449-59 (1990) (discussing confusion of admissibility with sufficiency that has helped to spawn the perverse doctrine of conditional relevance).

⁸⁵ Replying to criticisms of the conflation of admissibility with sufficiency, Professor Mueller argues that there is nothing wrong with excluding evidence because it is insufficient to meet the burden of production, *provided* the proponent has an adequate opportunity to advise the court about all additional evidence that may be forthcoming on the issue. Christopher B. Mueller, *Daubert Asks the Right Questions: Now Appellate Courts Should Help Find the Right Answers*, 33 SETON HALL L. REV. 987, 1001 (2003) (citing cases that illustrate such a practice). This, of course, assumes that the court is willing to consider all such evidence, expert and non-expert, in making its supposed admissibility ruling.

proponent is yet to introduce), the reliability requirement still amounts to the multiplication of events that trigger a summary determination, unless the admissibility of all the proponent's expertise is decided in a single pre-trial hearing. Otherwise, with each item of expertise offered on the issue by a party bearing the burden of production, the opponent gets, in effect, an opportunity to move for summary judgment. With more bites at the apple, so to speak, such a procedure entails more cumulative probability of summary determination, as well as more expense incurred by a proponent who, in some cases, will be at a sizeable resource disadvantage. It is unclear whether effectively increasing the burden of production is an intended or appropriate consequence, but no justification for doing so only in the presence of expert testimony has been provided.⁸⁶

Finally, the reliability requirement must be applied to parties not bearing a burden of production; Rule 702 does not distinguish on this ground. The obvious example is expertise offered by a criminal defendant to negate the prosecution's affirmative case. If, however, the same methodology is applied by analogy to such parties, then their evidence may be subjected to artificially high admissibility hurdles. Why, that is, should relevant expert testimony, unlike other evidence offered by an accused, be excluded simply because it is not enough to raise a reasonable doubt? Worse yet, why should it be excluded just because similar evidence was not enough to warrant a finding by a preponderance of the evidence in some civil case that is cited by analogy, or because it was not enough to warrant a finding beyond reasonable doubt when offered by the prosecution in a criminal precedent? This can be expected once these precedents are transmuted from sufficiency rulings into admissibility rulings.

Even worse possibilities exist. In their search for a manageable dichotomy to substitute for a determination of sufficient reliability, courts may try instead to determine which expert's testimony on a particular matter, the plaintiff's or the defendant's, is based on *more* reliable information, excluding the loser in this comparative

⁸⁶ In addition, sufficiency rulings are subject to *de novo* appellate review, while many admissibility rulings—including reliability determinations under *Daubert*—are subject to review only for abuse of discretion. See FAIGMAN ET AL., *supra* note 4, § 1-3.6 (arguing, however, that some reliability related issues should be considered *de novo*). Burying a sufficiency ruling within an admissibility ruling may thus have the effect of insulating the former from proper appellate review. Cf. John J. Gibbons, *Tenth Anniversary of the Supreme Court's Decision, Daubert v. Merrell Dow*, 34 SETON HALL L. REV. 127 (2003) (discussing confusion in the review of admissibility rulings when they are part of a summary or directed judgment).

evaluation.⁸⁷ Such a decision starts with a gradational conception of reliability but then transposes it into a binary one by way of the comparison. It also amounts to a decision on the merits by the trial judge.⁸⁸ The same is true if courts presume to think non-comparatively by excluding alleged expertise that the court believes is simply incorrect, or if they try to decide which side should win and then admit or exclude evidence based on its propensity to produce the “right” result. These various substitute decisions may be exactly what some interest groups would like to see by way of tort reform, but they are dangerous substitutions if the right to a jury trial is to be preserved, indeed if the right to a trial is to be preserved at all.⁸⁹

Reliability and Related Gradational Attributes. Yet there is a germ of insight in such substitute approaches. Many of the problems associated with a binary reliability regime can be avoided if we think in terms of *comparative* evaluation rather than *isolated* evaluation of reliability, *provided* the proper comparison is made. As with the use of analogies in substantive law, one can sometimes answer the comparative question more easily or more coherently than the isolated one.⁹⁰ Instead of asking whether a given item or type of proffered expertise is, relative to the suitably defined task at hand, reliable *vel non* (the isolated sense), or whether it is more reliable than that offered by the opponent (appropriately a question for the trier of fact), we might better ask whether a given item or type of proffered expertise is, in the context, more or less reliable than some

⁸⁷ See Beecher-Monas, *supra* note footnote 38, at 1073 n.151 (asserting that this confusion may occur). Once again, language can be found in *Kumho Tire* that would encourage this kind of thinking. See 526 U.S. at 155 (comparing the challenged expert’s report on certain foundational facts, concerning the remaining tire tread, with testimony by *opposing* experts).

⁸⁸ This possibility necessitated warnings from the drafters of the 2000 amendment: “When a trial court, applying this amendment, rules that an expert’s testimony is reliable, this does not necessarily mean that contradictory expert testimony is unreliable. The amendment is broad enough to permit testimony that is the product of competing principles or methods in the same field of expertise.” See FED. R. EVID. 702 advisory committee’s note to 2000 Amendments. The drafters did not, however, go on to explain in what sense of “reliable” this idea can be maintained.

⁸⁹ See Richard D. Friedman, *Squeezing Daubert Out of the Picture*, 33 SETON HALL L. REV. 1047, 1057 (2003) (stating that “a fair definition . . . is that an item of evidence is reliable proof of a phenomenon if the existence of the evidence indicates a high probability that the phenomenon is true” and arguing that under such a conception of reliability, “[t]he whole notion of a trial, and indeed the whole purpose of the factfinder, falls away if we say that a given piece of evidence can be presented to the factfinder only if it is reliable”).

⁹⁰ Cf. GOLDMAN, *supra* note 11, at 92 (indicating the advantages of comparative veritistic evaluation of social practices).

alternative presentation by that party, including the possibility of simply dropping the challenged expertise from the evidentiary mix.

To take advantage of the comparative evaluation of potential evidence packages, however, requires an acknowledgment of the different mode of thinking about reliability itself, for it depends, as already noted, on the ability to say that one package is *more reliable* or *less reliable* than another, which implicitly accepts that reliability is a matter of degree, like “probative value” and “credibility.” In Part III, I will suggest ways to incorporate explicitly the gradational conception of reliability into the law of admissibility. In so doing, I hope to further indicate that a gradational reliability concept is not only more appealing theoretically, but also more workable in practice.

Before doing so, however, we should pause to note that the exact relationships among reliability, credibility, and probative value, as these terms are used in the federal rules, are unclear. Common usage is vague, and the terms are not authoritatively defined. For example, expert credibility might be viewed as simply part of what determines expert reliability, or credibility of the expert might be distinguished from the reliability of the expertise itself.⁹¹ The latter formulation would reserve credibility assessments for the trier of fact by limiting reliability assessment under Rule 702 to matters unrelated to witness credibility. That might be desirable, but the oft-expressed concern about bias among experts suggests that the effects of bias must be taken into account in applying the reliability screening mandated by the rule.⁹²

Similarly, reliability might be considered one component of probative value, or the two might be considered distinct. As a theoretical matter, it has been suggested that the probative value of an item of evidence be identified with its likelihood ratio relative to the contending hypotheses in the case.⁹³ If so, then one might also

⁹¹ See, e.g., BERNARD ROBERTSON & G.A. VIGNAUX, INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM 7-8 (1995) (noting that the term reliability is variously used to refer to test sensitivity, quality control, discriminatory power, or witness honesty).

⁹² Bias seems to be a major concern, for example, when courts making reliability assessments take into account whether experts are “proposing to testify about matters growing naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying.” *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1317 (9th Cir. 1995).

⁹³ In general terms, the likelihood ratio for evidence, E, is the probability that such evidence would be presented under the hypothesis of guilt (or liability), P(E|G), divided by the probability such evidence would be presented under the

identify reliability, or rather *unreliability*, as a measure of the extent to which evidence causes divergence between the subjective likelihood ratio of the trier of fact and (some measure of) the objective likelihood ratio that the trier ought to attribute to that evidence.⁹⁴ The greater that divergence, the lower the reliability.⁹⁵ This suggestion seems to limit unreliability to a certain tendency to mislead.⁹⁶ But this misses an important point: evidence that is not misleading at all, at least in the indicated sense, can nonetheless be considered to have weak reliability; it may have a degree of unreliability that is perfectly well understood by the decision maker and thus reflected in its assessment of probative value. As I will argue below, focusing exclusively on the tendency to mislead the trier of fact, at least in the sense expressed in the suggested measure, gives the wrong focus to the reliability concern in the context of expert testimony.

For the purpose of developing a workable interpretation of Rule 702, therefore, it would be unwise to adopt a narrow and theoretically specific formalization of reliability. Instead, we should work with its unrefined—but gradational—meaning, indicating a degree of dependability or trustworthiness.⁹⁷ In comparative terms, one item of evidence is more reliable than another if the former is more dependable or trustworthy.⁹⁸

hypothesis of innocence (or non-liability), $P(E|\text{not-G})$. It is a measure of the relative consistency of the evidence with the two contending hypotheses. See Richard Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1023-24 (1977).

⁹⁴ For an interesting analysis of the idea of “objective” likelihood ratios, see Alvin I Goldman, *Quasi-Objective Bayesianism and Legal Evidence*, 42 JURIMETRICS J. 237 (2002).

⁹⁵ See REDMAYNE, *supra* note 5, at 118-20 (building on the work of philosopher Alvin Goldman).

⁹⁶ Redmayne discusses various problems with this model of reliability. See *id.* at 120-24.

⁹⁷ Cf. *id.* at 118 (“I think the common use of the term does have a coherent, if unarticulated, meaning, which treats ‘reliable’ as a synonym of ‘dependable’ or ‘trustworthy.’”).

⁹⁸ Professor Imwinkelried, in his comment on a draft of this Essay, chides me for not providing a specific definition of reliability. See Edward J. Imwinkelried, *The Relativity of Reliability*, 34 SETON HALL L. REV. 269, 282-83 (2003). To be sure, I believe that my claim that the reliability concept in Rule 702 is better interpreted as gradational rather than as binary (which is not the same thing as saying that it should be understood as “relative” to the inferential task at hand, though this is also true, see *id.* at 285) can be demonstrated without providing a canonical definition of reliability. Indeed, providing such a definition is not the role of someone offering an interpretation of an authoritative term. Such definition must be given by the courts or by Congress. In interpreting the rule, what commentators may do is to *investigate* or *refine* meaning, not to *prescribe* it. Nevertheless, in response to Imwinkelried’s understandable query, the best that I am currently able to do in this regard is to offer

III. TOWARD A POLICY-BASED BALANCING FRAMEWORK

In order to determine whether a differential in reliability warrants exclusion of the less reliable, it is crucial to look not only to the factors that affect the degree of reliability—factors that have been explored extensively in the extant literature—but also to the factors that would warrant the exclusion of relevant but insufficiently reliable evidence. One commentator has written:

[C]ourts and commentators have identified the varied considerations that determine the balance of probative value and prejudice of scientific evidence. Applying these to various types 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 of evidence.⁹⁹

This assertion is, I think, largely correct, except that courts and commentators so far have done a much better job of identifying “the varied considerations that determine . . . probative value” than of examining the factors related to “prejudice” that would inform the indicated balance. The latter is a major source of our continuing problems in understanding the reliability requirement. How then shall we complete the analysis in such a way as to give due respect to the teachings of the Supreme Court and the mandate of Congress in amended Rule 702?

