

Squeezing *Daubert* Out of the Picture

*Richard D. Friedman**

INTRODUCTION

In this essay, I will offer some thoughts on how we might reframe the issues governing the admissibility of expert evidence. My principal focus is not on any particular type of expert evidence but on broader questions: the extent to which we ought to rely on rulings of admissibility, the standards that should govern admissibility rulings, and the role of the trial and appellate courts in making those rulings. To some extent, I will concentrate on the context of criminal cases, but for the most part my conclusions apply in both civil and criminal litigation.

Here are my conclusions:

First, the standards for treatment of expert evidence should differ depending on the litigation context. Standards should be very lenient for criminal defendants, and tougher for prosecutors, with the standards for civil litigants somewhere in between.

Second, more than has been the case, the test that carries the bite with respect to expert evidence should be one of sufficiency rather than of admissibility.

Third, the model created by *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹ and *General Electric Company v. Joiner*,² and extended by *Kumho Tire Company, Ltd. v. Carmichael*,³ in which trial courts, reviewed under an abuse of discretion standard, act as gatekeepers to prevent jurors from being bamboozled by unreliable evidence, is not a useful one. Reliability is an inappropriate, misleading standard for testing the admissibility of expert evidence. It does not reflect the way we should think about admissibility issues,

* Ralph W. Aigler Professor of Law, University of Michigan Law School, Ann Arbor, Michigan 48109; rdfriedman@umich.edu; (734) 647-1078. Many thanks to Sam Gross, Michael Risinger, and Michael Saks for educating me, as best they were able, and no fault of theirs to the extent they were not.

¹ 509 U.S. 579 (1993).

² 522 U.S. 136 (1997).

³ 526 U.S. 137 (1999).

and it ought to be discarded. Moreover, trial courts are not well positioned to sort the good from the bad. Where it is appropriate to exclude evidence, the decision usually should be made as a matter of law. This is because the difficulties that are most likely to warrant exclusion usually share either or both of these characteristics: (i) they are recurrent matters (including the behavior of repeat players) for which a consistent resolution across cases is appropriate; or (ii) they involve errors, particularly inferential errors, in the way the evidence is presented that make it affirmatively misleading.

Fourth, in some settings, as a less restrictive alternative to exclusion, courts should admit expert evidence but explain to the jury factors limiting the weight that the jury should accord the evidence.

I believe that we should be considerably more generous than we have been with respect to expert evidence offered by the defense, and the net effect of my suggestions would probably be greater generosity with respect to expert evidence offered in civil cases. With respect to the principal focus of this article, expert evidence offered by a prosecution, my overall aim is not to argue that we should be more or less generous than we have been. But I do believe we need to restructure the way we think about the admissibility of such evidence. Sometimes the courts should exclude such evidence as a matter of law because it is affirmatively misleading, or because exclusion will induce the presentation of better evidence. Sometimes the trial court should admit the evidence but comment adversely on it. Sometimes the court should deem the evidence admissible but nevertheless hold that the evidence taken as a whole does not support a verdict. But if none of these principles apply and the expert evidence has significant probative value the court should not exclude it on the ground that it is unreliable, and rarely if at all should the court exclude it on the mere ground that the jury is likely to over-value it. Thus, I am suggesting that *Daubert* be squeezed out of the picture by other approaches to the problem.

I. DIFFERENTIAL STANDARDS

I am one of several participants in this symposium contending that the standards for treatment of expert evidence in criminal cases ought to be different from those in civil cases. Furthermore, I believe the standards in criminal cases for evidence offered by an accused should differ from those applicable to evidence offered by a prosecutor.

Although the rules of evidence are, for the most part, the same

for criminal and civil cases, at least on their face, there are numerous settings in which the rules apply differently depending on whether the case is a civil or a criminal one.⁴ In some of the most significant of these settings, the application of the rules in a criminal case depends on whether the offering party is the accused or the prosecutor.⁵ I do not necessarily mean to defend each of these rules or the distinctions they draw; some are surely good, others may not be. But they do suggest rather strongly that there is no principle that demands uniform application of evidentiary rules across types of cases or symmetry within criminal litigation. Indeed, asymmetries are rife throughout our criminal justice system.⁶

Moreover, the argument for a greater degree of differentiation in standards seems powerful. Part of the reason for distinguishing between civil and criminal evidentiary rules, and between the rules governing prosecutors and defendants, is that, as Paul Giannelli emphasizes in his contribution to this symposium, criminal and civil procedures are so different, and the position of a prosecutor is so different from that of an accused, in such crucial matters as discovery and resources of the parties. Part of the reason, which I shall discuss in more detail later, is the “best evidence” concern arising from the “repeat player” phenomenon. Prosecutors, and the experts whose evidence they present, are the classic repeat players, and exclusion of evidence may sometimes be warranted to induce them to present better evidence; this factor is less likely to be strong in the case of evidence with respect to civil litigants or criminal defendants.⁷ And part of the reason, the part that I regard as most important, is a matter of decision theory, the different consequences of various decisions.

I take it as common ground that the very high standard of persuasion in criminal cases—“beyond a reasonable doubt”—is based

⁴ FED. R. EVID. 301; 404(a); 412; 609; 803(8)(B) & (C); 804(b)(3).

⁵ FED. R. EVID. 404(a); 609; 803(8)(C); 804(b)(3). Professor Park argues that, for the most part, these asymmetries are based on the fact that evidence offered by the prosecutor may have less probative value than evidence offered by the defense.

⁶ See generally H. RICHARD UVILLER, *THE TILTED PLAYING FIELD: IS CRIMINAL JUSTICE UNFAIR* (1999); Robert Laurence, *Sovereignty and the Native American Nation: The Bothersome Need for Asymmetry in any Federally Dictated Rule of Recognition for the Enforcement of Money Judgments Across Indian Reservation Boundaries*, 27 CONN. L. REV. 979 (1995); Richard D. Friedman, *An Asymmetrical Approach to the Problem of Peremptories?*, 28 CRIM. L. BULL. 507 (1992).

