

# Trial and Error: The Supreme Court's Philosophy of Science

Susan Haack, PhD

Apparently equating the question of whether expert testimony is reliable with the question of whether it is genuinely scientific, in *Daubert v Merrell Dow Pharmaceuticals, Inc* (1993) the US Supreme Court ran together Karl Popper's and Carl Hempel's incompatible philosophies of science. But there can be no criterion discriminating scientific, and hence reliable, testimony from the unscientific and unreliable; for not all, and not only, scientific evidence is reliable.

In subsequent rulings (*General Electric Co v Joiner*, 1997; *Kumho Tire Co v Carmichael*, 1999) the Court has backed quietly away from *Daubert's* confused philosophy of science, but not from federal judges' responsibilities for screening expert testimony. Efforts to educate judges scientifically, and increased use of court-appointed experts are, at best, only partial solutions to the problems with scientific testimony. (*Am J Public Health*. 2005;95:S66–S73. doi:10.2105/AJPH.2004.044529)

*It seems to me that there is a good deal of ballyhoo about scientific method. I venture to think that the people who talk most about it are the people who do least about it. . . . No working scientist, when he plans an experiment in the laboratory, asks himself whether he is being properly scientific. . . . When the scientist ventures to criticize the work of his fellow scientist, he does not base his criticism on such glittering generalities as failure to follow the "scientific method," but his criticism is specific. . . . The working scientist is always too much concerned with getting down to brass tacks to be willing to spend his time on generalities.*

Percy Bridgman<sup>1</sup>

In *Frye v United States*<sup>2</sup> (1923), the Washington, DC court upheld the exclusion of testimony of the results of a then-new blood-pressure deception test on the grounds that novel scientific testimony "crosses the line between the experimental and the demonstrable," and so is admissible, only if it is "sufficiently established to have gained general acceptance in the particular field to which it belongs."<sup>2</sup> Ignored for a decade, rarely cited for a quarter-century, over time the "*Frye* test" became increasingly influential, until by the early 1980s it had been adopted by 29 states.

In 1975, however, newly enacted Federal Rules of Evidence had set a seemingly less restrictive standard: the testimony of a qualified expert, including a scientific expert, is admissible provided it is relevant (unless it is excluded, under Rule 403, on grounds of unfair preju-

dice, waste of time, or confusing or misleading the jury). In *Barefoot v Estelle*, a 1983 constitutional case, the Supreme Court affirmed that the rights of a Texas defendant were not violated by the jury's being allowed to hear psychiatric testimony of his future dangerousness at the sentencing hearing—even though an amicus brief from the American Psychiatric Association reported that two out of three such predictions are mistaken. Writing for the majority, Justice White observed that state and federal rules of evidence "anticipate that relevant, unprivileged testimony should be admitted and its weight left to the fact-finder, who would have the benefit of cross-examination and contrary evidence by the opposing party."<sup>3</sup> Justice Blackmun wrote an angry dissent.

In 1991, amid increasing public concern that the tort system was getting out of hand, Peter Huber argued in his influential *Galileo's Revenge* that, under the Federal Rules, worthless "junk science," which would have been excluded by the *Frye* test, was flooding the courts. In 1992, proposals to tighten up the Federal Rules were before Congress. In 1993, the Supreme Court issued its ruling in *Daubert v Merrell Dow Pharmaceuticals, Inc*,<sup>4</sup> the first case in its 204-year history where the central questions concerned the admissibility of scientific testimony. The *Frye* rule arose in a criminal case and had for most of its history been cited in criminal cases; but *Daubert* was a tort action in which the trial court had relied on

*Frye* in excluding the plaintiffs' experts' testimony that the morning-sickness drug Bendectin was teratogenic. So the Supreme Court was to determine whether the Federal Rules of Evidence had superseded *Frye*, and in particular how Rule 702 was to be interpreted.

Yes, Justice Blackmun wrote for the unanimous court, the Federal Rules of Evidence had superseded *Frye*; but the Rules themselves require judges to screen proffered expert testimony not only for relevance, but also for reliability. In doing this, he continued (in a part of the ruling from which Justice Rehnquist and Justice Stevens dissented), courts must look, not to an expert's conclusions, but to his "methodology," to determine whether proffered evidence is really "scientific . . . knowledge," and hence reliable. As to what that methodology is, citing law professor Michael Green citing philosopher of science Karl Popper, and quoting an observation of Carl Hempel's for good measure, the *Daubert* ruling suggested four factors that courts might use in assessing reliability: "falsifiability," that is, whether proffered evidence "can be and has been tested"; the known or potential error rate; peer review and publication; and (in a nod to *Frye*), acceptance in the relevant community.<sup>5</sup>

In partial dissent, however, pointing out that the word "reliable" nowhere occurs in the text of Rule 702, Justice Rehnquist anticipated difficulties over whether and, if so, how *Daubert* should be applied to nonscientific expert testimony; worried aloud that federal judges were being asked to become amateur scientists; and questioned the wisdom of his colleagues' readiness to get involved in philosophy of science. I think he was right to suspect that something was seriously amiss; in fact, what I shall have to say here might be read as an exploration, amplification, and partial defense of his reservations about that philosophical excursus.

## DAUBERT'S CONFUSIONS: POPPER AND HEMPEL

Apparently equating the question of whether expert testimony is reliable with the

question of whether it is genuinely scientific, taking for granted that there is some scientific “methodology” which, faithfully followed, guarantees reliable results, and casting about for a philosophy of science to fit this demanding bill, the *Daubert* Court settled on an unstable amalgam of Popper’s and Hempel’s very different approaches—neither of which, however, is suitable to the task at hand.