Avoiding Blind Alleys. Start by noting what should *not* be done. It does not help, and it might hurt, to invoke the notion of a “threshold of reliability,” above which expertise is considered acceptable in court and below which it is not. Unfortunately, prominent scholars—even some who disagree significantly about other aspects of expert testimony—often agree, explicitly or implicitly, that the post-*Daubert* regime rightly requires expertise to exceed such a threshold of

the following comparative conception as an interpretation of what I think lawmakers might be groping toward: E_1 is more reliable than E_2 , as evidence regarding proposition X, if and only if, without regard to whether either item of evidence favors X or not-X, but assuming (even if it is not true) that one favors X while the other favors not-X, a rational person required to wager on the truth of the proposition favored by E_1 or that favored by E_2 would bet on that which is favored by E_1 . The hypothetical veil of ignorance—deciding which evidence is worth the bet without regard to which side of the dispute it favors—represents the core of good sense of the statement in *Daubert* that “the focus, of course, must be solely on principles and methodology, not on the conclusions that they generate.” 509 U.S. at 595. I cannot say that anything in my argument in the text depends on the precise terms of this definition, but it may give the reader a better idea of what I have in mind. Or not. Professor Imwinkelried’s call for a definition may actually have been intended as a call for a formula with which to compute a real-valued measure of reliability. Obviously, I make no attempt to provide such a formula.

⁹⁹ I MCCORMICK ON EVIDENCE, *supra* note 5, at 737 (footnote omitted).

reliability.¹⁰⁰ Of course, if this mode of expression is understood simply as a metaphor for the use of another analytical framework, such as that of Rule 403, then it poses little danger, provided that framework is clearly identified.¹⁰¹ Yet, there are reasons to be skeptical of its employment even as metaphor.

One serious problem with this method of transforming continuous variation into dichotomous choice is that it suggests, although it does not explicitly require, an invariant threshold that applies across disciplines and across litigation contexts. The mental picture is an ascending scale of reliability with a mark that separates the insufficiently reliable from the sufficiently reliable. The suggestion is that the location of this mark on the scale, as distinguished from the question of whether the particular evidence exceeds the threshold it indicates, is identifiable, at least in principle, without an appreciation of either the nature and type of expertise or the alternatives to it that may exist in the context of a particular case. This pernicious idea is essentially a resurrection, in the limited context of expert testimony, of the now-discredited “plus-value” requirement that Wigmore described for all evidence.¹⁰² The idea is misleading in that it suggests that expertise can be determined to be “sufficiently reliable” by considering only the degree of reliability, without considering the particular counterweights that cut against admission in context.

Indeed, even without the imagery of the uniform minimum that

¹⁰⁰ Compare, e.g., Imwinkelried, *supra* note 77, at 42-45 (not explicitly using the “threshold” terminology, but arguing that “when the question posed is the sufficiency of a foundation [for an expert opinion], the issue is whether the foundation is adequate to support the specific opinion proffered, not whether the expert has relied on the best scientific technique available”), with David L. Faigman, et al., *How Good is Good Enough?: Expert Evidence Under Daubert and Kumho*, 50 CASE W. RES. L. REV. 645, 659, 664-65 (2000) (arguing for the use of a “better evidence” principle, but acknowledging the idea that the post-*Daubert* regime requires “a minimal threshold of reliability”).

¹⁰¹ Professor Faigman, in his comment on this paper, disclaims any suggestion that the idea of a “minimum threshold of reliability” can suffice to do the work of mapping gradational reliability into the dichotomous decision on admissibility, but he nonetheless argues that *Daubert* requires “some minimal amount of testing.” The only explanation that he provides for this, however, involves an example (prediction by “tarot cards, tea leaves and crystal balls”) that surely would result in exclusion under Rule 403 without any help from Rule 702. See Faigman, *supra* note 68, at 259. Perhaps, then, the invocation of a “minimum threshold” by Faigman et al., *supra* note 100, is really just an assertion that Rule 403 excludes purported expertise that is so weakly probative as to be a waste of time. If so, I would agree, but that does not speak to the interpretation of Rule 702.

¹⁰² See 1A JOHN H. WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 28 (Tillers rev. 1983).

makes the threshold imagery appealing, it remains vacuous. It merely restates the “sufficient reliability” idea without telling us anything affirmative about how to fill it out. Without more, one is left to suppose that the significance of the (unidentified) counterweights can be assessed by macroscopic gestalt, rather like identifying pornography by the “I know it when I see it” test.¹⁰³ The same would be true of unadorned exhortations to admit only that which is “reasonably reliable,” although this locution is less likely to call for the invisible marker imagery. Without giving attention to the parameters in the balance that determine whether a given degree of reliability is reasonable for admission, any such formulation effectively requires the judge to answer the question: “Does the degree of reliability of the expertise outweigh?” Unless something follows the word “outweigh,” the question is practically incoherent. Outweigh *what?*

Michael Graham’s interesting attempt to resolve the reliability puzzle is instructive in this regard.¹⁰⁴ Professor Graham identifies two interpretations of the reliability requirement that struggle for recognition in the courts and that illustrate dichotomous and gradational conceptions of reliability, although he does not use this terminology.¹⁰⁵ The first interpretation would require trial courts to determine “whether the [expert’s] explanative theory . . . produces a correct, accurate, truthful, valid conclusion.”¹⁰⁶ Under the second interpretation, trial courts would determine “whether there are sufficient assurances present . . . that the expert witness’s explanative theory produces an accurate result to warrant jury acceptance.”¹⁰⁷ The first interpretation, which Graham rightly rejects as too demanding of experts and of courts,¹⁰⁸ instantiates the dichotomous approach. Graham argues for the second interpretation as more within judicial capacities and better suited to the “‘liberal thrust’ to admissibility of expert witness testimony that was favored by *Daubert*.”¹⁰⁹ In doing so, he sometimes explicitly uses language that

¹⁰³ The quote is from Justice Stewart’s famous concurring opinion in *Jacobellis v. Ohio*, 378 U.S. 184, 197 (1964) (Stewart, J., concurring).

¹⁰⁴ See Michael 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).

¹⁰⁵ *Id.* at 336-39.

¹⁰⁶ *Id.* at 317.

¹⁰⁷ *Id.* at 317, 339.

¹⁰⁸ *Id.* at 356.

¹⁰⁹ *Id.* at 356-57.

accepts the gradational nature of reliability.¹¹⁰

Now, to ask whether there are sufficient assurances of accuracy to warrant jury *acceptance* (rather than jury *consideration*) of the expert's opinion or information appears on its face to fall into the trap already noted of conflating admissibility with sufficiency.¹¹¹ But the matter becomes less clear once one examines the details of such a determination as conceived by Graham. An expert's testimony would pass his test if either the expert's theory "has gained widespread acceptance in the particular field to which it belongs,"¹¹² or (if neither widely accepted nor widely rejected¹¹³) it "possesses particularized earmarks of trustworthiness."¹¹⁴ The first prong of this test seems to resurrect, modify, and extend the applicability of *Frye*, which may or may not be a good idea.¹¹⁵ However that may be, the idea of demanding, in default of such widespread consensus, only "particularized earmarks of trustworthiness" is as incomplete here as it is in the law of hearsay and confrontation, from which it is presumably drawn.¹¹⁶ If not understood as simply instrumentally related to a sufficiency ruling on the merits, itself problematic once again, this idea is just another way of articulating the essentially vacuous notion of a "threshold of reliability." And the problem remains: expertise cannot be known to pass a threshold—from insufficiently reliable to sufficiently reliable—*just* by identifying particularized earmarks of trustworthiness, no matter what they are or how many of them are accumulated. Something is necessarily missing from such an analysis; one must know, in some sense, where the threshold is in order to know whether it has been passed. Few would doubt the relevance of some such earmarks—presumably along the lines of the *Daubert* opinion and subsequent elaborations

¹¹⁰ *E.g.*, Graham, *supra* note 104, at 341 n.62.

¹¹¹ *See supra* notes 80-84 and accompanying text.

¹¹² Graham, *supra* note 104, at 340.

¹¹³ *Id.* at 346.

¹¹⁴ *Id.* at 340.

¹¹⁵ Graham acknowledges that "reliance on 'widespread acceptance' is subject to the limitation that the entire field to which the explanative theory belongs does not itself simply 'lack[] reliability.'" *Id.* n.61. This suggests that the widespread acceptance idea is instrumentally related to, and therefore subject to being trumped by, other considerations in regard to what is ultimately a sufficiency determination.

¹¹⁶ The idea of looking for particular indications of "trustworthiness"—Justice Blackmun's synonym of choice for reliability (*see supra* note 7 and accompanying text)—is also prominent in the law of hearsay and the prevailing constitutionalization of hearsay in the confrontation clause. On the inadequacy of this lore, see Richard D. Friedman, *Truth and its Rivals in the Law of Hearsay and Confrontation*, 49 HASTINGS L.J. 545 (1999). *See also* Richard D. Friedman, *Confrontation: The Search for Basic Principles*, 86 GEO. L.J. 1011 (1998).

thereof—in the assessment of reliability, but once again that assessment constitutes only one side of the balance that needs to be considered in order to make an appropriate admissibility decision.

The blind alley into which conventional analysis has wandered is nicely illustrated by a treatise statement made in the context of discussing the role of Rule 403 in screening expert testimony for reliability:

Under Rules 702 and 104(a), judges must decide whether the proponent of scientific evidence has demonstrated the validity of the scientific basis for the testimony by a preponderance of the evidence. In many cases, however, while judges might find scientific evidence to be “valid,” they might believe that it is not valid enough, in light of the dangers associated with its use.¹¹⁷

I do not take issue with this statement as a description of prevailing conceptions. My problem concerns those conceptions. While it is certainly true that Rule 403 serves as a residual exclusionary principle that is analytically distinct from Rule 702, the point of the discussion in Part II is that conventional understandings of Rule 702 do not allow for any usable, non-balancing, rule-like criterion to which Rule 403 can serve in such a capacity. Consequently, it makes no sense to say that the validity determination can be meaningfully separated from a consideration of the “dangers associated with [the evidence’s] use.” Once again, the binary conception of reliability suggested by the reference to a demonstration of validity “by a preponderance of the evidence” tends to conceal this problem.

Connecting Reliability to Counterweights. How, then, can we articulate a more coherent balancing analysis that gives explicit attention to the counterweights? In fact, we can draw on a rich body of theory about the rationale of admissibility rules, for many such rules attempt to answer, in particular contexts, whether evidence of a certain type is sufficiently reliable to be admitted.¹¹⁸ As I have detailed elsewhere, these rules respond to one or more of three basic concerns: (1) that the trier of fact is prone to err in its assessment of the probative value of certain types of evidence or in its application of the burden of persuasion when faced with such evidence (concerns that are expressed primarily in regard to jury trials); (2) that the offered item of evidence will be of too little probative value to warrant the necessary expenditure of time and resources, including

¹¹⁷ FAIGMAN ET AL., *supra* note 4, § 1-3.8, at 59-60.