⁷ Professor Park’s comment, however, forces me to acknowledge that the differences in this respect between prosecutors and defense lawyers, and between prosecution experts and defense experts, may not be sufficiently large to bear much weight.

largely on the shared perception that the social cost of an inaccurate judgment, given that the defendant is in fact innocent, is many times greater than the social cost of an inaccurate judgment given that the defendant is in fact guilty. The “more likely than not” standard that governs most civil litigation is similarly based on the sense that an error in favor of one civil litigant is about as bad as an error in favor of the other. This difference, I have argued elsewhere, should not only guide the jurors’ decision of the case once it is presented to them, but should also play out in the doctrines governing evidentiary and other procedural decisions: If we have a choice between two rules regimes, and the first yields more accurate pro-prosecution verdicts than the second, but at the price of a greater number of *inaccurate* pro-prosecution verdicts, we should be very wary. A rule that is pro-prosecution—say, admitting prosecution evidence or excluding defense evidence—is preferable to the alternative rule only if the increase that it causes in accurate pro-prosecution verdicts far exceeds the increase in inaccurate pro-prosecution verdicts.⁸

The arguments presented by Professors Park and Lillquist have forced me to recognize that the situation is considerably more complex than I had initially believed. One can rather easily imagine facts that would justify a standard for admissibility of a given type of evidence offered by the prosecution that is no greater than the standard applicable when the same type of evidence is offered by the defense. Assume, for example, that the evidence, whether offered by the prosecution or by the defense, has substantial probative value but will be over-valued by the jury. This means that when the prosecution introduces the evidence it will cause some guilty verdicts that would not occur absent the evidence, and that among these are some that are justified by a sound view of the evidence and some that are caused by the jury’s over-valuation. Similarly, when the defense introduces the evidence, it will cause some not-guilty verdicts that would not occur otherwise, and among these there will be some that are justified by the evidence and others as to which the over-valuation is an indispensable cause. Now, assume that most of the defendants as to whom the challenged type of evidence is offered—whether by the prosecution or by the defense—are in fact guilty. Thus, even though over-valuation of prosecution evidence causes some defendants to be found guilty who would not be found guilty absent over-valuation, the fact is that most of these defendants are in fact guilty. And even though proper use of defense evidence leads to acquittal of some

⁸ Richard D. Friedman, *Truth and Its Rivals in the Law of Hearsay and Confrontation*, 49 HASTINGS L.J. 545, 548-49 (1998).

defendants who otherwise would be found guilty, the fact is that most of these defendants are guilty as well. It is not clear that a differential standard is justified.

I find considerable force to this argument—though sketching it out in this brief and informal way does not give it full credence—but for at least three reasons it does not yet persuade me. First, however true may be the assumption that most defendants are guilty, I doubt the propriety of making the choice of rules dependent on it. Arguably, doing so violates the presumption of innocence, or at least the premises underlying that presumption. But even putting that problem aside, doing so seems to create a pernicious incentive. Using the assumption of probable guilt to choose rules makes it more likely that prosecutions will be successful. Thus, use of that assumption encourages prosecutors to bring more marginal cases, in which the probability of guilt is not so high. That is a harmful effect in itself, and ultimately it undermines the very assumption of probable guilt on which the choice of a rule relatively generous to the prosecution is based.

Second, it seems to me that the prosecution's standard of persuasion may take on independent importance of its own, not fully reflected in the weighing of probabilities.⁹

Third, if we are to speculate, I believe that the array of cases more likely than not counsels in favor of greater lenience with respect to defense evidence—or at least we cannot have substantial confidence that this is not so.¹⁰

⁹ Suppose, for example, we say that the weighing of social values is such that a guilty verdict is justified only when the odds of guilt are at least as great as 20:1. Once this standard has been adopted, it becomes significant in itself. Thus, though it is clearly not so bad to find for the prosecution when the odds of guilt are 19:1 than when they are only 1:1, it may be significantly worse, not just a little worse, to find for the prosecution when the odds of guilt are 19:1 than when they are 20:1.

¹⁰ The only cases that matter are the ones in which the challenged evidence causes a change of verdict. Use Good-Pros to represent cases in which the challenged evidence is offered by the prosecution and causes a beneficial change of verdict. This change in verdict would be a good one because, even though by hypothesis the jury over-values the evidence, a jury that gave it no more than an appropriate weight would still find the evidence decisive. Use Bad-Pros to represent cases in which the prosecution offers the challenged evidence and it is the jury's over-valuation that causes a change of verdict, a detrimental effect. Similarly, use Good-Def and Bad-Def to represent the counterpart cases where the evidence is offered by the defense.

Now, if the ratio of Bad-Pros to Good-Pros cases is greater than that of Bad-Def to Good-Def cases, that would indicate that prosecution evidence should satisfy a higher standard to be admissible. And there is some reason to believe that this relationship does exist. Bad-Pros and Bad-Def cases start, before the challenged evidence is offered, farther away from the line marked by the standard of persuasion

I will not discuss these reasons in depth here. The issue is not limited to the realm of scientific and expert evidence, the subject of this symposium, and it is sufficiently complex and interesting to warrant more focused and extensive attention than is appropriate here; furthermore, to make a full presentation satisfactory to myself I need to think the matter through more thoroughly than I have been able to do thus far. For now, then, I will leave a promissory note of sorts, a statement of intent to return to the matter in a separate article, where I can give the matter fuller consideration. Meanwhile, I acknowledge the force of at least some of the arguments of Professors Lillquist and Park. Nevertheless, I adhere to the belief that the differential in the costs of different types of error means that, whatever the standard of admissibility should be, it should admit defense evidence more readily than prosecution evidence,¹¹ with the test for civil litigation—in which the social cost of an inaccurate judgment for one party is perceived to be about as bad as an inaccurate judgment for the other—lying somewhere in between.