Popper describes his philosophy of science as “Falsificationist,” by contrast with the Verificationism of the Logical Positivists, because his key theme is that scientific statements can never be shown conclusively to be true, but can sometimes be shown conclusively to be false. Hence, his criterion of demarcation: to be genuinely scientific, a statement must be “testable”—meaning, in Popper’s mouth, “refutable” or “falsifiable,” i.e., susceptible to evidence that could potentially show it to be false (if it is false). Curiously, Popper acknowledged from the beginning that his criterion of demarcation is a “convention”; and in his 1959 introduction to the English edition of *The Logic of Scientific Discovery*, he affirmed that scientific knowledge is continuous with commonsense knowledge.<sup>6</sup> Nevertheless, his whole philosophy of science turns on his criterion of demarcation. Falsifiability is to discriminate real empirical science, such as Einstein’s theory of relativity, from prescientific myths, from nonempirical disciplines like pure mathematics or metaphysics, from nonscientific disciplines such as history, and from such pseudosciences as Freud’s and Adler’s psychoanalytic theories and Marx’s “scientific socialism.”<sup>7</sup> Falsifiability is also central to Popper’s account of the scientific method as “conjecture and refutation”: making a bold, highly falsifiable guess, testing it as severely as possible and, if it is found to be false, giving it up and starting over rather than protecting it by ad hoc or “conventionalist” modifications. (This readiness to accept falsification and eschew ad hoc stratagems is Popper’s “methodological criterion” of the genuinely scientific.)

Popper also describes his philosophy of science as “deductivist,” by contrast with “inductivism,” whether in the strong, Baconian form that posits an inductive logic for arriving at hypotheses or in the weaker, Logical Positivist form that posits an inductive logic of confirmation. According to Popper, Hume showed

long ago that induction is unjustifiable. But science doesn’t need induction; the method of conjecture and refutation requires only deductive logic—specifically, *modus tollens*, the rule invoked when an observational result predicted by a theory fails.

Theories that have been tested but not yet falsified are “corroborated,” the degree of corroboration at a time depending on the number and severity of the tests passed. That a theory is corroborated, to however high a degree, doesn’t show that it is true, or even probable; indeed, the degree of testability of a hypothesis is *inversely* related to its degree of logical probability.<sup>8</sup> Corroboration is not a measure of verisimilitude, but at best an indicator of how the verisimilitude of a theory *appears*, relative to other theories, at a particular time<sup>9</sup>; and the fact that a theory has been corroborated doesn’t mean it is rational to believe it. (It *does* mean, Popper writes, that it is rational to prefer the theory as the basis for practical action; not, however, that there are good reasons for thinking the theory will be successful in the future—there *can be* no good reasons for believing this.<sup>10</sup> So it seems that all this “concession” amounts to is that in deciding how to act, we can do no better than go with theories that we don’t so far know to be false.)

The first problem with the *Daubert* Court’s reliance on Popper is that applying his criterion of demarcation is no trivial matter; as Justice Rehnquist pointed out, observing wryly that, since *he* didn’t really know what is meant by saying that a theory is “falsifiable,” he doubted federal judges would, either.<sup>11</sup> Indeed, Popper himself doesn’t seem quite sure how to apply his criterion. Sometimes, for example, he says that the theory of evolution is not falsifiable, and, so, is not science; at one point, he suggests that “survival of the fittest” is a tautology, or “near-tautology,” and elsewhere that evolution is really a historical theory, or perhaps metaphysics. Then he changes his mind: evolution *is* science, after all.<sup>12</sup> It’s ironic; for Popper’s criterion of demarcation had already found its way into the US legal system, a decade before *Daubert*, in a 1982 First Amendment case, *McLean v Arkansas Board of Education*, where Michael Ruse’s testimony that creation science is not science, by Popper’s criterion, but the theory

of evolution is, apparently persuaded Judge Overton.<sup>13</sup>

But there is an even more serious problem with the *Daubert* Court’s reliance on Popper, of which Justice Rehnquist didn’t seem to be aware: Popper’s philosophy of science is signally inappropriate to the Court’s concern with reliability. When Popper describes his approach as “critical rationalism,” it is to emphasize that the rationality of the scientific enterprise lies in the susceptibility of scientific theories to criticism, that is, to testing, and potentially to falsification, *not* in their verifiability or confirmability. True, early on Carnap translated Popper’s word *Bewährung* as “confirmation”; for a while, thinking the issue merely verbal, Popper let it go—even occasionally using “confirm” himself. But in a footnote to the English edition of *The Logic of Scientific Discovery*, he comments that this had been a bad mistake on his part, conveying the false impression that a theory’s having been corroborated means that it is probably true.<sup>14</sup> Except for the weak moments when he condoned Carnap’s (mis)translation,<sup>15</sup> Popper insisted that corroboration must not be confused with confirmation.

The degree of corroboration of a theory represents its past performance only, and “says nothing whatever about future performance, or about the ‘reliability’ of a theory”; even the best-tested theory “is not ‘reliable.’”<sup>16</sup>—so scornful is Popper of the concept of reliability that he refuses even to use the word without putting it in precautionary scare quotes! Reiterating that he puts the emphasis “on *negative arguments*, such as negative instances or counter-examples, refutations, and attempted refutations—in short, criticism—while the inductivist lays stress on ‘*positive instances*’, from which he draws ‘non-demonstrative inferences’, and which he hopes will guarantee the ‘*reliability*’ of the conclusions of these inferences,” Popper specifically identifies Hempel as representative of those inductivists with whom he disagrees.<sup>17</sup>

Hempel is not, perhaps, the prototypical inductivist; he describes the method of science as “hypothetico-deductive,” he affirms that scientific claims should be subject to empirical check or testing, and he doesn’t follow Reichenbach and Carnap in explaining confirmation by appeal to the calculus of probabilities.