¹¹⁸ A similar strategy is employed, to somewhat different effect, in REDMAYNE, *supra* note 5. Redmayne argues that “it is inconsistent for the English legal system to apply reliability-based exclusionary rules to hearsay and confession evidence, while not applying one to expert evidence.” *Id.* at 126.

the cognitive resources of the trier of fact, necessary to incorporate it into the decision-making process (generally of concern in both jury and bench trials); and (3) that the trier of fact (judge or jury) should not be forced to assess the case based on the offered evidence when superior evidence is likely to be available.¹¹⁹

The first two concerns lead to an admissibility structure that prefers the exclusion of the challenged evidence *ceteris paribus*, that is, all other evidence in the case being (hypothetically) the same. They differ in that the first entails a much larger degree of epistemic paternalism than the second. The third concern, however, leads to an exclusionary structure that prefers the replacement of the offered evidence with something else, better evidence. It reflects more the problem of *advocate control* than the problem of *jury control* that tends to animate the first concern. Jury credulity, after all, is not the only potential source of inaccuracy in adjudication. Even in a bench trial, the court will want the best evidence reasonably available.¹²⁰ Indeed, even in a trial by a panel of experts, were one to be undertaken, the panel would want such evidence.¹²¹

Assuming that we are to deploy a sensitivity to these traditional concerns to interpret a “sufficient reliability” requirement, is there anything general and affirmative that one can say about how such an analysis should proceed? A few parameters frame the discussion to come. We are looking for an analysis that assumes testimony by a qualified expert that is relevant to a disputed issue and specifies an additional requirement related to reliability. The strategy is to develop the implications of a gradational reliability concept in the context of each of the three concerns articulated above. We want a standard, or set of standards, that is responsive to the concerns

¹¹⁹ See Nance, *supra* note 11, at 1554-60. The analysis sets aside rules, such as those privileging certain confidential communications against disclosure, that rest largely on non-veritistic concerns relating to incentives for conduct, or the protection of relationships, outside the courtroom.

¹²⁰ *Id.* In a series of articles, I have argued at length that the third concern is more pronounced in the legitimate rationales of admissibility rules than conventionally thought. See, e.g., Dale A. Nance, *A Theory of Verbal Completeness*, 80 IOWA L. REV. 825 (1995); Nance, *Conditional Relevance Reinterpreted*, *supra* note 84; Dale A. Nance, *Understanding Responses to Hearsay: An Extension of the Comparative Analysis*, 76 MINN. L. REV. 459 (1992). The most general treatment is Dale A. Nance, *The Best Evidence Principle*, 73 IOWA L. REV. 227 (1988) [hereinafter *The Best Evidence Principle*].

¹²¹ Of course, a trier of fact can try to make allowances for the absence of evidence, but this is a second best solution as compared to having the better evidence before the tribunal. See Dale A. Nance, *Evidential Completeness and the Burden of Proof*, 49 HASTINGS. L.J. 621, 633-39 (1998) (examining reasons not to rely upon jury discounting to handle problems of evidential incompleteness).

articulated above but not simply redundant of Rule 403.¹²² This presupposes that the Supreme Court did not simply intend to deliver the message that trial courts have been improperly applying Rule 403 in the context of expert testimony. More importantly, it reflects the general consensus that *Daubert* did not adopt the position advanced most conspicuously by Dean McCormick, well before the adoption of the federal rules, that would treat relevant testimony by a qualified expert in a way no different from most other kinds of evidence the admissibility of which is challenged, subjecting it to the standard balancing of probative value against competing concerns now reflected in Rule 403.¹²³

Nevertheless, because all three types of concerns identified above can be addressed by an analysis under Rule 403, and because analysis under that rule is at least reasonably well understood, it is possible to use that rule as a starting point in developing the kind of analysis needed for Rule 702. Indeed, one norm clearly serves to distinguish the contemplated balancing test under Rule 702 from balancing under Rule 403. Under 403, the burden is on the objecting party to convince the trial judge that the testimony's probative value is outweighed by the indicated risks.¹²⁴ In contrast, it is generally agreed that the burden is on the proponent of expertise

¹²² A familiar principle of statutory construction is that a body of rules should be interpreted so as not to render any one thereof wholly redundant of the others and therefore pointless. See 2A NORMAN J. SINGER, SUTHERLAND'S STATUTES AND STATUTORY CONSTRUCTION § 46:06 (6th ed. 2000).

¹²³ For McCormick's view, see CHARLES T. MCCORMICK, HANDBOOK OF THE LAW OF EVIDENCE 363 (1954), in which McCormick criticized *Frye* as more appropriately a test for taking judicial notice. He wrote, "Any relevant conclusions which are supported by a qualified expert witness should be received unless there are other reasons for exclusion. Particularly, its probative value may be overborne by the familiar dangers of prejudicing or misleading the jury, unfair surprise and undue consumption of time." *Id.* at 363-64. On *Daubert's* rejection of the McCormick view, see Paul C. Giannelli, *Daubert: Interpreting the Federal Rules of Evidence*, 15 CARDOZO L. REV. 1999, 2009-12 (1994), explaining that *Daubert* is more demanding than the standard Rule 403 analysis. See also 1 MCCORMICK ON EVIDENCE, *supra* note 5, § 203 (comparing *Frye* and *Daubert* with McCormick's preferred standard).

¹²⁴ Nominally, Rule 403 authorizes exclusion only when probative value is "substantially outweighed" by one of the indicated concerns. This presents a puzzle: How can it be that we should refrain from excluding evidence the probative value of which is *outweighed* by a concern, but not *substantially outweighed* by it? If the probative value *is* outweighed by the concern, then the evidence should be excluded, should it not? What else could it mean to say that it is "outweighed"? This puzzle afflicts all Rule 403 jurisprudence, and the answer must be this: "substantially outweighed" is a sloppy locution intended to mean "*clearly* outweighed," so that doubts about the balance of probative value and counterweights are resolved in favor of the proponent. This seems to be how courts interpret the rule. See CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, EVIDENCE § 4.9, at 190 (2d ed. 1999).

to establish (sufficient) reliability under 702.¹²⁵ One might leave it at that, switching the allocation of the burden but otherwise understanding the 702 reliability decision as a replication of the 403 balancing test.¹²⁶ That approach would certainly simplify the problem of articulating a coherent reliability test under Rule 702.¹²⁷ For reasons explained below, however, I believe it is less than optimal.

Offsetting Jury Credulity. Both before and after *Daubert*, what has often been stated as the primary concern warranting the exclusion of proffered expertise of dubious reliability is the idea that a lay jury will be misled by such evidence, giving it greater credence than it deserves.¹²⁸ This might seem to follow from one of the defining characteristics of expertise, that jurors lack specialized information available to the expert. It coheres with suspicions about the credibility of experts for hire, if it is thought that jurors are unable to discount for the effects of bias. As is well known, *Daubert* itself was decided within a climate of concern over jury credulity in the face of junk science.¹²⁹

This view of the matter has led to perceptive criticisms of the Court's solution to the problem. For example, Susan Haack poses the troubling rhetorical question: "[I]f judges need to act as gatekeepers to exclude scientific evidence which doesn't meet minimal standards of warrant because juries may be taken in by flimsy scientific evidence, how realistic is it to expect juries to discriminate the better from the worse among the half-way decent?"¹³⁰ Such criticisms carry undeniable force if juror credulity is the source of the

¹²⁵ See FAIGMAN ET AL., *supra* note 4, § 1-3.1.2.

¹²⁶ This is the solution favored by my colleague, Calvin Sharpe, in his comment on this Essay. See Calvin W. Sharpe, *Reliability Under Rule 702: A Specialized Application of 403*, 34 SETON HALL L. REV. 289, 307 (2003).

¹²⁷ This approach would provide something resembling a "minimum threshold of reliability" at least theoretically distinguishable from Rule 403. See *supra* note 101. Unlike the abstract use of that phrase, however, the present suggestion (a) identifies the counterweights that are to be assessed in context, and (b) requires a weighing of the probative value of the evidence, as affected by the degree of its reliability, against the identified counterweights.

¹²⁸ Pre-*Daubert*: see, e.g., Edward J. Imwinkelried, *Judge Versus Jury: Who Should Decide Questions of Preliminary Facts Conditioning the Admissibility of Scientific Evidence?*, 25 WM. & MARY L. REV. 577, 580 (1984); Moenssens, *supra* note 27, at 564-67. Post-*Daubert*: see, e.g., Donnelly & Friedman, *supra* note 53, at 974; Graham, *supra* note 104, at 356; Clifton T. Hutchinson & Danny S. Ashby, *Daubert v. Merrell Dow Pharmaceuticals, Inc.: Redefining the Bases for Admissibility of Expert Scientific Testimony*, 15 CARDOZO L. REV. 1875, 1878-79 (1994).

¹²⁹ See, e.g., PETER W. HUBER, *GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM* (1991).

¹³⁰ Haack, *supra* note 80, at 233.

Court's concern.¹³¹

Strikingly, however, the majority opinion in *Daubert* was at pains to disavow any serious skepticism about jury competence, opining that it is “overly pessimistic about the capabilities of the jury” to think that abandonment of *Frye*'s general acceptance test would result in litigation “in which befuddled juries are confounded by absurd and irrational pseudoscientific assertions.”¹³² Although the Court was anything but clear about why a reliability inquiry *is* needed, other than to say that—needed or not—it is somehow mandated by the language of Rule 702 (which at the time contained no explicit reference to reliability), the Court seemed fairly clear in distinguishing jury credulity as a source of its concerns.¹³³ Neither the opinion in *Kumho Tire*, nor the advisory committee's explanation of the 2000 amendment speaks to the contrary.¹³⁴

In this, the law's promulgators were right. The available empirical evidence points to jurors being remarkably conscientious in their work and not demonstrably less accurate in their inferences than judges.¹³⁵ More specifically, so far we have no empirical basis to conclude that jury credulity in over-crediting expert testimony is a serious or pervasive problem. After a careful review of the available studies, Neil Vidmar and Shari Diamond recently concluded simply: “Empirical data do not support a view that juries are passive, too-credulous, incompetent, and overawed by the mystique of the expert.”¹³⁶

¹³¹ See also Michael S. Jacobs, Testing the Assumptions Underlying the Debate About Scientific Evidence: A Closer Look at Juror “Incompetence” and Scientific “Objectivity,” 25 CONN. L. REV. 1083, 1086-93 (1993) (maintaining that arguments against juror competence are often intuitively implausible or logically inconsistent).

¹³² *Daubert*, 509 U.S. 595-96 (expressing confidence in the adversary system's ability to handle this sort of problem in most cases).

¹³³ See Andrew Taslitz, *Daubert's Guide to the Federal Rules of Evidence: A Not-So-Plain-Meaning Jurisprudence*, 32 HARV. J. ON LEGIS. 3, 62-65 (1995).

¹³⁴ One can try to dig further. For example, in an earlier constitutional case, not involving the interpretation of Rule 702, Justice Blackmun, the author of the majority opinion in *Daubert*, made a statement that suggests a concern about jury credulity. See *Barefoot v. Estelle*, 463 U.S. 880, 916 (1983) (Blackmun, J., dissenting) (“In a capital case, the specious testimony of a psychiatrist, colored in the eyes of an impressionable jury by the inevitable untouchability of a medical specialist's words, equates with death itself.”). But is hard to know how seriously to weight such a statement as a proposition warranting the reliability requirement in *Daubert*. In any event, we need not base interpretation of amended Rule 702 on undisclosed concerns of the judges in *Daubert*.

¹³⁵ See generally JOHN GUNTHER, *THE JURY IN AMERICA* (1988); HARRY KALVEN, JR. & HANS ZEISEL, *THE AMERICAN JURY* (1966).

¹³⁶ Neil Vidmar & Shari Seidman Diamond, *Juries and Expert Evidence*, 66 BROOK. L. REV. 1121, 1180 (2001); see also REDMAYNE, *supra* note 5, at 107, 109-12 (reviewing the

Ironically, if there is distortion in jurors' evaluation of expertise, extant empirical studies suggest that the problem is likely to be the opposite of what is usually assumed. It appears that jurors sometimes *undervalue* complex expert testimony of relatively high probative value, especially when its significance for the disputed issues is not well explained.¹³⁷ Based on such results, one might argue that the reason to exclude expertise of poor reliability is not because it will be overvalued, but rather because contrary expertise of much greater reliability offered by the opponent will be undervalued, leading to an inaccurate weighing of the opposing proofs. The problems with such an argument, however, are pretty clear: it conflates admissibility with sufficiency and distorts the latter. It attempts to deal with a sufficiency problem—that jurors might improperly weigh competing evidence—by excluding evidence that is not itself subject to demonstrated overvaluation. And it prevents the jury from combining the evidence of low reliability with independent evidence, if there is any, that might be enough in combination rationally to convince the jury. The better response to the indicated problem, when it occurs, is to explore ways to reduce the jurors' undervaluation of the superior expertise by improving their comprehension of its significance.¹³⁸

The foregoing observations are consistent with, but do not entail, the more general proposition that juries have greater difficulty as the scientific or technical complexity of the evidence increases. There is some limited evidence to support such a general proposition, although the evidence is subject to various interpretations.¹³⁹ To be sure, the idea is not particularly surprising,

empirical studies and concluding that “the focus on jury competence is somewhat misplaced”); Jacobs, *supra* note 131, at 1094-98 (arguing that doubts about juror competence with respect to scientific evidence fail to take into account the growing body of empirical evidence to the contrary).