So far as expert evidence offered by an accused is concerned, I believe the question of what test the courts should apply is a rather simple one to answer: The test should be very lenient. As I have just suggested, the overwhelming fear must be that unjustified exclusion will lead to an inaccurate guilty verdict, not that unwarranted admission will lead to an inaccurate verdict of innocence. That means that the accused should be allowed to introduce evidence even if the court is persuaded that it has very little merit. Such an approach not only makes sense as a matter of measuring the expected costs of error, but it also serves a valuable symbolic function, demonstrating that an accused who has been convicted has not been denied any plausible opportunity to leave the jury with a reasonable doubt as to his guilt. If the evidence is so clearly worthless—phrenology, perhaps—that its presentation would be a

than do Good-Pros and Good-Def cases, respectively, because it is the increment of over-valuation that carries them across that line. The standard of persuasion in a criminal case requires an extremely high probability of guilt. Distributions tend to be declining as one moves towards the extremes—suggesting that, among cases in which challenged evidence will be decisive, the proportion that is close to the line before that evidence is introduced is greater where it is the defense rather than the prosecution that is offering the evidence.

¹¹ If I am right about the importance in this realm of the difference in costs of error, I think this factor would likely overwhelm some of the concerns raised by Professor Park. He is concerned, for example, that a differential standard would distort factfinding and be hard to explain. But if the differential admissibility standard best takes into account the different costs of pro-prosecution and pro-defendant error, these concerns are likely to be secondary.

waste of time or an abuse of judicial process, exclusion would be warranted. But rarely will defendants be motivated to offer such egregious evidence, because it tends to make them look foolish and desperate. Moreover, there is no need to finance folly: Defendants should not benefit from public financing of expert evidence that the court regards as having little value.

Determining the appropriate standard for admissibility with respect to prosecution evidence or evidence offered in a civil case is, I believe, a much more difficult matter.¹²

II. SUFFICIENCY RATHER THAN ADMISSIBILITY

Daubert, *Joiner*, and *Kumho* were all decisions about admissibility, but they all involved sufficiency considerations, and perhaps they should have been explicitly about sufficiency rather than about admissibility. In each case, the trial court held the challenged evidence inadmissible and in light of that decision granted summary judgment for the defendant. I think it would have been better in each case to hold the evidence admissible and yet grant summary judgment for the defendant, as some of the courts in *Bendectin* cases have done.¹³ Perhaps in *Joiner* and *Kumho* it would have been better yet to hold the evidence admissible and *not* grant summary judgment for the defense, but I put that question aside. My main points are that what I think really motivated the courts in these cases was assessments of the evidence as a whole, that the decisions would have been better if they had been articulated in that way, that treating them as matters of admissibility tends to distort the standards for admissibility, and that it is consistent with the judicial function to grant judgment as a matter of law in cases of this sort.

I assume that in each of these three cases there was insufficient evidence, apart from the challenged expert evidence, to support the plaintiff's case. Inadmissibility therefore implied insufficiency; by holding the challenged evidence inadmissible, the courts accomplished nothing different from what they would have done if

¹² It may be, as Professor Giannelli contends, that in fact courts have been more lenient with respect to prosecution evidence than with respect to defense evidence. See Paul C. Giannelli, *The Supreme Court's "Criminal" Daubert Cases*, 33 SETON HALL L. REV. 1071 (2003). If so, that is a perverse result.

¹³ See, e.g., *Elkins v. Richardson-Merrell, Inc.*, 8 F.3d 1068, 1071 (6th Cir.1993), cert. denied, 510 U.S. 1193 (1994); Michael D. Green, *The Road Less Well Traveled (And Seen): Contemporary Lawmaking in Products Liability*, 49 DEPAUL L. REV. 377 (1999); Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation*, 86 NW. U. L. REV. 643 (1992).

they had held the challenged evidence admissible and yet held the entire body of evidence insufficient as a matter of law.

But now alter the cases somewhat. Suppose in *Daubert* there was substantial and plausible, but disputed, epidemiological evidence that Bendectin was a human teratogen. It seems clear to me that in such a case the animal and *in vitro* studies that were ultimately held inadmissible in the *Daubert* litigation should be admitted. Or suppose that in *Joiner* or in *Kumho* the testifying expert gave his testimony by deposition and then died, and that subsequently the plaintiffs were able to develop sufficient evidence, apart from the challenged testimony of the deceased experts, to support a verdict. In such a case, the admissibility decision would look very different from the way it looked in the actual case. In *Joiner*, the attitude of the courts might be not that Dr. Daniel Teitelbaum was trying to accomplish by *ipse dixit* an inferential leap that was unjustified by evidence but that his view as an experienced toxicologist deserved some consideration among the mix of evidence bearing on the question of whether PCBs promote cancer. Similarly, in *Kumho* Dennis Carlson's testimony was obviously riddled with problems, but the courts might have regarded it as worth some consideration given that there was a genuine factual dispute, that he had considerable expertise in the area of dispute, and that his assessment pointed in the same direction as a substantial body of evidence. If this analysis is right, it suggests that cases like *Daubert*, *Joiner* and *Kumho* distort the admissibility standard to achieve a sufficiency result, and that the admissibility standard that they articulate is too stringent.

But if a court holds expert testimony asserting a proposition admissible, could it then grant judgment as a matter of law against the party that bears the burden of producing evidence of that proposition? I suspect that a good deal of the resistance to greater reliance on admissibility standards stems from the perception that this would be inappropriate. After all, if a percipient witness testifies to personal knowledge of a given fact, that is sufficient evidence for the jury to find that fact to be true. It is within the province of the jury to find the witness credible, and so to conclude that the witness is accurately reporting what she observed. But I do not believe the same analysis applies with respect to an expert witness. An expert who is not testifying as a percipient witness is, by hypothesis, stating inferences that she hopes to lead the jury to adopt. It is standard fare for judgments as a matter of law to hold that the evidence does not support a given inference. I believe there is no good reason why a court should not make such a ruling with respect to expert evidence. In effect, the court would be ruling,

I thought that this expert's evidence would be of assistance to the jury, and if there were more evidence supporting her conclusions it probably would have been. But looking at this evidence together with all the other evidence that the proponent has presented, the jury could not reasonably conclude that the proponent has met its burden of persuasion. And so I am granting judgment as a matter of law.