Nevertheless, Popper is surely right to see Hempel's approach as very significantly at odds with his own: Hempel is not centrally concerned with demarcating science; he questions the supposed asymmetry between verification and falsification, and argues that Popper's criterion "involves a very severe restriction of the possible forms of scientific hypotheses," for example in ruling out purely existential statements<sup>18</sup>; when he speaks of "testing" he envisages both disconfirmation and confirmation of a hypothesis; and one of his chief projects was to articulate the "logic of confirmation," i.e., of the support of general hypotheses by positive instances.

Apparently the Supreme Court hoped, by combining Hempel's account of confirmation with Popper's criterion of demarcation, to craft a crisp test to identify genuine, and hence reliable, science. But although Hempel's philosophy of science is more positive than Popper's, it isn't much more helpful regarding the question of reliability. For one thing, the confirmation of generalizations by positive instances that preoccupies Hempel is just too simplified to apply to the enormously complex congeries of epidemiological, toxicological, etc. evidence at stake in a case such as *Daubert*. For another, Hempel himself seems eventually to have concluded (rightly, I believe) that the "grue" paradox shows that confirmation isn't a purely syntactic or logical notion after all,<sup>19</sup> and late in life began to think that maybe Thomas Kuhn had been on the right track in focusing on historico-politico-sociological, rather than logical, aspects of science.<sup>20</sup>

But the most fundamental problem is that what Hempel offered was an account of supportiveness of evidence or, as he said, of "relative confirmation," the *relation between* observational evidence and hypothesis, expressible as "E confirms H [to degree n]," or "H is confirmed [to degree n] by evidence E." This, as Hempel acknowledged, falls short of an account of "absolute confirmation," the warrant of a scientific claim, which would be expressed in nonrelative terms as "H is confirmed [to degree n], period." To discriminate reliable from unreliable testimony, however, would require an account of the absolute concept—which Hempel doesn't supply.

### DAUBERT'S CONFUSIONS: "SCIENTIFIC" AND "RELIABLE"

So, the *Daubert* Court mixes up its Hoppers and its Pempels. But isn't this just a slip, of merely scholarly interest? No; it is symptomatic of the serious misunderstanding of the place of the sciences within inquiry generally revealed by the Court's equation of "scientific" with "reliable."

So successful have the natural sciences been that the words "science," "scientific," and "scientifically" are often used as generic terms of epistemological praise, meaning vaguely "strong, reliable, good"—as, in television advertisements, actors in white coats urge viewers to get their clothes cleaner with new, "scientific" Wizzo. This honorific usage is unmistakably at work in the *Daubert* ruling; indeed, it seems to be implicit even in the way Justice Blackmun writes of "scientific . . . knowledge," strategically excising three not insignificant words from the reference in Federal Rules of Evidence 702 to "scientific or other technical knowledge," apparently signaling an expectation that a criterion of the genuinely scientific will also discriminate reliable testimony from unreliable.

If "scientific" is used honorifically, it is a tautology that "scientific" equals "reliable"; but this tautology, obviously, is of no help to a judge trying to screen proffered scientific testimony. If "scientific" is used descriptively, however, "scientific" and "reliable" come apart: for, obviously, physicists, chemists, biologists, medical scientists, etc., are sometimes incompetent, confused, self-deceived, dishonest, or simply mistaken, while historians, detectives, investigative journalists, legal and literary scholars, plumbers, auto mechanics, etc., are sometimes good investigators. In short, not all, and not only, scientists are reliable inquirers; and not all, and not only, scientific evidence is reliable. Nor is there a "scientific method" in the sense the Court assumed: no uniquely rational mode of inference or procedure of inquiry used by all scientists and only by scientists. Rather, as Einstein once put it, scientific inquiry is "nothing but a refinement of our everyday thinking,"<sup>21</sup> superimposing on the inferences, desiderata, and constraints common to all serious investigation a vast variety of constantly

evolving local ways and means of stretching the imagination, amplifying reasoning power, extending evidential reach, and stiffening respect for evidence.

Every kind of empirical inquiry, from the simplest everyday puzzling over the causes of delayed buses or spoiled food to the most complex investigations of detectives, of historians, of legal and literary scholars, and of scientists, involves making an informed guess about the explanation of some event or phenomenon, figuring out the consequences of its being true, and checking how well those consequences stand up to evidence. This is the procedure of all scientists; but it is not the procedure *only* of scientists. Something like the "hypothetico-deductive method," really is the core of all inquiry, scientific inquiry included. But it is not distinctive of scientific inquiry; and the fact that scientists, like inquirers of every kind, proceed in this way tells us nothing substantive about whether or when their testimony is reliable.

The sciences have extended the senses with specialized instruments; stretched the imagination with metaphors, analogies, and models; amplified reasoning power with numerals, the calculus, computers; and evolved a social organization that enables cooperation, competition, and evidence-sharing, allowing each scientist to take up his investigation where others left off. Astronomers devise ever more sophisticated telescopes, chemists ever more sophisticated techniques of analysis, medical scientists ever more sophisticated methods of imaging bodily states and processes, and so on; scientists work out what controls are needed to block a potential source of experimental error, what statistical techniques are needed to rule out a merely coincidental correlation, and so forth. But these scientific "helps" to inquiry are local and evolving, not used by *all* scientists.<sup>22</sup>

You may object that, since I have acknowledged that scientific inquiry is continuous with everyday empirical inquiry, I have in effect agreed with Popper that science is an extension of common sense. Indeed, I think science is well described, in Gustav Bergmann's wonderfully evocative phrase, as the "Long Arm of Common Sense."<sup>23</sup> But the continuity is not between the content of scientific and of commonsense knowledge, but between the

basic ways and means of everyday and of scientific inquiry; and it is precisely because of this continuity that the Popperian preoccupation with the “problem of demarcation” is a distraction.