¹³⁷ See, e.g., Dale A. 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) (confirming with summoned jurors the results of earlier research conducted mostly with students: while a few jurors seem to make identifiable mistakes that might lead them to overvalue scientific “trace” evidence with a quantifiable random match probability, the greatly predominant problem seems rather to be that juries *excessively discount* the probative value of such evidence).

¹³⁸ See, e.g., *id.* (finding that variation in the manner of presenting the random match probability affects the extent of jury undervaluation of trace evidence relative to Bayesian norms).

¹³⁹ See, e.g., Richard O. Lempert, *Civil Juries and Complex Cases: Taking Stock After Twelve Years*, in VERDICT: ASSESSING THE CIVIL JURY SYSTEM 181-247 (Robert E. Litan ed., 1993) (reviewing the available evidence and concluding (at p. 235) that “the weight of the evidence indicates that juries can reach rationally defensible verdicts in

for it is difficult to imagine a plausible definition or criterion of “complexity” that would not render the proposition true to some degree, if not tautologically so. Almost by definition, complex cases are difficult cases. Consequently, an analogous proposition will be true for any decision makers, including judges or panels of experts. And even if judges’ veritistic acumen falls off less steeply than juries’ as complexity increases—which we cannot as yet conclude—that would seem to speak to the question of the wisdom of jury trials in complex cases, not to the question of the advisability or criterion of excluding evidence.¹⁴⁰

In his present contribution Joseph Sanders tries to bridge this gap, developing the best argument yet identified for grounding the Rule 702 reliability determination in epistemic paternalism toward the jury.¹⁴¹ Building on the slender reed of the jury’s asserted difficulties in complex cases, he concludes that “on balance the empirical research does lend some support to the paternalistic justification for restrictions on the admissibility of unreliable expert testimony.”¹⁴² As I understand his argument, however, the most that can be inferred from the available data is that, *if* judges are capable of making the necessary discriminations (which we do not know), then “close scrutiny of testimony [by judges] is potentially most beneficial in complex cases where peripheral processing [by jurors] is most likely.”¹⁴³ And Sanders acknowledges that we do not have evidence from which to conclude that judges are capable of identifying evidence that is “unreliable” relative to the goal of minimizing erroneous verdicts (or the costs thereof) due to either (a) jury misunderstanding of admitted expertise or (b) jury error induced by exclusions.¹⁴⁴

complex cases”); *cf.* SANDERS, *supra* note 9, at 193 (“If there is one lesson to be drawn from these cases, one single overarching problem revealed by the Bendectin litigation, it is that in cases involving complex scientific evidence juries have a difficult time reaching the truth.”).

¹⁴⁰ Even with regard to the choice between bench trial and jury trial, such a showing would be incomplete. If, for example, juries are much better than judges in relatively uncomplicated cases, then juries’ reduced acumen in complex cases might still be higher than judges’. See LEMPERT, *supra* note 139, at 234-35 (concluding that there is still no evidence that judges will perform better than juries in complex cases).

¹⁴¹ See Joseph Sanders, *The Paternalistic Justification for Restrictions on the Admissibility of Expert Testimony*, 33 SETON HALL L. REV. 881 (2003).

¹⁴² *Id.* at 938.

¹⁴³ *Id.* at 937. “Peripheral processing” refers to the jury’s use of factors such as credibility cues as opposed to an assessment of the rational strength of the inference suggested by the expert. *Id.* at 909.

¹⁴⁴ *Id.* at 937-38.

This valiant effort, therefore, does not close the argumentative gap. Because the evidence does not show whether such judicial scrutiny is on balance beneficial or harmful, even in complex cases, the claim that judicial scrutiny is “potentially most beneficial” in certain cases is the same as saying that it is “potentially least harmful” in such cases. In other words, at best Sanders has shown that if judges are successfully trained to make good assessments of the degree of reliability and exclude expertise on paternalistic grounds when they think the evidence has low reliability (a big “if”), then the expected benefit of such a practice increases as the degree of complexity of the case increases, but the expected benefit (for all we know) might not be *positive* over any identifiable range of complexity. Thus, the argument necessarily falls short of a justification for a reliability-based exclusionary practice applied in all cases involving expert testimony.

Until we have more evidence of jury credulity, as well as more evidence that judges are in fact capable of offsetting such incompetence by their decisions to exclude expertise, respect for the political significance of the jury and for those citizens who are called to service on juries dictates that we should not readily exclude supposed expertise *on this ground*. Instead, we should leave this kind of problem to be handled under the pro-admissibility standard articulated in Rule 403, as in fact prescribed by the majority in *Daubert*.¹⁴⁵ Indeed, before excluding evidence on this ground, courts should at least consider and rule out less drastic remedies such as judicial comment on the evidence.¹⁴⁶

Monitoring Adversarial Presentations. Much more prominent in the *Daubert* and *Kumho Tire* opinions is a concern about regulating the introduction of expert testimony so that it will properly *assist* the trier of fact.¹⁴⁷ The question is how *excluding* proffered expert testimony can be of assistance, if the concern is not about juror credulity. Obviously, expertise that is very unreliable may be excluded under Rule 403 as simply a waste of time. This is not an insignificant limitation on the use of experts, given the far-fetched proffers that occasionally find their way into courtrooms. But the

¹⁴⁵ 509 U.S. at 595; *see also In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 746-47 (3d Cir. 1994) (per Becker, J.) (accepting that *Daubert* altered prior law in the circuit by taking consideration of the danger of overwhelming or confusing the jury out of the Rule 702 analysis and committing it to the Rule 403 analysis); FOSTER & HUBER, *supra* note 38, at 207-09 (noting that Rule 403 is the appropriate tool for dealing with problems of jury credulity).

¹⁴⁶ *See* Friedman, *supra* note 89, at 1064.

¹⁴⁷ *See Kumho Tire*, 526 U.S. at 149, 156; *Daubert*, 509 U.S. at 591, 592, 593.

question here is whether we can go further to find exclusionary principles that are more demanding than this obvious time-saving norm.¹⁴⁸ What is there in expert testimony that might call for such a heightened burden?

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,¹⁴⁹ 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.¹⁵⁰ 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.¹⁵¹

The implication is that greater judicial management or monitoring of evidence might be needed in this context in order to prevent adversarial incentives from undermining the accuracy of trials. To be clear, this implication derives not so much from a concern that jurors cannot or do not rationally take these process considerations into account, but rather from the view that they should not be required to do so when a better alternative can be found. The next two sections examine possible arguments for intervention in the adversarial exchange on reliability grounds, suggesting how such interventions might be structured to assist the

¹⁴⁸ I put aside the intensely skeptical suggestion that would exclude all conflicting expert testimony before the trier of fact because neither a judge nor a jury has the background to decide between conflicting experts, a suggestion that can only be tied to some sort of court appointed expert system. See, e.g., Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40, 51-56 (1901). The suggestion may have some merit, but it exceeds the scope of my present inquiry.

¹⁴⁹ See FED R. EVID. 602, 701.

¹⁵⁰ See RICHARD O. LEMPERT ET AL., A MODERN APPROACH TO EVIDENCE 1013-16 (3d ed. 2000).

¹⁵¹ One need not embrace an extreme fact skepticism to see in these phenomena the ability to undermine verdict accuracy. See REDMAYNE, *supra* note 5, at 5-35 (discussing “case construction” using forensic science in criminal cases). To be sure, concerns about experts for hire being especially prone to distort their testimony are probably exaggerated, at least in civil cases, given the market mechanisms that will influence experts. See RICHARD A. POSNER, FRONTIERS IN LEGAL THEORY 401-04 (2001).

trier of fact.

Marshaling Cognitive Resources. One plausible line of argument begins with the idea of helping the trier of fact (whether judge or jury) to focus its attention—its cognitive resources—on the most reliable forms of expertise offered by each side on a disputed issue. When multiple proffers of expertise address a single issue, one tool to be considered is to remove from the trial the least reliable evidence offered by a party on that issue.

Suppose, for example, that a party needs to prove the value of a vacant parcel of land. The party offers several different forms of expertise that relate to the question. One is a survey of prices at which twenty vacant parcels, of identified acreage, within a mile of the subject parcel were sold. The opponent raises questions about the methodology of the survey. The proponent also offers the testimony of three experienced realtors who have visited the subject parcel and made “gestalt” assessments based only on their visits (and their experience) and not, for example, on the information contained in the survey of prices. Each “gestalt expert” places the per-acre value of the disputed land at a figure higher than the average per acre figure from the survey. The opponent also raises questions about the reliability of such gestalt assessments. In such a case, it would be plausible, depending on the details of the challenges raised, for a trial court to conclude that the survey of prices is substantially more reliable, as a measure of the value of the disputed parcel, than the gestalt testimony. The court might then assist the trier of fact by excluding the latter, forcing the parties to focus, before the jury, on the details of the challenge to the reliability of the survey.¹⁵²

How would this be of assistance? If we put aside the possibility of distortion due to jury credulity, why would not the jury always be better off with more relevant information rather than less, subject to the limits of cumulativeness? It may be difficult for lawyers or experts, accustomed in their reflective moments to an all-things-considered, use-every-piece-of-available-information type of decision

¹⁵² An alternative resolution might be to insist, as a condition of admissibility, that the “gestalt” witnesses make less subjective appraisals by incorporating the survey data in the basis of their opinions. This kind of response is addressed in the following subsection. Another possibility, of course, is that the gestalt witnesses might be considered to be so weakly reliable as to make consideration of them a waste of time and inadmissible under Rule 403. It is, of course, difficult to come up with an example for which the application of Rule 403 is incontestable, so for present purposes, assume that the gestalt witnesses would not be properly excluded pursuant to Rule 403.

making, to accept the possibility that “less is more” in the adjudicative context. Yet there are reasons to entertain the idea. Decision theory has shown that sometimes basing decision on only selected features of a decision environment can be as successful or more successful, and less costly in time and other resources, than a more nuanced decision scheme that takes all pertinent factors into account.¹⁵³ Even restricting consideration to a single factor can be remarkably accurate if that factor is the best single factor for the decision.¹⁵⁴ These findings present at least the possibility of a strategy that can generate improvements in the accuracy of verdicts as well as a reduction in the cost of trials. Such a strategy would give concrete meaning to that portion of the *Daubert* opinion that rejects the need to have as complete and wide-ranging debate on the matters addressed in litigation as is needed in the scientific process.¹⁵⁵

Unfortunately, the efficacy of particular decision heuristics is context dependent. To work well, a simple decision heuristic must take advantage of information structures in particular decision environments.¹⁵⁶ At this point, we have no experiments or empirical studies that can establish how this plays out in the context of adversarial trials. For the time being, therefore, a trial judge would best approach the problem by placing herself in the position of the jury and asking whether the challenged expertise is so unreliable, in comparison to other expertise offered by the same party, that *the judge* would ignore the challenged proffer if she were the trier of fact. This viewpoint takes advantage of any applicable decision heuristics

¹⁵³ See generally GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART (1999).