If this approach to judgment as a matter of law appears rather aggressive, it is important to bear in mind that it is actually approved by *Daubert*. In the portion of the opinion rejecting the *Frye* test, and so tending to minimize rather than maximize the gatekeeping role of the court, *Daubert* emphasized the power of courts to grant judgment as a matter of law against a party when all the evidence considered together would not support judgment in that party's favor.¹⁴

One further advantage of this approach is that it shows greater respect to the jury. If the challenged evidence has any value, as I am arguing it did in cases like *Joiner* and *Kumho*, then ruling that it is inadmissible often amounts to a determination that, notwithstanding its value, the danger that the jurors will over-value it is so great that it is better to shut their eyes and ears to it than to let them use it for what it is worth. By contrast, a ruling of insufficiency entered before verdict says that because there is no legitimate way that the jury can reach a given conclusion there is no need to give them the opportunity. The same ruling entered after a verdict says that because the jury could have reached its conclusion only by failing to perform its function properly it must have so failed; unlike the ruling of inadmissibility, it does not presume that the jury is likely to act erroneously or incompetently, but rather concludes on the basis of the jury's actual decision that it must have done so.

I have concentrated in this part of my discussion on the recent Supreme Court cases, which are all civil ones. But it seems to me the argument for greater reliance on sufficiency rulings is, if anything, stronger with respect to prosecution evidence. Given that the prosecution's standard of persuasion is so high, there is plenty of room for a ruling that the jury could not reasonably find that the standard has been met. Often in criminal cases expert evidence is offered to prove identity. Courts ought to be willing to rule that, even though the expert evidence offered by the prosecution is admissible, the entire body of prosecution evidence does not support a finding that beyond a reasonable doubt the defendant was the perpetrator.

¹⁴ *Daubert*, 509 U.S. at 595-96.

Can we have confidence that the courts will perform very well the function of determining when the evidence taken as a whole, of which disputable forensic evidence is an important part, supports a finding of guilt beyond a reasonable doubt? Perhaps not, but if the courts do not perform this function well, what basis is there for believing that they will perform well the function of holding inadmissible evidence that will be detrimental to truth determination? There is some reason to believe that the courts can perform sufficiency determinations better than admissibility determinations (as those decisions are now made), because sufficiency is determined as a matter of law, and that means that the trial court's decision is subject to *de novo* review. I will speak further about the advantages of such review later.

Note that I have said that greater emphasis should be put on sufficiency rulings. I do not contend that such rulings should do the whole job—that is, that courts should freely admit expert evidence on the thought that if it creates a problem a sufficiency ruling will address it. Such an approach would leave wide open the possibility that, although there is enough evidence to support a verdict for the plaintiff or prosecutor, the evidence in question will be detrimental to truth determination. Thus, it is possible that, although the jury could find the accused guilty on the basis of the evidence apart from the challenged evidence, the jury would not do so, but the jury would be persuaded to convict by evidence that should not have such a strong persuasive impact. As my colleague Sam Gross has pointed out in conversation, the typical situation may be one in which there is weak witness identification evidence and vivid forensic evidence of identification that if fully understood is less persuasive than it might appear. In such a case, the witness's testimony is enough to preclude a judgment that the prosecution evidence is insufficient as a matter of law, but it may be that the jury puts little weight on that testimony. And yet, if the forensic evidence is admitted, the jury may determine on the basis of that evidence that the prosecution has proven guilt beyond a reasonable doubt. Some mechanism should be in place to minimize this possibility.

III. RELIABILITY

Daubert pronounced that “under the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.”¹⁵ This doctrine seems to fit well with

¹⁵ *Id.* at 589.

the Court's approach to hearsay and confrontation, which likewise emphasizes the reliability of evidence. In my view, it is badly mistaken.

What does reliability mean? I think a fair definition, consonant with the use of the term in everyday use, 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. But this is too rigorous a standard.

Note that reliability is not the standard generally applied for the admissibility of evidence. Eyewitness testimony, for example, is notoriously unreliable, but we do not exclude it for that reason. The purpose of a trial is to give a factfinder the opportunity to decide the facts on the basis of disparate items of evidence that might point in conflicting directions; there is no necessity that any single item of evidence be particularly reliable, or even very powerful. The 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, because that would effectively mean that the evidence could only be presented if it appears to the court that the proposition that the evidence is offered to prove is highly probable.

I must be careful not to argue against a strawman; we should consider the possibility that some softer sense of reliability that could be applied in this context states a useful standard for admissibility. I do not believe such a softer standard is useful here, but let me try two alternative meanings of the term.

One softer meaning of reliability would be particularly applicable to trace evidence. Trace evidence, in the broad sense in which I will use the term, is evidence that was, at least on the proponent's account, left by the events at issue. Most forensic evidence in criminal trials consists of trace evidence. One could say that trace evidence is a reliable indicator of a phenomenon if its likelihood ratio is very high—that is, if the evidence is many times more likely to arise given the phenomenon than given the absence of the phenomenon. This is a softer definition because if the prior probability of a phenomenon (that is, the probability of the phenomenon as assessed absent the evidence in question) is very low, then the posterior probability (the probability of the proposition given the evidence) may not be very high even though the evidence has a high likelihood ratio. But even this definition is too rigorous as a standard for admissibility.

Consider microscopic hair comparison. This may be a waning

type of evidence, because if microscopic comparison yields a result mitochondrial DNA analysis usually (though not inevitably) will as well, and the mtDNA analysis is generally considered to be more accurate. But right now we are in a time window in which evidence based on microscopic comparison is still often being offered and its accuracy can be tested as against that of mtDNA. In a recent study based on 170 examinations of questioned hairs and known samples, examiners reached conclusions (either “association” or “exclusion”) by both methods of analysis in 95 cases. Those cases broke down this way:¹⁶

TABLE 1

MtDNA		
	Association	Exclusion
Association	69	9
Microscopic Comparison		
Exclusion	0	17

Let us assume that the mtDNA analyses were all accurate, and further that these data indicate the probabilities of finding an association by microscopic comparison given that the hairs are, or are not, actually from the same source.¹⁷ Then the probability that if the hairs are actually from the same source a microscopic comparison will so indicate is very close to one hundred percent (69/69). The probability that if the hairs are *not* from the same source the comparison will indicate that they are is much lower, but still quite substantial, 34.6 percent (9/26).