Or you may object that the *Daubert* Court’s Popperian advice that courts ask whether proffered scientific testimony “can be and has been tested” surely is potentially helpful. This is true; but it is no real objection. “Check whether proffered testimony has been tested” is very good advice when a purported expert hasn’t made even the most elementary effort to check how well his claims stand up to evidence: such as the knife-mark examiner in *Ramirez v State*,<sup>24</sup> who testified that he could infallibly identify this knife, to the exclusion of all other knives in the world, as having made the wound—though no study had established the assumed uniqueness of individual knives, and his purported ability to make such infallible identifications had never been tested. This is not, however, because falsifiability is the criterion of the scientific, but because *any* serious inquirer is required to seek out all potentially available evidence and to go where it leads, even if he would prefer to avoid, ignore, or play down information that pulls against what he hopes is true.

Yes, this is a requirement on scientists; as Darwin recognized when he wrote in his autobiography that he always made a point of recording recalcitrant examples and contrary arguments in a special notebook, to safeguard against his tendency conveniently to forget negative evidence.<sup>25</sup> But it is no less a requirement on other inquirers, too; as we all realized a few years ago when a historian who announced that he had evidence that Marilyn Monroe had blackmailed President Kennedy turned out to have ignored the fact that the supposedly incriminating letters were typed with correction ribbon, and the address included a zip code—when neither existed at the time the letters were purportedly written!<sup>26</sup>

“Nonscience” is an ample and diverse category that includes the many human activities other than inquiry, the various forms of pseudo-inquiry, inquiry of a nonempirical character, and empirical inquiry of other kinds than the scientific; and of course there are plenty of mixed and borderline cases.

The honorific use of “science” and its cognates tempts us, as it did the *Daubert* Court, to criticize poorly conducted science as not really science at all; but “not scientific” is as unhelpful as generic epistemic criticism as “scientific” is as generic epistemic praise. The pejorative tone of the phrase “pseudoscience,” which presumably refers to activities that purport to be science but aren’t really, derives in part from its imputation of false pretenses, generally, and in part from the favorable connotations of “scientific,” specifically. But rather than sneering unhelpfully that this or that work is “pseudoscientific,” it is always better to get down to those “brass tacks” Bridgman talks about, and specify what, exactly, is wrong with it. Is it dishonestly or carelessly conducted? Does it rest on flimsy or vague assumptions—assumptions for which there is no good evidence or which aren’t even susceptible to evidential check? Does it seek to impress with decorative or distracting mathematical symbolism or elaborate-looking apparatus? Does it fail to take essential precautions against experimental error? And so on.

### DAUBERT’S LEGAL PROGENY

So, the *Daubert* Court’s philosophy of science was muddled; but haven’t subsequent Supreme Court rulings cleared things up? Not exactly: it would be more accurate to say that in *General Electric Co v Joiner* (1997) and *Kumho Tire Co v Carmichael* (1999), the Supreme Court quietly backed away from *Daubert*’s confused philosophy of science.<sup>27</sup> At any rate, those references to Hepper, Pompel, falsifiability, etc., so prominent in *Daubert*, are conspicuous by their absence from *Joiner* and *Kumho*. But there are points of epistemological interest.

In *Joiner*, there was a bit of a kerfuffle about “methodology”: *Joiner*’s attorneys had argued that the lower court erred in excluding their proffered expert testimony because, instead of focusing exclusively on their experts’ methodology—which, they maintained, was the very same “weight of evidence” methodology used by the other party’s (General Electric’s) experts—improperly concerned itself with the experts’ conclusions. Apparently anxious to sidestep this argument,

the *Joiner* Court (with the exception of Justice Stevens) flatly denied the legitimacy of the distinction between methodology and conclusions. Opining that this is No Real Distinction, the Court sounded like nothing so much as a conclave of medieval logicians; but given their citation to *Paoli*,<sup>28</sup> it seems likely that they didn’t really intend to make a profound metaphysical pronouncement, only to acknowledge, as Judge Becker had, that if an expert’s conclusions are problematic enough, this alerts us to the possibility of some methodological defect.

This focus on “methodology”—an accordion concept expanded and contracted as the argument demands<sup>29</sup>—obscured a much deeper epistemological question. *Joiner*’s attorneys proffered a collage of bits of information, none sufficient by itself to warrant the conclusion that exposure to polychlorinated biphenyls promoted *Joiner*’s cancer, but which, they argued, *taken together* gave strong support to that conclusion. General Electric’s attorneys replied, in effect, that piling up weak evidence can’t magically transform it into strong evidence. In response, *Joiner*’s attorneys referred to the EPA guidelines for assessing the combined weight of epidemiological, toxicological, etc. evidence. But no one addressed the key question: Is there a difference between a congeries of evidence so interrelated that the whole really is greater than the sum of its parts, and a collection of unrelated and insignificant bits of information between true consilience and the “faggot fallacy”<sup>30</sup>—and if so, what is it?

There *is* a difference. Evidence of means, motive, and opportunity may interlock to support the claim that the defendant did it much more strongly than any of these pieces of evidence alone could do. Similarly, evidence of increased incidence of a disease among people exposed to a suspected substance may interlock with evidence that animals biologically similar to humans are harmed by exposure to that substance and evidence indicating what chemical mechanism may be responsible, to support the claim that this substance causes, promotes, or contributes to the disease, much more strongly than any of these pieces of evidence alone could do. However, the interlocking will be less robust if, for example, the animals

are unlike humans in some relevant way, or if the mechanism postulated to cause damage is also present in other chemicals not found to be associated with an increased risk of disease, and so on.