¹⁵⁴ *Id.* at 73-168 (discussing the “Take the Best” heuristic).

¹⁵⁵ Justice Blackmun wrote:

It is true that open debate is an essential part of both legal and scientific analyses. Yet there are important differences between the quest of truth in the courtroom and the quest for truth in the laboratory. Scientific conclusions are subject to perpetual revision. Law, on the other hand, must resolve disputes finally and quickly. The scientific project is advanced by broad and wide-ranging consideration of a multitude of hypotheses, for those that are incorrect will eventually be shown to be so, and that in itself is an advance. Conjectures that are probably wrong are of little use, however, in the project of reaching a quick, final, and binding judgment—often of great importance—about a particular set of events in the past.

Daubert, 509 U.S. at 596-97.

¹⁵⁶ The successfulness of a heuristic depends on its “ecological rationality,” the domain-specific match between the heuristic and the information environment in which it is employed. See Gerd Gigerenzer, *The Adaptive Toolbox*, in BOUNDED RATIONALITY: THE ADAPTIVE TOOLBOX 38, 46-48 (Gerd Gigerenzer & Reinhard Selten eds., 2001).

available to the judge to economize cognitive resources. In jury trials, it also helps to prevent slipping into a consideration of jury credulity, since it is unlikely that the judge will consider herself to be subject to irrational overvaluing of the challenged evidence. This viewpoint is one of *representation*: The trial judge represents the jury, whose silence cannot operate to exclude evidence that its members would consider comparatively too unreliable to be of use.¹⁵⁷

Here, then, is one place where the stricter demands of *Daubert*, *Kumho Tire*, and Rule 702 might be put into effect. As noted above, conventional wisdom is that the effect of these developments is to place the burden on the proponent to establish “reliability.”¹⁵⁸ 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 403.¹⁵⁹ Distinguishing this analysis from that of Rule 403 would help to maintain the representational viewpoint described earlier.¹⁶⁰

Despite this default principle favoring exclusion, an important limitation of the present analysis is that it presupposes there is expertise of discernibly superior reliability offered by the proponent of the challenged evidence. Sometimes no such judgment is possible. For example, returning to our land appraisal case, suppose the proponent’s experienced realtors base their otherwise gestalt appraisals on an unusual characteristic of the parcel not shared by the twenty parcels included in the survey of sale prices. In such a

¹⁵⁷ Cf. *The Best Evidence Principle*, *supra* note 120, at 291 (suggesting a broader use of the representative perspective).

¹⁵⁸ See *supra* note 125 and accompanying text.

¹⁵⁹ This would reverse the argumentative burden only when items of expertise of significantly different reliability are involved. If, for example, the opponent’s complaint is simply that too many witnesses are offered to testify to the same effect, then the burden would remain on the opponent, under Rule 403, to convince the judge that proffered witnesses would be cumulative.

¹⁶⁰ To provide notice to the proponent and to sharpen the issues, the opponent should have the initial burden of providing a reason to think that the challenged expertise is less reliable than other expertise offered by the proponent on the point. But then the proponent would have to surmount the persuasive hurdle for the challenged expertise to be ruled not inadmissible on this ground. Fed. R. Evid. 104(a) would govern the determinations, as indicated in *Daubert*, 509 U.S. at 592.

case, a trial judge might properly conclude that both types of evidence should be considered by the trier of fact. The important point, in any event, is that the assessment should proceed not in terms of whether a particular proffer is sufficiently reliable or reasonably reliable, *considered by itself*, but rather whether the proponent's package of expertise on the issue is as reliable, and thus as helpful to the jury, as it might be if narrowed by the exclusion of some of what is offered. The comparative nature of the evaluation is crucial.¹⁶¹

There is a further implication, one that will not sit well with those who see in the reliability requirement a quick route to summary judgment. If the challenged testimony is the *only* expertise offered by a party on the disputed fact, the judge would not be warranted in excluding it, at least not under the theory posited. For example, in our hypothetical real estate appraisal case, if the proponent presented only the gestalt witnesses, they could not be excluded in the interest of marshaling cognitive resources. The reason for this limitation is that the present rationale of exclusion is to assist the jury in focusing on the best expertise each party has on the issue. With no other expertise on the proponent's side, there is nothing presented with which to make a comparative reliability assessment except the expertise, if any, offered by the opponent. To use that as the basis for comparison would be, for practical purposes, to address the merits of the case. Consequently, if the party's only expertise on the matter is to be excluded, it would have to be done on some other ground, such as to avoid mistakes due to jury credulity or to avoid wasting the jury's time with very unreliable expertise—each of which would require meeting the more demanding (for the opponent) standard of Rule 403—or to induce the proponent to present better expertise, about which more follows in the next section.

It is important, therefore, not to overstate the impact of adopting the foregoing analysis as compared to a scheme that would dispose of a separate reliability inquiry under Rule 702 and leave such matters to be handled under Rule 403. Obviously, some

¹⁶¹ The demands of this analysis may not be circumvented by a proponent's gratuitous inclusion of extremely weak expertise that would otherwise not have been submitted, on the theory that its exclusion under the analysis suggested here would leave the proponent's other evidence untouched. That will not work, since the principle discussed in this section may be iterated; with the strategically inserted expertise out of the way, the court could proceed to test the proponent's next weakest evidence as if the subterfuge had not taken place. There would be, therefore, no incentive for a proponent to throw in sacrificial junk expertise, unless of course some judges systematically misapplied the principle suggested here.

comparatively unreliable expert testimony could and would be excluded under Rule 403 anyway. Still, the suggested analysis might effectively, if marginally, narrow the scope of offered expertise in a way that would assist the trier of fact to focus on the best expert evidence offered by each side. Whether the limited veritistic gain to be derived from such a marshaling of cognitive resources is worth the added complexity in the reliability inquiry is difficult to assess *a priori*. I have stated the case for such an exclusionary principle as well as I can. Nevertheless, I am skeptical that this is the kind of reliability inquiry for which we are searching.

My skepticism derives from three main considerations. First, relying (as we must, for the time being) on the trial judge's intuitions about what will facilitate cognitively economical decision-making might produce as much heuristic error as it does veritistic improvement. Judges, after all, are not immune from cognitive illusions.¹⁶² There is, moreover, considerable controversy among those who study the accuracy of human decision-making about the usefulness of commonly employed decision heuristics.¹⁶³ And, as noted above, none of the extant research speaks directly to the question of selection of evidence, expert or otherwise, in an adversarial trial setting where admissibility decisions are separated from evaluations of evidentiary weight. Because heuristics can be relatively successful or not, depending on the particular decision environment, it is hard to say with any confidence that significant improvement will result from a practice in which judges exclude the least reliable expertise that a party offers.

Relatedly, one might question whether it is necessary to *exclude* evidence to achieve the contemplated benefits, assuming they are non-trivial. Will not the trier of fact employ its own heuristics, placing its serious attention on the subset of admitted expertise that will allow it to render a reasonably accurate and prompt verdict?¹⁶⁴ This is possible, and to the extent that it is true, the exclusion of

¹⁶² There is evidence, for example, that judges are subject to heuristic errors such as anchoring and hindsight bias. See Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777 (2001) (reporting the results of a study of 167 federal magistrate judges).

¹⁶³ Compare Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Daniel Kahneman et al. eds., 1982) (presenting a relatively pessimistic assessment), with GIGERENZER ET AL., *supra* note 153 (presenting a more optimistic assessment).

¹⁶⁴ The evidence that ordinary people actually employ particular "fast and frugal" heuristics that have been theoretically identified and tested is as yet sketchy. Peter M. Todd, *Fast and Frugal Heuristics for Environmentally Bounded Minds*, in BOUNDED RATIONALITY, *supra* note 156, at 63-67.

expertise serves primarily to streamline trials and potentially improve accuracy by saving the cognitive resources that would be devoted by the trier of fact to selecting those forms of expertise that should receive more detailed attention. On the other hand, what is saved in terms of the jury's time could be lost in terms of the judge's time.¹⁶⁵ Perhaps future research will shed light on these matters in a way that will make the potential benefits less speculative.

The third consideration relates to the workability of the exclusionary principle and arises primarily from the asymmetry that the proposed structure would put in place. Suppose, for example, in our hypothetical property valuation case, that the plaintiff's gestalt witnesses are excluded because she has presented more reliable survey evidence as well; the trial judge believes that the gestalt witnesses will only be distracting. Suppose, then, that defendant offers its own gestalt witnesses as its only expertise on the issue. As indicated above, the present principle of exclusion would not be available upon plaintiff's objection, with the consequence that defendant would be able to introduce expertise with the same or a lesser degree of reliability as that which was foreclosed to the plaintiff upon defendant's objection. Many, including a party in the plaintiff's situation, will consider this result unfair. This sense of unfairness, to the extent that it is well-founded, will be largely attributable to concerns about the way that a jury will react to the situation.¹⁶⁶ There is a danger that the jury will (rationally!) draw a negative inference from plaintiff's failure to present counter evidence in the form of gestalt witnesses, or something similar thereto, witnesses who, by the trial court's admission thereof, have been seemingly certified as reliable enough to warrant the jury's consideration.¹⁶⁷

¹⁶⁵ As a further factor, it is possible that without judicial assistance in winnowing the evidence to be considered, the jury might feel obligated to neglect its own common sense in favor of a searching consideration of every item of evidence admitted, an effort that could be distracting. But this remains highly conjectural.

¹⁶⁶ In a bench trial, the plaintiff will have substantial solace in the fact that the trial judge has already gone on record in ruling survey evidence more reliable than gestalt witnesses.

¹⁶⁷ Scholars have discussed the idea that jurors may be misled when they expect to receive evidence that a party has tried unsuccessfully to introduce. See generally Bruce A. Green, "The Whole Truth?": *How the Rules of Evidence Make Lawyers Deceitful*, 25 LOYOLA L.A. L. REV. 699, 699-704 (1992); Stephen Saltzburg, *A Special Aspect of Relevance: Countering Negative Inferences Associated with the Absence of Evidence*, 66 CAL. L. REV. 1011 (1978). Empirical evidence confirms that negative inferences from missing evidence can occur. See, e.g., Bettyruth Walter, *The Civil Juror: A Research Project Sponsored by the Roscoe Pound Foundation*, in GUNTHER, *supra* note 135, at 285, 313; Tina M. Webster et al., *Voices from an Empty Chair: The Missing Witness Inference and the Jury*, 15 LAW & HUM. BEHAV. 31, 39-41 (1991).

There are a number of possible solutions to this kind of problem, including informing the jury that the judge excluded some of plaintiff's offered expertise and the reasons therefor, or allowing the plaintiff to use her gestalt witnesses in rebuttal once the defendant has "opened the door" to such. But these solutions entail some extension of the jury's consideration to the very evidence the exclusion of which was supposed to streamline the trial. To be sure, this kind of problem will not arise in all cases affected by the exclusionary principle articulated,¹⁶⁸ but its potential reduces the overall benefit to be derived by adding the complexity of a test based on marshaling of cognitive resources.

For the time being, then, the *most* that Rule 702 should do, in the name of marshaling cognitive resources, is to reverse the burden of Rule 403 regarding whether a proponent's evidence is sufficiently reliable as not to constitute a waste of the tribunal's time and resources. Upon objection, the proponent would be required to show that this balance clearly favors admission. Even if a more discriminating exclusionary structure like that discussed above is premature, the trial judge under this simpler framework might at least be sensitive to the importance that the judge as trier of fact would attribute to the various items of expertise offered by the same proponent and to any rational heuristics the judge would employ as trier of fact.