Now suppose that a prosecutor introduces evidence of a positive result yielded by microscopic hair comparison. That does not appear to me to satisfy even the alternative definition of reliability that I have

¹⁶ Max M. Houck & Bruce Budowle, *Correlation of Microscopic and Mitochondrial DNA Hair Comparisons*, 47 J. FORENSIC SCI. 964, 966 (2002). In four cases, either the hair was insufficient for mtDNA analysis or the analysis was inconclusive and yet microscopic comparison yielded a result, in two cases association and in two exclusion. Far more frequently, the mtDNA test yielded a result, but either the microscopic hair comparison was inconclusive or upon inspection the questioned hair was not deemed suitable for microscopic examination.

¹⁷ These are simplifying assumptions. Thus, the probability of finding an association by comparison given that the hairs are, or are not, from the same source may depend on the context in which the comparison is made; the comparisons in this study involved not hairs selected at random but questioned hairs involved in actual investigations.

given based on the likelihood ratio.¹⁸ That definition focuses our attention on the rate of false positives, and it seems apparent that, at least for evidentiary purposes, a test that yields a positive result in nearly thirty-five percent of the cases in which the facts are negative cannot usefully be described as reliable. This is especially so if the factfinder begins (as the presumption of innocence requires before the introduction of any evidence¹⁹) with prior odds strongly indicating that the facts *are* negative.

And yet this evidence *if properly explained* has substantial probative value. The likelihood ratio of a positive result by microscopic comparison is 2.89 ($100 \div 9/26$). Assuming it is independent of any other evidence, this evidence should cause the factfinder to increase quite substantially the odds the factfinder attributes to the proposition that the two hairs came from a common source. Indeed, the evidence should cause the factfinder to multiply those odds by 2.89.²⁰ That is more powerful evidence than proof that both the defendant and the perpetrator had blood type A, which has a likelihood ratio of about 2.5 and would be admissible.²¹ Even evidence that has a much lower impact on the probability of a material proposition may have ample probative value to warrant admissibility.

I do not believe it much matters whether we characterize the reason for so many false positives as inherent unreliability of microscopic comparisons or imperfect performance by those who perform it, or some combination of the two. The bottom line is that microscopic comparison is very unreliable, even as measured by the likelihood ratio, and yet if properly understood it can be very useful.

A final definition of reliability would equate it with consistency. Under this standard, which is the way the term is used by scientists, a test is reliable to the extent it generates the same results on repeated

¹⁸ A negative result, however, might be considered reliable. *See infra* note 20.

¹⁹ *See* Richard D. Friedman, *A Presumption of Innocence, Not of Even Odds*, 52 STAN. L. REV. 873 (2000).

²⁰ The FBI apparently finds microscopic comparison useful as an investigatory device. Though Houck and Budowle decline to call microscopic comparison a "screening" test, *id.* at 967, it appears that the FBI uses it in this way, because "very few hairs that are microscopically excluded in typical casework are submitted for mtDNA analysis." *Id.* at 965. The data from their study indicate that this is a rational decision: of the 17 pairings for which both analyses yielded a result and the result of microscopic comparison was exclusion, in all 17 mtDNA analysis also indicated exclusion.

²¹ *See* *People v. Mountain*, 495 N.Y.S.2d 944, 946-48 (1985) (overruling a prior case that held that proof that the accused and the perpetrator had Type A blood, a common type, is inadmissible because lacking probative value).

runs under similar circumstances.²² But this definition hardly seems useful as a guidepost to admissibility of evidence. Again, consider the hair comparison data. The results of hair comparison do not, on their face, appear to be consistent: When hairs not from a common source are compared, a conclusion of no association is reached about sixty-five percent of the time and a conclusion of association is reached about thirty-five percent of the time. On its face, that is hardly consistent. Conceivably, there is some latent explanation that accounts for the apparent inconsistency. For example, suppose that an omniscient observer recognizes that microscopic hair comparison indicates that the two hairs are from the same source if and only if the two hairs are (a) indeed from the same source, or (b) both from males over 20 years old. Then the omniscient observer could say that the comparison yields consistent results. But this consistency would be of no avail to us if the explanation was not apparent to us; given a positive result from the comparison, we would have no way of knowing whether it was because the hairs were from a common source or for some other unknown factor that tended to yield that result.

One could devise other meanings of reliability. But if the term needs to be stretched out of shape to yield a useful standard it obstructs rather than advances analysis. I therefore regard the reliability standard enunciated by *Daubert* as inappropriate, and the incorporation of that standard in the 2000 amendment to Federal Rule of Evidence (FRE) 702—requiring expert testimony to be “the product of reliable principles and methods,” which the expert “has applied . . . reliably to the facts of the case”—as unfortunate. The better standard is the one that originally stood by itself in FRE 702 and still shares the Rule with the reliability standard—whether the specialized knowledge in question “will assist the trier of fact to understand the evidence or to determine a fact in issue.” *If properly explained*, it appears that microscopic hair comparison evidence satisfies this standard, and that should suffice for admissibility.

²² See, e.g., David Medoff, *The Scientific Basis of Psychological Testing: Considerations Following Daubert, Kumho, and Joiner*, 41 FAM. CT. REV. 199, 200 (2003); D. Michael Risinger & Jeffrey L. Loop, *Three Card Monte, Monty Hall, Modus Operandi and “Offender Profiling”: Some Lessons of Modern Cognitive Science for the Law of Evidence*, 24 CARDOZO L. REV. 193, 245 (2002) (equating reliability, as a technical term of art, to consistency).

IV. PREJUDICE

If the evidence creates an intolerable chance that it will cause the factfinder to reach an inaccurate pro-prosecution conclusion, then it should be excluded. One may reach this conclusion under the basic decision-theoretic approach I have outlined above. It is also suggested by the general standard of FRE 403, that if “the danger of unfair prejudice, confusion of the issues, or misleading the jury,” among other costs, substantially outweighs the probative value of the evidence, it ought to be excluded. And the conclusion is also indicated by the original standard of Rule 702, which I have just quoted; if it is too likely that the challenged expert evidence will lead the trier of fact astray, then the court should not deem that the challenged evidence “will assist the trier of fact to understand the [other] evidence or to determine a fact in issue.”