“Interlocking” is exactly the right word; for evidence is structured like a crossword puzzle, with warranted claims anchored by experiential evidence (the analogue of clues) and enmeshed in reasons (the analogue of completed intersecting entries). How reasonable a crossword entry is depends on how well it is supported by the clue and completed intersecting entries, how reasonable those other entries are, independent of this one, and how much of the crossword has been completed; similarly, how warranted a claim is depends on how supportive the evidence is, how secure the reasons are, independent of this claim itself, and how much of the relevant evidence the evidence includes.<sup>31</sup> Because of the ramification of reasons, the desirable kind of interlocking of evidence gestured at in *Joiner* is subtle and complex, not easily captured by any mechanical weighting of epidemiological data relative to animal studies or toxicological evidence. Nor, moreover—as Justice Rehnquist already saw in the context of *Daubert*—can its quality readily be judged by someone who lacks the necessary background knowledge.

In *Kumho*, the Supreme Court made a real epistemological step forward. In this products-liability case, which focused on the proffered testimony of an expert on tire failure, the Court tried to sort out the problems with non-scientific experts which, as Justice Rehnquist had anticipated, soon arose in the wake of *Daubert*; and ruled that judges can’t evade their gatekeeping duty on the grounds that proffered expert testimony is not science; the key word in Federal Rule of Evidence 702, after all, is “knowledge,” not “scientific.” No longer fussing over demarcation, recognizing the gap between “scientific” and “reliable,” in *Kumho* the Supreme Court acknowledged that *what matters is whether proffered testimony is reliable, not whether it is scientific*. Quite so.

Far from backing away from federal courts’ gatekeeping responsibilities, however, the *Joiner* Court affirmed that a judge’s decision to allow or exclude scientific testimony, even though it may be outcome-determinative, is

subject only to review for abuse of discretion, not to any more stringent standard; and the *Kumho* Court, pointing out that, depending on the nature of the expertise in question, the *Daubert* factors may or may not be appropriate, held that it is within judges’ discretion to use any, all, or none of them. A year later, revised Federal Rules of Evidence made explicit what according to *Daubert* had been implicit in Rule 702 all along: admissible expert testimony must be based on “sufficient” data, the product of “reliable” testimony “reliably” applied to the facts of the case. Federal judges now have large responsibilities and broad discretion in screening not only scientific testimony but expert testimony generally—but very little guidance about how to perform this difficult task.

Post-*Daubert* courts have apparently been significantly tougher than before on expert testimony proffered by plaintiffs in civil cases. This isn’t the place for a full-scale discussion of the frequently heard criticism that *Daubert* and its progeny tend to favor defendant corporations over plaintiffs; but I will say that I think things are a lot more complicated than this criticism suggests. No doubt there are heartless and unscrupulous companies more concerned with profit than with the dangers their products may present to the public; and it is certainly easier to sympathize with poor Jason Daubert or with poor Mr. Joiner than with a vast, impersonal outfit like Merrell Dow or General Electric. But no doubt there are also greedy and opportunistic plaintiffs and plaintiffs’ attorneys—and the people thrown out of work when meritless litigation forces a company to downsize or close also deserve our sympathy. Moreover, although we certainly hope that the tort system will discourage the manufacture of dangerous substances and products, we also want it *not* to discourage the manufacture of safe and useful ones. And I will add that, although it seems that since *Daubert* courts have not—at least not yet—been as tough on expert testimony proffered by prosecutors in criminal cases as they have on plaintiffs’ experts in civil cases, we surely also want to avoid convicting innocent criminal defendants on flimsy forensic testimony—and leaving the real offenders at liberty.<sup>32</sup> That said, I will leave it to others to pursue

*Daubert*’s policy ramifications, and pick up the epistemological thread once more.

## WHERE DO WE GO FROM HERE?

So, since *Kumho*’s epistemological step forward, the other problem Justice Rehnquist worried about—that judges generally lack the background knowledge that may be essential to a serious appraisal of the worth of scientific (or other technical) testimony—looms larger than ever. But hasn’t the legal system by now found ways to help judges handle their quite burdensome responsibilities for keeping the gate against unreliable expert testimony? Up to a point; but only up to a point. Ways have been explored to give judges some of the background knowledge they may need, and to enable them to call on the scientific community for help; but these have been relatively small steps, and sometimes (understandably) fumbling.

*Daubert* prompted various efforts to educate judges scientifically. In May 1999, for example, about two dozen Massachusetts Superior Court judges attended a 2-day seminar on DNA at the Whitehead Institute for Biomedical Research. A report in the *New York Times* quoted the director of the Institute assuring readers that, while in the O.J. Simpson case, lawyers had “befuddled everyone” over the DNA evidence, after a program such as this judges would “understand what is black and white . . . what to allow in the courtroom.”<sup>33</sup> To be candid, this report leaves me a little worried about the danger of giving judges the false impression that they *are* qualified to make subtle scientific determinations; when it is hardly realistic to expect that a few hours in a science seminar will transform judges into scientists competent to make subtle and sophisticated scientific judgments—any more than a few hours in a legal seminar could transform scientists into judges competent to make subtle and sophisticated legal determinations.