This simplified structure, however, would pose greater risk of the exclusion of *all* of a party's expertise, as its potential reach extends to even the party's most reliable expertise. This in turn presents again the risk of judicial misuse of Rule 702 to exclude expertise when, and just because, the opposing expertise is more reliable. And it is problematic whether judges would be able to maintain the distinction between this analysis and the pro-admissibility standard that, I have argued, is appropriate when jury credulity is the expressed concern. It would be easy to slide from a consideration of what would be useful to the judge (and *therefore* to the jury) to what would be useful to the judge but not to the more gullible jury. Consequently, while this framework might be workable, I remain skeptical. I would rather leave the matter to the conventional Rule 403 analysis, with the burden of argument on the opponent of the expertise, until such time as a more discriminating framework becomes warranted by the empirical evidence.

¹⁶⁸ The defendant in the example might not attempt to introduce pure gestalt witnesses, but might instead offer one or more witnesses whose expertise is not discernibly less reliable than the survey evidence admitted for the plaintiff.

Securing Better Evidence. One exclusionary principle remains. 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. This idea is present in a wide variety of rules,¹⁶⁹ including Rule 403.¹⁷⁰ Here again, we may ask whether the “produced” nature of expert testimony requires that a heavier burden be placed on a proponent thereof than is imposed by Rule 403 in this and other contexts. The post-*Daubert* reliability requirement arguably entails an affirmative answer.

An illustration can be found in the regulation of the use of pretrial experiments:

In passing on the admissibility of pretrial experiments, courts often assert that the experiment must be “substantially similar” to the conditions of the litigated event. However, what constitutes substantial similarity is all but impossible to discern from the case law without inquiring into the possibility that a better experiment could have been done. The best explanation of the results in these cases seems to be that “whenever the marginal benefits of a more refined experiment do not exceed the marginal costs, and the experiment is probative, the substantial-similarity requirement should be deemed satisfied.”¹⁷¹

Several distinguished scholars have recognized the pertinence of this kind of consideration in implementing the reliability requirement for expert testimony, emphasizing that 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.¹⁷² I

¹⁶⁹ For example, secondary evidence of the contents of a document is excluded in preference for the original, FED. R. EVID. 1002. For numerous other examples, see the articles cited *supra* note 120.

¹⁷⁰ One factor a court should consider in deciding whether to exclude evidence challenged under Rule 403 is the “availability of other means of proof.” FED. R. EVID. 403 advisory committee’s note. This is amply demonstrated by the decision in *United States v. Old Chief*, 519 U.S. 172 (1997), holding that in some circumstances evidence of prior crimes offered by the government may be excluded in favor of a stipulation to the conviction when the stipulation involves no less legitimate probative value but less prejudicial potential than the proffered evidence.

¹⁷¹ Faigman et al., *supra* note 100, at 654 (footnotes omitted).

¹⁷² See FAIGMAN ET AL., *supra* note 4, § 1-3.5.1. The matter was debated in recent issues of the Case Western Reserve Law Review. See Faigman et al., *supra* note 100; Edward J. Imwinkelried, *A Final Comment—The Importance of the Procedural Framework*, 50 CASE W. RES. L. REV. 669 (2000); Imwinkelried, *supra* note 77. The important point to be derived from this exchange is that the better evidence idea need not be taken as entailing some strict hierarchy of types of evidence, whether expert or

would go further: this kind of consideration represents the primary, if not exclusive argument structure appropriate for the reliability inquiry of Rule 702. For want of a better alternative structure, the principal question here is whether this consideration can be elaborated into a workable framework for decision. In the following paragraphs, I briefly explore the parameters of such a framework.

The most important consideration, once again, is the viewpoint to be adopted by the judge when making the decision. The judge should ask whether a reasonable jury, sensitive to the delays and costs associated with a demand for more reliable expertise, would express such a preference. Once again, the judge acts to facilitate good inferences by speaking on behalf of the silent jury, not to channel the jury's inferences out of suspicion of jury credulity. On behalf of the jury, the judge will want the best evidence that is or should have been reasonably available, with due regard to the adversarial structure of the trial, in knowledge of which the trial judge has a distinct comparative advantage relative to the jury.¹⁷³

One very important consequence is the negative implication of the principle: when contemplated alternative expertise is not discernibly more reliable, or when such discernibly more reliable expertise is not reasonably available to the proponent, either because it is impossible to acquire or because the costs of its acquisition would be disproportionate to the stakes involved, this principle provides no exclusionary authority. The best that is reasonably available should be admitted, at least so far as the reliability requirement of Rule 702 is concerned. Beyond that, an important limiting principle again places the better evidence idea within an adversarial context. Specifically, when the discernibly more reliable expertise is reasonably available *to the opponent* to present if the opponent should deem it worthwhile, then no exclusion of evidence is warranted; in most cases adversarial incentives should suffice to provide the necessary assistance to the jury.¹⁷⁴ Thus, upon a challenge that

otherwise, but can, and in this context should, involve a context-dependent assessment of what can be gained by excluding proffered evidence. See Dale A. Nance, *Conditional Probative Value and the Reconstruction of the Federal Rules of Evidence*, 94 MICH. L. REV. 419, 454-56 (1995) (emphasizing the importance of this distinction in the context of doctrines supposed to rest on the concept of conditional relevance).

¹⁷³ That is part of the reason that it is sub-optimal to leave the question of whether better evidence is reasonably available to the jury's evaluation of evidential weight. Of course, the more important reason not to leave the matter to the jury is that the jury can be assisted by having the better evidence. See *supra* note 121.

¹⁷⁴ For the unusual case in which the more reliable expertise, although available to the opponent (thus making exclusion of proponent's less reliable expertise

identifies the potentially better evidence that should be brought to bear and convinces the trial judge that such evidence is not reasonably available to the challenger, the proponent would bear the burden of convincing the trial judge that such evidence either is not significantly more reliable or is not reasonably available to the proponent. Once again, it is this placement of a burden on the proponent, coupled with the facilitative judicial viewpoint explained above, that serves to distinguish the reliability inquiry of Rule 702 from the somewhat similar balancing test of Rule 403.

In order to be more concrete, and to indicate qualifications to the foregoing general principles, it is necessary to distinguish two contexts in which an objection might be raised. As a general proposition, expert testimony can be divided into two components, each of which might be subject to a reliability challenge: (a) the non-case-specific information or general explanatory theory or technique that is to be conveyed or applied—essentially the matters addressed by Rule 702(1) & (2); and (b) the application of such information to the specifics of the litigated case, usually by presenting an opinion—addressed by Rule 702(3).¹⁷⁵ When an objection based on the better evidence idea is raised with regard to the first component, the issue concerns whether or not to expand the informational or theoretical resources brought to bear on the case. The potential costs thereof are greater than when the objection concerns the form or content of the expert's opinion. A problem with the latter can be corrected with relatively little cost by adjusting the expert's testimony. While the opponent may be able to induce concessions on cross-examination as to the weakness of a particular opinion, there is sometimes considerable inconvenience to the jury, and no compelling need, to require such a delay.¹⁷⁶ The following subsections consider these two contexts in more detail.

— *Testimonial Form and Content.* Current practice is often overly

unwarranted), is not presented by either side, the trial court retains the option of seeking court-appointed expert assistance. See FED. R. EVID. 706.

¹⁷⁵ In the interests of both accuracy and uniformity, the deference to trial court discretion mandated by *General Electric Co. v. Joiner*, 522 U.S. 136, 145 (1997), in terms of the standard of appellate review, should be construed to apply only to the latter, and even in the latter context, some issues may transcend the particular case and require rulings as a matter of law that are reviewed *de novo*. See Saks, *supra* note 39, at 231-35.

¹⁷⁶ The same point lies behind the exclusion of lay opinion in favor of more concrete factual reports. See 1 MCCORMICK ON EVIDENCE, *supra* note 5, § 11 (characterizing the modern lay opinion rule as a rule of excusable preference that is enforced notwithstanding the opportunity of an opponent to develop the more concrete facts on cross-examination).

generous to proponents in allowing opinions on case-specific material facts, ultimate or not, when those facts are not within the personal knowledge of the expert. Although Rule 704 rightly swept away the *per se* exclusion of opinions on ultimate issues,¹⁷⁷ courts have failed adequately to heed an important caveat to that change:

The abolition of the ultimate issue rule does not lower the bars so as to admit all opinions. Under Rules 701 and 702, opinions must be helpful to the trier of fact, and Rule 403 provides for exclusion of evidence which wastes time. These provisions afford ample assurances against the admission of opinions which would merely tell the jury what result to reach¹⁷⁸

Overly generous allowance of opinions on such ultimate issues is a major contributing cause of many perceived problems in the use of experts.¹⁷⁹

Modern developments in forensics theory point the way to improvements in conventional practice consistent with the better evidence inquiry. Specifically, it would be better to require an expert, whenever practicable, to frame testimony in a way that effectively communicates a likelihood ratio for the case-specific facts to which the expert's explanatory theory applies.¹⁸⁰ If the data are not available that would allow quantitative or even qualitative measures of the likelihood ratio, then the expert, again to the extent practicable, should be limited to providing specialized information that the jury can use to reach its own, typically intuitive sense of the likelihood ratio.¹⁸¹ Such restriction of testimony recognizes that, for the expert to give an opinion on an ultimate issue, the expert must implicitly weigh the other evidence in a case, including evidence that goes beyond any expert's asserted expertise.¹⁸²

¹⁷⁷ See FED. R. EVID. 704 (now prohibiting opinions on ultimate issues only in the context of testimony regarding the mental state or condition of a criminal defendant).

¹⁷⁸ FED. R. EVID. 704 advisory committee's note.

¹⁷⁹ See Risinger, *supra* note 73, at 767 (citing generous treatment of ultimate issue opinions as well as liberal rules of evidentiary sufficiency).

¹⁸⁰ Of course, the best way to communicate a likelihood ratio, when it can be at least partly quantified, is a difficult matter. See, e.g., Nance & Morris, *supra* note 137 (investigating different presentation formats for evidence with a quantifiable random match frequency); Samuel Lindsey, et al., *Communicating Statistical DNA Evidence*, 43 JURIMETRICS J. 147 (2003) (investigating additional formats).

¹⁸¹ Some such developments are indeed occurring, in areas such as handwriting identification. See DAVID L. FAIGMAN ET AL., SCIENCE IN THE LAW: FORENSIC SCIENCE ISSUES § 3-1.4.3 (2002) (discussing post-*Kumho* cases in which courts have admitted expert testimony about similarities in handwriting but excluded the expert's opinion on the identity of the writer).

¹⁸² See Richard Friedman et al., *Expert Testimony on Fingerprints: An Internet Exchange*,

It would unduly prolong this paper to explain these positions in detail, and with the necessary qualifications. It implicates the recurring dilemma of whether expert testimony should provide appropriate education for the jury or, instead, provide simply the occasion for the jury to defer to the opinions of the more credible expert.¹⁸³ But at least I can say this: If an expert's opinion is rejected under Rule 702 (as distinct from Rule 403) on the ground that it goes too far beyond what the foundational facts and principles would seem to warrant, that rejection should not be justified on the ground that the proffered opinion is "unwarranted," while a more conservative opinion is "warranted," except insofar as this is taken as an indirect way of saying that the more conservative opinion is *better* warranted within the scope of the expertise than the opinion originally proffered.¹⁸⁴

A second, surely less controversial application of preferential regulation of the form of testimony provides the most plausible argument for a certain deference to non-legal institutional norms. When an expert witness *asserts*, explicitly or implicitly, that the basis for his testimony is a body of knowledge that meets the standards of some discipline, then it is only fair to hold the witness to such standards. Thus, a witness who claims his field is "scientific" may

43 JURIMETRICS J. 91, 94, 98 (2002) (indicating my arguments regarding this point in the context of fingerprint evidence). See generally ROBERTSON & VIGNAUX, *supra* note 91, § 4.4 (arguing that experts ordinarily should not give opinions on the ultimate issue or any material fact in a case not directly observed by the expert, but should rather provide testimony that communicates the strength of the evidence to which their expertise relates by providing its likelihood ratio).