I am sure that in some cases bad expert evidence does lead juries to inaccurate pro-prosecution results. I do not believe, however, that we have a good gauge on how rampant the problem is. Barry Scheck, Peter Neufeld, and Jim Dwyer, in a popularly oriented book, *Actual Innocence*, present some interesting data gathered by the Innocence Project purporting to show factors leading to convictions that have been established as wrongful.²³

TABLE 2

Data from Scheck, Neufeld & Dwyer: Factors Leading to Wrongful Convictions in 62 U.S. Cases

Mistaken ID	52
Serology Inclusion	32
Police Misconduct	31
Prosecutorial Misconduct	26
Defective or Fraudulent Science	21
Microscopic Hair Comparison	18
Bad Lawyering	17
False Witness Testimony	15
Informants/Snitches	15
False Confessions	15
Other Forensic Inclusions	5
DNA Inclusions	1

²³ BARRY SCHECK ET AL., *ACTUAL INNOCENCE: FIVE DAYS TO EXECUTION AND OTHER DISPATCHES FROM THE WRONGLY CONVICTED* 263 (2000).

The data have been updated (with the addition of new cases from the Innocence Project's files) and reinterpreted (by combination of some of the categories) by Michael Saks and his colleagues:²⁴

TABLE 3
Factors Leading to Wrongful Conviction in 81 Cases

Erroneous Forensic Science	53
Fraudulent, Tainted Evidence	25
Mistaken Eyewitness Identification	60
Police Misconduct	26
Prosecutorial Misconduct	32
Bad Lawyering	23
False Confessions	15
Informants/Snitches	14
False Witness Testimony	14

It is difficult to know just how much to make of the data. I have the greatest respect for the principal authors of both studies—Barry Scheck, for example, is an old friend and a former colleague. But *Actual Innocence* is a book of advocacy written by advocates, and it presents the data in very conclusory form. Even assuming that a given factor was present in a particular case, which may not always be self-evident (what are the standards for determining “defective” forensic science?), how do we know that the factor was one “leading to” a wrongful conviction? How do we know, for example, that the jury did not discount forensic science evidence that these writers later describe as erroneous? And even if the factor did play a role in jury decision-making, do we know whether it was in fact decisive, in the but-for sense? I am not saying that these matters were not determined, or that they are not determinable, just that the methods are not made transparent.

Beyond this, how bad a problem do the data suggest expert evidence presents? In effect, we have only an imperfect indication of the numerator of a fraction, and nothing about the denominator. At worst, the data reveal a few dozen terrible outcomes attributable to forensic science. Of course, this may be the bare tip of the iceberg,

²⁴ Michael J. Saks et al., *Toward a Model Act for the Prevention and Remedy of Erroneous Convictions*, 35 NEW ENG. L. REV. 669 (2001).

and there may be many other such outcomes about which we are still ignorant—but we really do not have a good basis for knowing how many there are. Moreover, the data tell us nothing about how many accurate pro-prosecution outcomes such evidence caused that might not have been achieved but for such evidence. Note, for example, that according to Scheck, Neufeld, and Dwyer, mistaken identification is the leading cause of wrongful convictions—but we would never seriously consider a rule broadly excluding identification testimony.

Nevertheless, let us assume that the data indicate that forensic science evidence poses a serious problem that requires some attention. That is a plausible conclusion, and the re-combination of categories by Saks et al. reinforces it. At the same time, such an aggregation of data should not cause us to be too reductionist. There are problems of different types, and they may warrant different responses.

For example, consider “serology inclusion.” Scheck, Neufeld, and Dwyer explain that this “refers to ABO and protein blood typing of semen, saliva, and bloodstains.”²⁵ These are well-established techniques, and I do not suppose anybody poses a serious objection to their use; this is not a matter of “defective or fraudulent” science or of junk science of any sort. So then what is the problem? If all that happened is that the serology results strengthened the prosecution case by indicating that the defendant was within a limited category of persons who could have left an incriminating stain, then perhaps there is no problem, or at least we cannot say that anybody on the prosecution side did anything wrong. A rational factfinder fully understanding the situation would conclude that this evidence increased the probability that the defendant was guilty, even though given fuller information the factfinder would conclude that the defendant was not in fact guilty. Perhaps, though, the problem is that the expert witness over-claimed the significance of the serological result. If that is the problem, then the solution would not be to exclude evidence of the underlying serological phenomenon, but to try to prevent the over-claiming, or at least to ensure that the jury is not misled by it. It does not seem to me that the *Daubert* framework is of much help here.

The same analysis applies to microscopic hair comparison.²⁶ I

²⁵ SCHECK ET AL., *supra* note 23, at 263.

²⁶ It also applies to at least some of the “Other Forensic Inclusions,” a category that, according to Scheck, Neufeld, and Dwyer, “refers to the comparisons of fingerprints, fibers, and other physical evidence,” and is significantly less

have argued above that microscopic hair comparison does appear to have significant probative value. There does not seem to be any warrant for a general rule keeping such evidence from the jury. If there is a problem it is presumably that the proponent of the evidence claims that the evidence is more probative than it is. Ideally, then, the problem can be prevented by avoiding such over-claims—not by excluding evidence of the comparison.

I assume that if jurors are given misinformation they are prone to reach inaccurate results because of it. But I do not believe that we have any good basis for concluding that if evidence such as microscopic hair comparison—evidence that is imperfect but that nevertheless has significant probative value—is accurately described jurors will nevertheless so likely over-value it that it should be deemed more prejudicial than probative. And even if jurors are likely to over-value some evidence to that extent, I wonder how we can have confidence that the courts will be able to sort out cases where this is likely to happen from cases in which it is not. Furthermore, as I will now argue, even if courts can sort out such cases, exclusion is not necessarily the most appropriate remedy.