It really isn’t feasible to bring—let alone keep—judges up to speed with cutting-edge genetics, epidemiology, toxicology, and so on. (This is not in the least to denigrate judges’ abilities, but rather to draw the analogy with expecting a few lessons to turn a professional football player into a ballet dancer, or me

into a concert pianist.) It ought to be possible, however, to educate judges in the elements of probability theory, to give them a sense of how samples may be mishandled or this or that kind of mistake made at the laboratory, and to explain how information about the probability that the lab made a mistake is such-and-such affects the significance of a random-match probability. More generally, it seems both feasible and useful to try to ensure that judges understand the more commonly employed scientific ideas they are likely to encounter most frequently: the role of suggestion, for example, and its significance for how DNA samples or suspect knives, etc. should be presented or how photo-arrays or lineups should be conducted. Of course, when the issues are subtle, the subtleties need to be conveyed; one would hope that judges understand the concept of statistical significance, for instance—but also grasp the element of arbitrariness it involves.

Since 1975, under Federal Rule 706 and many state equivalents, courts have had the power to appoint experts of their own selection. Used in a number of asbestos cases in 1987 and 1990,<sup>34</sup> the practice came to public attention in the late 1990s in the context of a wave of lawsuits against the manufacturers of silicone breast implants, when it was adopted by Judge Jones in *Hall*,<sup>35</sup> and most notably by Judge Samuel Pointer, who in 1996 appointed a National Science Panel to help him sift through the scientific evidence in the several thousand federal silicone breast implant cases that had been consolidated to his court. And it seems that, as their gate-keeping responsibilities have grown, more judges have been willing, as Justice Breyer urged in *Joiner*, to call directly on the scientific community for help<sup>36</sup>: court-appointed experts have advised judges on the potential dangers of seatbelt buckles, the diet drug fenphen, and the antilactation drug Parlodel; and, in the Court of Appeals in Michigan, on Bendectin.<sup>37</sup> At the American Academy for the Advancement of Science, the Court-Appointed Scientific Experts Project makes available “independent scientists . . . [to] educate the court, testify at trial, assess the litigants’ cases, and otherwise aid in the process of determining the truth”; Duke University’s Registry of Independent Scientific and Tech-

nical Advisors also provides the names of independent experts.

It has been said that the use of court-appointed experts is “elitist” and “undemocratic,” even “totalitarian”<sup>38</sup>; but this strikes me as something of an exaggeration. Certainly, trial by jury is a better way of getting at the truth than trial by oath or ordeal; certainly citizens’ service on juries is an expression of the democratic ethos (though it would be strange to deny that the Netherlands, say, is a democracy, simply because the Dutch judicial system routinely relies on experts appointed by the courts). Still, especially considering how tiny the proportion of federal cases decided by juries now is,<sup>39</sup> it seems reasonable to be willing to consider adapting the adversarial culture a little in this way,<sup>40</sup> if and when this would better serve the fundamental purpose of protecting against arbitrary and irrational determinations of fact.

Sometimes it is thought that there *are no* neutral experts. If neutrality is taken to mean freedom from all preconceptions, it is true that there are few if any neutral experts: anyone competent to the task of a court-appointed scientist is virtually certain to have some view at the outset. And if neutrality is taken to mean freedom from all contact, direct or indirect, with either party, again there probably won’t be many neutral scientists; for, given the dependence of much medical research on drug company funding,<sup>41</sup> most scientists competent to the task will probably know people involved with one party or the other. But it doesn’t follow, and it isn’t true, that some experts aren’t, in the essential sense, more neutral, less biased than others: that is, more willing to go where the evidence leads, even if it pulls against what they were initially inclined to believe.

Bias, in the sense at issue here, is not the same as conflict of interest; nevertheless, we certainly want to avoid conflicts of interest, both because they may lead to bias in the relevant sense, and because, even if they don’t, we want to avoid the appearance of such bias. But we should be conscious that there is a broad continuum from a court-appointed scientist’s being financially supported in some way by a defendant company or plaintiffs’ attorneys, to his discussing his court-appointed work with an acquaintance

who is supported in some way by a defendant company or by plaintiffs’ attorneys, to his simply having such acquaintances . . . to his being completely out of any professional loop in the field in question.

Yes, it is disturbing that, while serving on Judge Pointer’s panel, one scientist signed a letter asking for financial support for another project from one of the defendant companies; and worrying that just four scientists were, in effect, responsible for the disposition of several thousand cases. Moreover, given that even competent and honest scientists will sometimes legitimately disagree, we need to think about what will happen when court-appointed scientists are not of one mind. Both legal issues and practical questions need to be addressed, among them<sup>42</sup>: Should court-appointed experts help judges with their *Daubert* screening duties, or should they testify before juries, along with the parties’ experts? How could court-appointed experts best be selected? Who should pay for their services? How should they be instructed about conflicts of interest? We could learn a lot from Judge Pointer’s experience—and (if we are careful to avoid the pitfalls of facile crosscultural comparisons) from the experience of other legal systems—about how and when court-appointed experts might be most helpful.

Such experts are potentially very useful in some kinds of cases; but of course they are no panacea—in fact, I don’t suppose for a moment that there *is* a panacea. Rather, there is a range of possibilities worth pursuing. Thinking about the unhappy interaction of the FDA and the tort system in the silicone breast implants affair, for example, you might wonder how the FDA could have acted to prevent the panic in the first place<sup>43</sup>; thinking about the willingness of the American Association for the Advancement of Science to help, one might wonder about other ways of making the scientific community more responsive when legal disputes turn on scientific issues irresolvable by the presently available evidence; thinking of the weaknesses of other techniques of forensic identification, and the mistakes made by crime labs, etc., revealed in the wake of those dramatic DNA exonations, one might wonder how we could make the forensic science business more rigorous

(the temptation to say “more scientific” is strong; but I shall resist it!).