¹⁸³ See generally Ronald J. Allen & Joseph Miller, *The Common Law Theory of Experts: Deference or Education?*, 87 NW. U. L. REV. 1131 (1993). This, of course, is also a matter of degree, not a categorical choice. See Allen, *supra* note 51, at 1160 n.9. Educational presentations should be preferred over ones that merely invite deference except to the extent that the nature of the expertise makes this impracticable. See generally D. Michael Risinger, *Preliminary Thoughts on a Functional Taxonomy of Expertise for the Post-Kumho World*, in FAIGMAN ET AL., *supra* note 4, §§ 2-2.2 to 2.3 (expressing preference, *ceteris paribus*, for "summarizational" expertise over "translational" expertise).

¹⁸⁴ Cf. Imwinkelried, *supra* note 100, at 44-45 (describing the more aggressive opinion, prone to a higher error rate, as unsupported by its foundation). The acid test would be presented by a case in which the more aggressive opinion comes from an unavailable expert declarant who based his opinion on non-replicable observations and whose opinion is presented as hearsay not excluded as such. Under the binary approach, an unwarranted opinion does not become warranted by the declarant's being unavailable and the test being non-replicable. Under my comparative preference approach, such hearsay would not be inadmissible by virtue of the reliability requirement of Rule 702 as adumbrated in this section. Cf. 1 MCCORMICK ON EVIDENCE, *supra* note 5, § 18 (arguing that lay opinions from unavailable hearsay declarants should be exempt from a preference for more concrete facts that would apply if the declarant testified).

properly be held (in part) to standards of the scientific community, to the extent they can be identified. This is no small problem:

A large element of politics drives many ostensibly scientific findings. The quality of much of the science that is urged on courts is, to put it mildly, weak. In fact, it is not science at all. Many claims to science are really assertions of policy wrapped in the guise of science.¹⁸⁵

In such contexts, a consideration of scientific standards of reliability is appropriate, but even here it will only rarely be a binary determination. Rather, the degree of scientific validity in fact present will matter as part of the process of comparing the probative value of the offered testimony and its tendency to mislead with that of a more modest testimonial assertion.¹⁸⁶ And this is true even if, but for such an assertion by the proponent, the standards of the scientific community are not those that the law should embrace as pertinent to the litigation. For example, even if the law in a particular context need not demand the level of certainty associated with conventional scientific standards of statistical significance (e.g., a 95% confidence level),¹⁸⁷ an expert who falsely presents to the jury as working with knowledge that does meet that standard should ordinarily not be allowed to attempt in this way to mislead the trier of fact. This, however, also explains why a more modest claim by the expert witness might rationally lead to admission of the testimony.¹⁸⁸ The point is not to exclude expertise just because it is not scientifically sound, whatever that may mean, but rather to assure that the testimony is as helpful to the trier of fact as it can be, whether the testimony is commonly regarded as scientific or not.

Attributing such significance to the expert's assertions is reasonably compatible with both *Daubert* and *Kumho Tire*.¹⁸⁹ The

¹⁸⁵ FAIGMAN, *supra* note 46, at 71; *see also* FOSTER & HUBER, *supra* note 38, at 209-24 (illustrating rhetorical strategies used by parties and their experts in an effort to misrepresent the available scientific knowledge).

¹⁸⁶ Reference to a tendency to mislead does not mean that we have moved back into the realm of jury credulity as the grounds of exclusion. The concern here is with the *attempt* to mislead. There is no need to countenance such attempts, even if they are only rarely or never successful. *See The Best Evidence Principle, supra* note 120, at 291-92.

¹⁸⁷ *See supra* note 51.

¹⁸⁸ *See* FAIGMAN, *supra* note 46, at 77.

¹⁸⁹ The idea of invoking the deferential form of the *Daubert* inquiry only in situations in which the expert asserts a scientific basis has been "in the air" at conferences, but I am unable to attribute the idea to its originator. It would appear to be a position taken in print by Professor Michael Saks. *See* Saks, *supra* note 39, at 237-38 ("The particular regime for scrutinizing 'scientific expert testimony' detailed in *Daubert* emerged because that was the type of *asserted* expertise before the Court.")

opinion in *Daubert* emphasized somewhat ambiguously that its deference to norms of scientific validity was premised on the “nature of the expertise offered” in that case,¹⁹⁰ and stated its holding only in the context in which a trial judge is “[f]aced with a proffer of expert scientific testimony.”¹⁹¹ And the concern expressed in *Kumho Tire*, that the existence of a reliability requirement not depend on the category at issue, is not seriously implicated in relying on the expert’s assertions. It is one thing to require trial judges to categorize all expertise as scientific or non-scientific, independent of the witness’s claims;¹⁹² it is another to hold the expert to the standards that the expert asserts underlies the testimony. Furthermore, a witness cannot escape all reliability scrutiny just by disclaiming any scientific or other methodological basis.¹⁹³ Regardless of assertions, such expertise would have to satisfy the reliability requirements described in the following paragraphs.

—*Testimonial Bases.* With regard to the general data, techniques, and theories brought to bear by the expert—the issues addressed by Rule 702(1) & (2)—in a very real sense the question is not about excluding anything; it is about *including* something. That is, exclusion of proffered expertise, under this prong of the analysis, would have the purpose of expressing a preference for the proponent to expand the data considered by the court, or to include an additional, more reliable technique, and so forth. Importantly, once that additional information is presented, if it is, the exclusionary principle here expressed loses its point, and the originally challenged testimony, if not inherently altered by the addition, may be admissible.¹⁹⁴ For example, if a civil plaintiff in a toxic tort case offers structure-activity and *in vitro* studies, and the court is considering excluding the plaintiff’s evidence in order to express a preference for epidemiological evidence of human teratogenicity, there would then

(emphasis supplied).

¹⁹⁰ 509 U.S. at 590 n.8.

¹⁹¹ *Id.* at 592. The ambiguity is that the Court might mean either (1) that the proponent happened to offer evidence that the Court regards as scientific, or (2) that the proponent happened to offer evidence *as* scientific.

¹⁹² See *Kumho Tire*, 526 U.S. at 151 (“We do not believe that Rule 702 creates a schematism that segregates expertise by type while mapping certain kinds of questions to certain kinds of experts.”)

¹⁹³ This was a dangerous trend in the wake of *Daubert* and before *Kumho Tire*. See *id.* at 237.

¹⁹⁴ While this may seem counterintuitive, it is actually a familiar admissibility structure. Cf. 4 JOHN HENRY WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 1190 (Chadbourn rev. 1972) (noting that preference for original under original document rule is exhausted by introduction of original, so that copy may then be admitted, even to prove contents of the document, subject only to other rules).

be no point in excluding the challenged evidence under the present analysis if the indication is that such preferred evidence will be introduced, whether by plaintiff or defendant.¹⁹⁵

When it is not clear to the court that the preferred evidence has been or will be presented without excluding that which is proffered, exclusion is appropriate on this theory only when more reliable expertise is (a) reasonably available to the proponent, and (b) not reasonably available to the opponent. As gross generalizations, such a test will place greater demands on the prosecution than on the accused in criminal cases, and it will place greater demands on powerful civil defendants than on impecunious civil plaintiffs. The question of the resources that a party can bring to bear on improving expertise is thus of crucial importance, rather than the question of the burden of proof against which that party is operating, although the two are not of course totally unrelated. I consider imprecise, therefore, maxims that would vary the level of judicial scrutiny according to the burden of proof borne by the proponent.¹⁹⁶ The latter principle also tempts us, once again, to conflate admissibility with sufficiency.

The indicated assessments of availability should not exalt form over substance. A seemingly impecunious civil plaintiff may be associated with a resourceful repeat player, such as a subrogating insurance company or a well-heeled personal injury firm. These facts should matter in the trial court's assessment. Nevertheless, it would not be implausible to create one categorical privilege against the operation of this principle, a privilege protecting the accused in a criminal case.¹⁹⁷ Even in the unusual case in which the accused is

¹⁹⁵ I find problematic, therefore, those cases that exclude plaintiff's toxicology evidence just because substantial contrary epidemiological evidence is present. See FAIGMAN ET AL., *supra* note 4, § 7-1.3.1[1], at 290. Once again, this is a sufficiency question, pure and simple, and should be handled as such. See *id.* § 8-1.6.6. To be sure, exclusion on grounds other than the reliability requirement of Rule 702, including the ground that the toxicology evidence is simply a waste of time, might be warranted, provided of course such an exclusionary rationale is not a sufficiency determination in disguise. See *id.* § 8-1.6.5. When the issue is whether to exclude an expert's opinion in order to express a preference for more reliable background information in the foundation of the opinion, then the exclusion also implicates the previous discussion of exercising control over the testimonial form and content, at least insofar as the expert might be required to opine based on the augmented information base. In many cases, this can be done adequately on cross-examination, and there is then no need to exclude the opinion originally proffered.

¹⁹⁶ "All things being equal, the higher the standard of proof applicable to the issue upon which the expertise is offered, the higher the required threshold dependability should be." Risinger, *supra* note 183, § 2-4.0, at 89.

¹⁹⁷ This would mute concerns over the unfairness of strict application of *Daubert* to evidence offered by the accused. See Christopher Slobogin, *The Structure of Expertise in*

wealthy, it may be unwise to allow courts to apply a nominally symmetrical admissibility principle to exclude relevant expertise offered by the accused, in view of the apparent tendency of the courts to apply such principles more strictly to the accused than to the prosecution.¹⁹⁸ Such a privilege would also obviate concerns over the constitutionality of excluding expertise offered by the accused.¹⁹⁹ Somewhat more debatably, but of more practical importance, the accused might be privileged not to have to show, as a condition of excluding the prosecution's expertise on this theory, that the more reliable expertise cannot reasonably be presented by the accused.

Three further refinements are worth attention here. First, constraints on the development of more reliable expertise may be non-monetary. Ethical and legal restrictions on human experimentation are obvious constraints that do more than raise the monetary costs of developing new information.²⁰⁰ The point may have surprisingly broad implications. Even established medical practices may have evolved over time or been articulated in consensus-based "practice guidelines" without being subjected to rigorous empirical testing.²⁰¹ In recent years this has helped to spawn a movement for "evidence-based medicine,"²⁰² in some respects the medical profession's version of *Daubert*.²⁰³ In contemplating an exclusionary response under the principle of demanding better evidence, courts may encounter ethical or legal restrictions on

Criminal Cases, 34 SETON HALL L. REV. 105 (2003) (arguing that the "positivist push" of *Daubert*, if taken seriously, will unfairly disadvantage the accused, in part because of resource disparities).

¹⁹⁸ See D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99 (2000) (reporting results of an empirical analysis of decisions).

¹⁹⁹ Cf. *United States v. Scheffler*, 523 U.S. 303 (1998) (rejecting, over one dissent, a constitutional argument under the Compulsory Process Clause that accused should be privileged to introduce favorable polygraph evidence in the face of an explicit exclusionary rule in the Military Rules of Evidence).

²⁰⁰ See FAIGMAN ET AL., *supra* note 4, § 7-2.3.2, at 313 (noting ethical limitations on the use of randomized controlled trials in epidemiology).

²⁰¹ See, e.g., Fiona Godlee, *Getting Evidence Into Practice*, 317 BRIT. MED. J. 6 (1998) ("[M]any of the questions that arise in daily clinical practice remain unaddressed by well designed research."); Arnold J. Rosoff, *Evidence-Based Medicine and the Law: The Courts Confront Clinical Practice Guidelines*, 26 J. HEALTH POL., POL'Y & L. 327, 328-29 (2001) (noting that even some clinical practice guidelines have not been developed using evidence-based medicine).