V. COMMENT AS A LESS RESTRICTIVE ALTERNATIVE

The primary remedies that our system offers for evidence that might otherwise over-persuade the factfinder are responses by the adversary—cross-examination, rebuttal, and argument. Assume that a court believes that a given piece of expert evidence has substantial probative value but that there is too great a chance that, even after taking these responses into account, the evidence will over-persuade the jury. Exclusion is not the inevitable remedy. Judicial comment, expressing reasons to limit the significance of the evidence, seems more appropriate. Coming from the court, such comment will not likely be ignored. And yet it allows the jury to give the evidence some

troublesome. *Id.*

Fingerprint identification, by the way, seems to be a trivial source of mistaken identification. Scheck, Neufeld and Dwyer seem not to have any particular difficulty with it; they compare it favorably to hair identification, pp. 161-62, and in their extensive series of recommendations as to how to prevent erroneous convictions, pp. 254-60, they make no suggestions as to fingerprints. I am inclined to believe that *United States v. Lera Plaza*, 2002 WL 27305 (E.D. Pa. Jan. 7, 2002), *vacated*, 188 F. Supp. 2d 549 (E.D. Pa. 2002), addressed a problem that is not pressing in reality. If the treatment of fingerprint evidence does not square with *Daubert*, as some, including my collaborators in a recent exchange, have suggested, Richard Friedman et al., *Expert Testimony on Fingerprints: An Internet Exchange*, 43 JURIMETRICS J. 91, 92 n.9 (2002), I believe the problem lies with *Daubert* rather than with the way fingerprint evidence has been treated for decades.

weight.

Courts traditionally have the power to comment on the evidence. And yet they are hesitant to use the power, for fear of encroaching on the jury's function. But surely it encroaches less on the jurors' function to admit evidence that has some probative value and give them reasons why they should not put too much weight on it than to exclude the evidence altogether because they are too likely to over-value it.

Comment rather than exclusion makes sense only if the evidence has significant probative value. The availability of the comment option leaves less room in which exclusion is the optimal solution. In some cases, though, exclusion of at least part of a proffer of expert evidence is warranted; perhaps the underlying phenomenon has too little probative value to justify admissibility, perhaps the expert's assessment of it is misleading, or perhaps the exclusion of this evidence will induce the production of better evidence. The question then arises as to the responsibilities of trial and appellate courts in making the decision to exclude.

VI. THE ALLOCATION OF DECISION-MAKING AUTHORITY

In *Joiner* the Supreme Court held that the job of applying the *Daubert* standard lies primarily with the trial court; its decisions are to be reviewed for abuse of discretion. I believe that the role of appellate courts should be greater in this realm than *Joiner* suggests. I have already suggested one reason why: To a large extent, weak expert evidence should be deemed to create a sufficiency problem rather than an admissibility problem, and sufficiency issues are treated as matters of law. Here, I will go further: When expert evidence is to be held inadmissible, it should not usually be because of a discretionary judgment by the trial court that the evidence is more likely prejudicial than probative, but rather because admission would be deemed erroneous as a matter of law. The trial court should make the initial decision as to whether exclusion is appropriate, of course. But that decision should often be reviewed *de novo* by the appellate court.

I have already suggested an argument on one side of the coin: I do not believe the type of weighing that we are used to leaving to trial courts works well in this context. If expert evidence is not affirmatively misleading, it may well have significant probative value, and in the absence of a demonstrated reason to believe that the jury will substantially over-value it there is probably no basis on which to conclude that the evidence is more likely to hurt than to help truth

determination.

Now, if we consider factors that sometimes *do* justify the exclusion of expert evidence, I think the best conclusion is that the determination should most often be made as a matter of law.

First, if the admissibility of evidence depends on the truth of a proposition that is asserted to be scientific, then by definition that proposition is one that has applicability beyond the bounds of the case at hand, and probably it is one that will recur from case to case. There is no good reason why such a proposition of scientific fact should be deemed to be true in one case and not true in another case. The proposition appears to be a matter of legislative fact or perhaps, in the language of Professors Monahan and Walker, a social framework;²⁷ it is not in any event an adjudicative fact the significance of which is confined to the case at hand. A ruling that runs across cases seems appropriate.

Second, such recurrent matters tend to involve repeat players. Indeed, with respect to prosecution evidence there is, almost by definition, one significant repeat player—the prosecutor herself, the party offering the evidence. Beyond this, other repeat players, laboratories and testifying forensic experts, are almost always involved in the creation of prosecution forensic evidence. The prominence of repeat players makes it especially appropriate to apply best evidence rules—that is, rules excluding evidence even if it would be more probative than prejudicial because exclusion would likely induce the creation of better evidence. Sometimes a best evidence rule is focused largely on the case at hand: If unauthenticated evidence is inadmissible, the proponent will probably be induced to present authenticating evidence, and if a leading question is held improper the proponent will probably ask non-leading questions. The involvement of repeat players makes it especially appropriate to apply best evidence rules for broader reasons going beyond an individual case: If the proffered evidence is excluded on the grounds that the procedures in creating it were suboptimal, or that the proponent could with sufficient effort make a showing of such matters as proficiency, so that the factfinder could better assess the probative value of the evidence, then better evidence is likely to be presented in future cases. Such a rule of exclusion is based on a significant policy determination, and it should be decided as a matter of law rather than of discretionary judgment by the trial court.

Third, often the problem with scientifically-based evidence is not

²⁷ See, e.g., Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559 (1987).

in the phenomenon but in the explanation. That is, the phenomenon in question may indeed be an indicator, more or less strong, of the proposition in question, but the difficulty occurs if the testifying expert makes a misleading statement about what the evidence shows. As discussed above, this appears to be the principal problem with respect to evidence of microscopic hair comparison, and it appears as well to be a significant problem with respect to serological evidence. Moreover, it is a recurrent and serious problem with respect to DNA evidence, for prosecution witnesses often commit the “prosecutor’s fallacy” in describing the impact of the evidence.²⁸ This fallacy arises when the accused’s DNA matches that of a crime scene sample and there is a small probability that such a match would have arisen had the accused not been the source of that sample. The fallacy consists of describing this evidence as if it demonstrates that there is a small probability that the accused was not the source of that sample. The witness has effectively treated the probability of the evidence given the negation of the prosecutor’s hypothesis as equal to the probability of that negation given the evidence—but the two are not the same and may be different by a very large order of magnitude.