Justice requires not only just laws, and just administration of those laws, but also factual truths—which, increasingly often, courts must rely on science to discover. As Learned Hand once put it, “No one will deny that the law should in some way effectively use expert knowledge wherever it will aid in settling disputes. The only question is as to *how it can do so best*.”<sup>44</sup> Now, more than a century after Hand posed the essential question, and more than a decade after *Daubert*, we are still fumbling towards an answer. ■

### About the Author

The author is Professor of Law and of Philosophy at the University of Miami, Miami, Fla.

Requests for reprints should be sent to Susan Haack, PhD, School of Law, University of Miami, 1311 Miller Drive, Coral Gables, FL 33124 (e-mail: shaack@miami.edu).

This article was accepted July 27, 2004.

### Acknowledgments

This work was supported in part by the Project on Scientific Knowledge and Public Policy.

The author is grateful to Mark Migotti for helpful comments on a draft.

### References

- Bridgman P. On “scientific method.” Originally published in 1949; reprinted in: Bridgman P. *Reflections of a Physicist*. New York, NY: Philosophical Library; 1995:81–83.
- Frye v United States*, 54 App DC 46, 293 F 1013,1014 (1923).
- Barefoot v Estelle*, 463 US 880,898, 103 SCt 3383,3397 (1983). Barefoot was executed in 1984.
- Daubert v Merrell Dow Pharmaceuticals, Inc*, 509 US 579, 113 SCt 2786 (1993).
- The *Daubert* Court did not itself scrutinize the disputed testimony; on remand, Judge Kozinski again excluded the plaintiffs’ proffered experts, this time under *Daubert* rather than *Frye*. Because of litigation costs, Merrell Dow had already taken Bendectin off the market in 1984. In 2000, the FDA again declared the drug safe.
- Popper KR. *Logik der Forschung*. Vienna, Austria: Julius Springer, 1934. English edition, *The Logic of Scientific Discovery*. London, UK: Hutchinson, 1959. The observation that his criterion of demarcation is a convention, found in the original German edition, appears on page 37 of *The Logic of Scientific Discovery*; the observation that science is continuous with commonsense knowledge appears only in the new Preface added to the English edition, page 18.
- See Popper KR. Philosophy of science: a personal report. In: Mace CA, ed. *British Philosophy in Mid-Century*. London, UK: Allen & Unwin; 1957; reprinted under the title *Science: conjectures and refutations* in Popper KR. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York, NY: Basic Books; 1962: 33–69; and in part, under the title *Falsificationism*, in Klee R, ed. *Scientific Inquiry: Readings in the Philosophy of Science*. Oxford, UK: Oxford University Press; 1999: 65–71; see also Popper KR. The problem of demarcation (1974) in: Miller D, ed. *A Pocket Popper*. London, UK: Fontana; 1983:118–130.
- Popper KR. *The Logic of Scientific Discovery*. London, UK: Hutchinson; 1959: section 83.
- Popper KR. *Objective Knowledge: An Evolutionary Approach*. Oxford, UK: Oxford University Press; 1972: 102.
- Popper KR. *Objective Knowledge: An Evolutionary Approach*. Oxford, UK: Oxford University Press; 1972: 22.
- Daubert v Merrell Dow Pharmaceuticals, Inc*, 509 US 579,600; 113 SCt 2786,2800 (1993). Some federal judges evidently understand falsifiability better than others. In *US v Havvard*, 117 FSupp2d 848,854 (2000), admitting fingerprint identification testimony, Judge Hamilton observes that “the methods of latent print identification . . . have been tested . . . for roughly 100 years . . . in adversarial proceedings.” But in *Llera-Plaza I*, 2002 WL 27305 (ED Pa 2002), 27310 imposing restrictions on fingerprint identification testimony, Judge Pollak points out that “‘adversarial’ testing in court is not . . . what the Supreme Court meant when it discussed testing as an admissibility factor.” (Shortly thereafter, Judge Pollak reconsidered and revised his ruling, but on grounds unrelated to the point at issue here.)
- See Popper KR. Natural selection and its scientific status; reprinted from sections 1 and 2 of a lecture of 1977 in: Miller D, ed. *A Pocket Popper*. London, UK: Fontana; 1983:239–246.
- McLean v Arkansas Board of Education*, 529 FSupp 1255 (1982). Judge Overton’s ruling and Ruse’s testimony, along with Larry Laudan’s properly scathing critique, can be found in Ruse M, ed. *But Is It Science? The Philosophical Question in the Creation/Evolution Controversy*. Amherst, NY: Prometheus; 1996.
- Popper KR. *The Logic of Scientific Discovery*. London, UK: Hutchinson, 1959: 251–252, Note number \*1, added in the English edition. When Popper uses “confirm” for “corroborate,” as he does in his *Philosophy of science: a personal report* (1957), the effect is powerfully confusing.
- I am being deliberately noncommittal about whether this really is a mistranslation. *Pons’ Global-wörterbuch Deutsch-Englisch* (1983), explains *Bewährung* as “proving one’s/its worth”; a secondary meaning is “probation.”
- Popper KR. *Objective Knowledge: An Evolutionary Approach*. Oxford, UK: Oxford University Press; 1972: 18, 22.
- Popper KR. *Objective Knowledge: An Evolutionary Approach*. Oxford, UK: Oxford University Press; 1972: 20; the reference to Hempel is in footnote 29.
- Hempel CG. Studies in the logic of confirmation. *Mind*. 1945; 54:1–26, 97–121; reprinted in Hempel CG. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, NY: Free Press; 1965: 3–46, 43–45. See also Hempel CG. Empirist criteria of cognitive significance: problems and changes. Adapted from two papers originally published in 1950 and 1951. In: Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, NY: Free Press; 1965: 101–120, with the addition of Postscript (1964) on cognitive significance, 120–122.
- Hempel CG. Postscript (1964) on confirmation. In: Hempel CG. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, NY: Free Press; 1965:47–51, 51.
- Hempel CG. The irrelevance of truth for the critical appraisal of scientific theories (1990). Reprinted in: Jeffrey R, ed. *Selected Philosophical Essays [by] Carl G. Hempel*. Cambridge, UK: Cambridge University Press; 2000: 75–84. Kuhn T. *The Structure of Scientific Revolutions*. Chicago, Ill: University of Chicago Press; 1962.
- Einstein A. Physics and reality. *Journal of the Franklin Institute*, 221, No. 3 (1936). Reprinted in: Bargmann S. *Ideas and Opinions of Albert Einstein*. New York, NY: Crown Publishers; 1954: 290–323, 290.
- For a detailed development of the conception of scientific method on which I have relied here, see Haack S. *Defending Science—Within Reason: Between Scientism and Cynicism*. Amherst, NY: Prometheus; 2003: 93–121.
- Bergmann G. *Philosophy of Science*. Madison, Wis: University of Wisconsin Press; 1957: 20.
- Ramirez v State*, 542 So2d 352 (Fla 1989); *Ramirez v State*, 651 So2d 1164 (Fla 1995); *Ramirez v State*, 8120 So2d 836 (Fla 2001). Florida remains officially a *Frye* state, but it seems to be rapidly evolving in the direction of (as Michael Saks puts it) “*Fryebert*.”
- Darwin F, ed. *Charles Darwin: Autobiography and Letters*. New York, NY: D. Appleton and Company; 1893:45. Reprinted New York, NY: Dover; 1952.
- See Thomas E, Hosenball M, Isikoff M. The JFK–Marilyn hoax. *Newsweek*. June 6, 1997:36–37.
- General Electric Co v Joiner*, 522 US 136, 118 SCt 512 (1997); *Kumho Tire Co v Carmichael*, 526 US 137, 119 SCt 1167 (1999).
- In re: Paoli R.R. Yard PCB Litig.*, 35 F3d 717 (3d Cir 1994).
- The term “accordion concept” was introduced in Sellars W. Scientific realism or irenic instrumentalism? In: Cohen R, Wartofsky M, eds. *Boston Studies in the Philosophy of Science*. Vol 2. New York, NY: Humanities Press, 1965:171–204.
- The word “consilience,” meaning etymologically “jumping together,” was coined by the 19th century philosopher of science William Whewell, and recently made famous as the title of a bestselling book by E.O. Wilson, *Consilience: The Unity of Knowledge*. New York, NY: Knopf, 1998. The phrase “faggot fallacy,” adopted by General Electric’s attorneys in *Joiner*, was introduced in *Follies and Fallacies in Medicine* by P. Skrabanek and J. McCormick. Originally published in 1989; reprinted Amherst, NY: Prometheus Books, 1997.
- I first introduced the analogy in Haack S. Rebuilding the ship while sailing on the water. In: Gibson R, Barrett R, eds. *Perspectives on Quine*. Oxford, UK: Blackwell; 1990:111–128. It is articulated in more detail in Haack S. *Evidence and Inquiry: Towards Reconstruction in Epistemology*. Oxford, UK: Blackwell; 1993: 73–94; and developed further in Haack S. *Defending Science – Within Reason*. Amherst, NY: Prometheus; 2003: 57–91.