²⁰² See generally D.L. SACKETT ET AL., *EVIDENCE-BASED MEDICINE: HOW TO PRACTICE AND TEACH EBM* (2d ed. 2000).

²⁰³ See Daniel W. Shuman, *Expertise in Law, Medicine, and Health Care*, 26 J. HEALTH POL., POL'Y & L. 267, 287 (2001) ("Science-based medical evidence and the *Daubert* trilogy reflect unorchestrated parallel movements in medicine and law about how to assess expertise critically.").

human experimentation, including the requirement of informed consent, that limit the ability of researchers to test the efficacy of now conventional practices or the causal assumptions on which they were originally based.²⁰⁴ Considerations such as these will sometimes warrant the admission of expertise that otherwise might fail the reliability test, especially in regard to establishing the standard of care in professional malpractice.²⁰⁵ In any event, the “cost” consideration involved in making such determinations is, always in theory and often in practice, a richly textured normative consideration.

The second refinement to be noted is that greater reliability might be unavailable to a party within the context of a particular case (even considering the matter retroactively from the beginning of the litigation), yet reasonably available to that party within the context of repeated litigation of the same or similar issue. At the outer reaches of the better evidence idea, repeat players, such as the state in regard to forensic science techniques, may plausibly be considered in regard to the long run of cases, rather than based on what is reasonably available in the short run to address a particular case.²⁰⁶ In such contexts, it would be important that the prosecution, or other repeat player, be given reasonable notice of the courts’ intention to reject expertise. Such notice need not come in the form of the rejection of expertise offered in a particular case, however, for a court might admit expertise while at the same time serving warning that in the future such expertise will be rejected for want of sufficient reliability.

The third refinement once again takes account of the adversarial nature of the proceedings. Suppose that the prosecution offers

²⁰⁴ In addition to problems of informed consent, there is significant controversy over the proper restrictions on the use of placebo-controlled experiments, which would be necessary to test the efficacy of conventional practices. See, e.g., Sharona Hoffman, *The Use of Placebos in Clinical Trials: Responsible Research or Unethical Practice?*, 33 CONN. L. REV. 449, 451 (2001) (noting that the principle of beneficence militates against the inclusion of placebos in clinical trials under most if not all circumstances in which a standard therapy exists, and arguing for a compromise position significantly limiting the use of placebo-controlled studies). But see ROBERT J. LEVINE, *THE ETHICS AND REGULATION OF CLINICAL RESEARCH* 4-5 (2d ed. 1986) (noting that even established practices may be, and in appropriate cases should be, considered “unvalidated” for purposes of applying ethical and legal principles controlling experimentation, thus muting the principle of beneficence).

²⁰⁵ This, of course, assumes that courts will come around to applying the reliability requirement of *Daubert/Kumho* to conventional medical testimony. See Shuman, *supra* note 203, at 277-82 (observing that *Daubert* has been largely ignored in civil cases outside the contexts of toxic torts and products liability). It is difficult to see how jurisdictions controlled by the 2000 amendment of Rule 702 can avoid this extension much longer.

²⁰⁶ See REDMAYNE, *supra* note 5, at 126 (emphasizing the peculiar force of the best evidence principle when applied to the state in a criminal case).

questioned document examiner testimony to identify the accused as the writer of a ransom note, and assume that the court holds this testimony inadmissible because (1) error rates have not been established for this expert or even for the average comparably-trained expert, (2) tests to determine such error rates could be easily developed by or at the insistence of the state, and (3) such tests could not be practically undertaken by the accused in this case. The state proceeds without this expertise, because it cannot conduct the tests (or cause other agencies to do so) within the time frame of the present trial. Suppose, then, that the accused offers exactly the same kind of handwriting identification testimony, except of course that it is exonerative. Applying the principles as announced above, the trial judge concludes that the latter testimony should not be excluded because the better expertise (that which is accompanied by error rate information) cannot be reasonably obtained by the accused or because the accused is privileged not to have to make a showing of unavailability. This would present a problem. The accused would be allowed to present evidence of the same degree of reliability as that which the prosecution is precluded from introducing. That might seem unfair to the state, even if adequate notice of the court's concern was given to the state in previous decisions.

Is this a serious problem? At one level, the answer might be, "No." If the goal of the admissibility structure were solely to induce improvements in the quality of expertise, then one can hardly imagine a better system. Faced with the prospect that the accused will be able to provide exonerative testimony while the prosecution will be unable to present inculpatory testimony from the same discipline, the state will have a very powerful incentive—stronger than if *neither* side or *both* sides could use such evidence—to investigate further the reliability of that form of expertise in order to assure either that the prosecution could introduce such evidence or that the defense could not.²⁰⁷ However, the goal is not to improve future reliability at any cost; accuracy in the present case ought not be completely subordinated to accuracy in future cases. A balance must be struck, and I suggest a compromise that would allow the prosecution, in the hypothetical described above, to offer rebuttal evidence using expert testimony of the same type, in this case handwriting identification testimony that is not significantly less

²⁰⁷ The latter would be possible if, for example, the research showed facts rendering the defense testimony inadmissible under Rule 403 or conditionally admissible if the defense expert provides a more conservative opinion, as explained above.

reliable than that which the accused has introduced.²⁰⁸ The accused would, however, retain the option to foreclose this line of inquiry by not offering such testimony.²⁰⁹ The same rebuttal structure would be employed in civil cases when a similar asymmetry of evidentiary options occurs. Such a balance would retain significant pressure for the long-run improvement of the reliability of under-validated expertise without creating unfairness and potentially distorted interim verdicts.²¹⁰

The foregoing considerations obviously do not exhaust the kinds of issues that would need to be addressed. I have tried only to present a framework that can be developed, refined, or modified as further thought on the matter proceeds. Many of the details cannot be perfectly anticipated and would need to be addressed in the traditional common-law manner as particular issues arise.²¹¹ But the

²⁰⁸ The problem is similar to, but not the same as, the problem discussed above in connection with the previously considered principle for marshaling cognitive resources. *See supra* notes 152-68 and accompanying text. The similarity lies in the risk of negative inferences by the jury from a party's (here, the prosecution's) failure to introduce evidence the jury would expect to receive. The main differences lie in the greater need for a remedy—because here the prosecution has no other expertise on the issue—and in the fact that a remedy is available in the present context that does not so dramatically undermine the purpose of the exclusion.

²⁰⁹ The structure would parallel that used for character evidence, in which the accused has the option of opening the door to character evidence with the understanding that the prosecution can reply in kind. *See* FED. R. EVID. 404(a)(1), (2).

²¹⁰ The degree of pressure can be adjusted, moreover, by choices about the timing of the contemplated “rebuttal.” For example, we can choose whether to allow the prosecution to offer the undervalidated handwriting expertise only at the formal rebuttal stage, which would mean the prosecution would have to survive a motion to dismiss at the end of its case-in-chief without the benefit of that testimony, or else to require the defense to indicate its intention to use its handwriting expertise and then allow the prosecution to anticipate that use by introducing its undervalidated experts in its case-in-chief. I am inclined to believe the former solution gives a better balance, but others may disagree.

²¹¹ Professor Imwinkelried sees in the resolution of these issues a multiplication of preliminary issues affecting admissibility, and he fears that this will operate to the disadvantage of financially weaker parties. *See* Imwinkelried, *supra* note 98, at 283-85. The issues are complex, to be sure, but this is not substantially more so under my proposed interpretations than under current practice. Reliability determinations, using the *Daubert* factors and similar factors articulated after *Daubert*, already are complicated matters. How could they be otherwise? My framework adds certain factors to be considered, to be sure, but if my argument—that some such proposal must be developed to make reliability standards coherent—is correct, then any marginal increase in complexity is in the nature of the beast; conceptually incoherent reliability determinations certainly cannot be justified on the ground that they are simpler to state. If the costs of complexity are not worth the gains from a coherent reliability inquiry, then the answer lies in rejecting the reliability requirement itself.

underlying theme should be clear enough: courts should not exclude expert testimony (under Rule 702) just because it is “unreliable” *simpliciter*, but only because (and when) by doing so, the reliability of the expert testimony, in this or future cases, can be noticeably improved.²¹²

IV. CONCLUSION

The general thrust of *Daubert*, that the judiciary should take a more active role in monitoring the use of expertise in court, may be entirely sound. If so, however, it might have been better had the law of expert testimony not taken the path it has. Perhaps the demand for reliability, as a condition of admissibility, is a conceptually poor substitute for judges taking more seriously their responsibility to grant directed verdicts and summary judgments when expert testimony, taken in conjunction with non-expert evidence relating to the same essential element of a cause of action or defense, is too weak to support a rational verdict favorable to its proponent.²¹³ Personally, I am inclined to think so, at least so long as the burden of production is rightly understood.²¹⁴

²¹² This, it will be noticed, is a *necessary* condition for exclusion, not a *sufficient* condition. It might be the case, for example, that excluding expertise offered by a party without significant economic or political power, such as a criminal defendant or impecunious civil plaintiff, would cause the community of experts, as repeat players themselves, to test their methods so as to improve their degree of reliability in a cost justified manner. Professor Faigman seems to believe that this would be enough to warrant exclusion, without regard to the resources of the proponent. See Faigman, *supra* note 68, at 261-64. I am indebted to him for pressing me on this point and thus requiring me to articulate the reason for my focus on party resources rather than expert-community resources. The reason is this: To use the present litigant solely as a means to the end of improving expertise offered in subsequent trials involving other parties is unfair, a violation of a categorical imperative never to use others solely as means to an end. For the litigant who must take what he or she finds in terms of the reliability of available expertise, Rule 702 should allow the litigant to use the best of what is available. If such use of available expertise is considered problematic nonetheless, it would seem more properly to be the subject of direct regulation of the expert community, which would invoke a different justificatory framework with its own set of issues. On the other hand, to the extent that a litigant bears some responsibility for the current level of reliability of the challenged expertise—a responsibility arising out of access to or creation of data, investigative economies of scale, or political power—it is fair to deprive the litigant of expertise that does not attain the degree of reliability that those resources make possible and desirable, given the nature of the case.

²¹³ Professor Friedman argues forcefully along these lines. See Friedman, *supra* note 89.

²¹⁴ See Nance, *supra* note 121 (arguing that the burden of production should be understood to embrace an obligation to assure that evidence is reasonably complete).

Be that as it may, *Daubert*, *Kumho Tire*, and amended Rule 702 have marked out a different, rockier conceptual path, and at least for the time being, we must try to make the most of it. The use of a dichotomous concept of reliability, whether or not founded on deference to a non-legal discipline, simply poses too many problems. These problems, in turn, encourage an epistemically invasive use of admissibility to monitor the weight of the evidence. It is time to turn away from that kind of thinking and try another tack, one that draws on more manageable comparative reliability assessments that build on a gradational concept of reliability.

In this paper I have argued that if we set aside the generally unfounded concerns about juror credulity, we can still fashion criteria of admissibility under Rule 702 that reflect the gatekeeping mandate of the post-*Daubert* regime while employing such a comparative assessment methodology. These criteria could assist the trier of fact by (1) marshaling its cognitive resources around the relatively reliable forms of expertise each party offers, and/or (2) providing appropriate incentives for parties to discover, develop, or present more reliable expertise than would otherwise be presented in court. While the case for the former is still relatively weak and undeveloped, the case for the latter is stronger. Careful use of the latter, if not also the former, could effect a change for the better, both in terms of the substance of the admissibility rulings and in terms of the consistency and coherence of those rulings.