A witness should be barred from committing such fallacies in describing scientifically-based evidence. She should be barred on a ground more definite than that the evidence is unreliable, or that it is more prejudicial than probative, or that it is not helpful to the jury, and she should be barred by a firmer standard than the discretion of the trial court. She should be barred because her testimony is demonstrably *wrong*, as a matter of logic—and therefore that testimony, as compared to testimony that does not suffer from the same illogic, can only hurt the truth-determination process. In conjunction with this demonstrable wrongness, several other factors warrant exclusion as a matter of law:

(a) The testimony is wrong on a basis that transcends the case at hand.

(b) The witness is likely to be a repeat player—and even if she is not, other witnesses who are repeat players will heed a ruling of exclusion and learn to conform their testimony to its demands.

(c) Indeed, as rulings come from high courts all the professional players in the system are likely to pay attention and be aware of the

²⁸ See, e.g., William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor’s Fallacy and the Defense Attorney’s Fallacy*, 11 L. & HUM. BEHAVIOR 167 (1987); Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence at Trial*, 34 JURIMETRICS J. 21 (1993).

problem so that it is less likely to arise, and more likely to be corrected, in future cases. Thus, even if the prosecutor does not, judges and defense lawyers are likely to ensure that the testifying witness adheres to prescribed standards.

(d) Because the problem is a matter of logic, and the logic of inference, it is likely that courts are at least as well situated as the testifying expert to detect and understand the problem. Typically, the testifying expert is not someone with a strong understanding of probability theory. I do not claim that judges usually have such a strong understanding, either. But logic should be something judges are good at, at least if they put their minds to it, and they are motivated to do a good job in monitoring the inferential process.

Finally, I believe rulings that will durably appear to be in accordance with scientific truth will more likely result from procedures in which appellate review is *de novo* rather than for abuse of discretion. Matters are more likely to be briefed with greater care in the appellate courts, especially because the incentives are greater if a decision will determine the rule for a whole class of cases. Appellate judges tend to have greater time for careful decision-making. The participation of a panel of judges—and ultimately of the jurisdictions' court of last resort—increases the chance that the decision will be infused by good understanding of the situation. If more than one trial court has gathered information on the matter, the appellate court will be in a good position to draw on it.

The one significant advantage of trial procedure in this area, as matters now stand in the United States, is that trial judges, unlike appellate judges, are in a position to have an oral interchange with experts, which might increase their understanding. But this, it seems to me, is a problem easily enough corrected. Experts occasionally appear before British appellate courts, and there is no reason in principle why this should not happen in American appellate courts. Indeed, I cannot find anything in the Federal Rules of Appellate Procedure that would bar the practice; I leave to the side whether the expert should testify under oath, as in the trial court, or argue like an appellate attorney, or merely be present to answer questions of the court. In most settings, there is no reason for a witness to testify before the appellate court, or for anybody but the attorneys to participate in argument, because the appellate court is not the province for factfinding. But here our concern is legislative facts, or matters that should be treated as the equivalents of legislative facts. This means that appellate courts are the ultimate factfinders. I do not see any reason that they should be limited to what they can glean from the trial court record, the submissions of the parties, and their

own independent research. If they have questions of scientific fact, why should these questions not be openly explored before them?

Is there a danger that greater reliance in this area on lawmaking by the appellate courts will be top-heavy and inefficient? I do not believe so. Because we are talking about recurrent decisions, it is more efficient to have a uniform rule, so that trial courts do not have to go through the exercise each time of exercising their discretion on a given point.

Is there, then, a danger that appellate decision-making will be too ossified, too unlikely to change in response to changing scientific knowledge? There should not be. If an appellate court rules a given way on an issue of admissibility based on its perception of scientific fact, new evidence as to scientific truth would create a new situation. Presumably appellate courts would be reluctant to alter the impact of recent decisions too rapidly, because that would undercut the purpose of appellate decision-making. But an appellate court provided with fuller or more updated information than was available to a prior court can reach a different result without overruling the prior court. There is no inconsistency if a factfinder reaches different results in two cases that on the facts are largely similar but different in some material respect. And the same is true with respect to decisions by a court in cases in which the legislative facts bear some material dissimilarity.

CONCLUSION

I have suggested in this article that we restructure the way we think about the admissibility of expert evidence. I have focused largely on criminal cases, but my conclusions are more general.

I have suggested that *Daubert* should be squeezed, from several directions.

In some cases, the courts should simply admit the evidence, notwithstanding doubts about its reliability, and let the jurors do with it as they will; that the jury might reach different conclusions from the court is the whole reason why we have a jury. This is especially so with respect to defense evidence.

In some cases, even though the court admits the evidence it might decide to make a comment expressing reasons why the evidence should not be given too much weight. Such a comment is less restrictive than exclusion and generally preferable if the evidence has substantial probative value.

In some cases, the court should decide that the challenged evidence is admissible and yet conclude that as a matter of law the

evidence taken as a whole does not satisfy the burden of proof imposed upon the prosecutor (or, in a civil case, upon the claimant).

When we have taken account of all such cases, there will still be others in which the expert evidence should be held inadmissible. But when this is appropriate it should be, at least most often, as a matter of law. If the underlying phenomenon has too little value as an indication of the proposition at issue to warrant admissibility, or the expert's explanation is affirmatively misleading, or best evidence considerations counsel exclusion to induce the production of better evidence in this or subsequent cases, then exclusion should be justified as a matter of law. And if these conclusions do not hold, then exclusion is probably not warranted.

At the end of the day, taking all these possible resolutions into account, I do not believe there is any scope, or at least not much, in which it is appropriate or necessary for the *Daubert* approach to apply. That should not be surprising. It is by no means clear how significant a problem juror over-valuation of expert evidence is, or whether courts can effectively sort out evidence that jurors are likely to over-value from evidence that they are not. So long as it is not misleadingly explained, even evidence that is quite unreliable may play a highly useful role in the truth-determination process.