32. As one character says to another in a cartoon for which I have a particular fondness, “Politically, I suppose you could say I’m a member of the lunatic middle.”
33. Goldberg C. Judges’ unanimous verdict on DNA lessons: Wow! *New York Times*. April 24, 1999:A10.
34. See Rubin CR, Ringenback L. *The Use of Court Experts in Asbestos Litigation*, 137 FRD 35 (1991).
35. *Hall v Baxter Healthcare Corp*, 947 FSupp 1387 (D Ore 1996).
36. See Erichson HM. Mass tort litigation and inquisitorial justice. *Geo Law J*. 1999;87:1983–2024.
37. *DePypier et al. v Paul V. Navarro*, No. 19149, 1998 WL 1988927 (Mich App. Nov. 6, 1998); *Denial of Expert Witness Testimony Violates Daubert, Appeal States*. DES Litig Rep., Dec. 1998;
38. Howard MN. The neutral expert: a plausible threat to justice. *Crim Law Rev*. 1991; 98–105; cited by: Van Kampen P. Expert evidence compared. In: Malsch M, Nijboer JF, eds. *Complex Cases: Perspectives on the Netherlands Criminal Justice System*. Amsterdam, The Netherlands: Thela Thesis; 1999.
39. Only 4.4% of federal criminal cases end in a jury verdict, and only 1.4% of federal civil cases are resolved by juries. Glaberson W. Juries, their powers under siege, find their role is being eroded. *New York Times*. March 2, 2001: A1.
40. I have written at greater length about tensions between science and the culture of the law in “Inquiry and advocacy, fallibilism and finality: culture and inference in science and the law,” *Law, Probability and Risk*. 2003; 2:205–214.
41. See Angell M. Is academic medicine for sale? *N Engl J Med*. 2000; 342(20):1516–1518.
42. See, for example, Cecil JS, Hooper LJ, Willging TE. Assessing causation in breast implant litigation: the role of science panels. *Law Contemp Prob*. 2001; 64:139–189. Monahan J, Walker L. Scientific authority: the breast implant litigation and beyond. *Va Law Rev*. 2002;86:801–833.
43. The wave of litigation began after the FDA banned silicone breast implants, formerly “grandfathered in”; they were not known to be unsafe, but the manufacturers had failed to submit evidence of their safety, as they had been required to do.
44. Learned Hand. Historical and practical considerations regarding expert testimony. *Harv Law Rev*. 1901; 15:40–58, 40 (my emphasis).