

The Gatekeeping Role in Civil Litigation and the Abdication of Legal Values in Favor of Scientific Values

Neil B. Cohen*

INTRODUCTION

Beginning almost two decades ago, I wrote a series of law review articles about the nature of legal proof in the context of probabilistic evidence or analysis.¹ In these articles, I explored, in passing, some of the differences between the legal system's concept of proof of facts, especially in the context of civil litigation, and the same subject as seen by scientists, especially those analyzing empirical data.

I put that endeavor aside for what turned out to be an extended period while my career took a different turn, focusing on the intricacies of commercial law. In the last decade, much of my time has been devoted to the problems of international commercial transactions. In particular, a major portion of my work during this time has involved the negotiation of various principles to govern international transactions. In one case, I was the chair of a task force within the drafting committee preparing Revised Article 9 of the Uniform Commercial Code which developed Article 9's rules governing international secured transactions from the perspective of domestic United States law. In another case, I spent several years as a member of the United States delegation to the working group of the United Nations Commission on International Trade Law ("UNCITRAL") which developed an international convention (now

* Professor of Law, Brooklyn Law School; S.B., Massachusetts Institute of Technology; J.D., New York University. The author wishes to acknowledge the assistance of a Brooklyn Law School Summer Research Grant in the preparation of this article.

¹ See Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385 (1985) [hereinafter *Confidence in Probability*]; Neil B. Cohen, *The Costs of Acceptability: Blue Buses, Agent Orange, and Aversion to Statistical Evidence*, 66 B.U. L. REV. 563 (1986) [hereinafter *Costs of Acceptability*]; Neil B. Cohen, *Conceptualizing Proof and Calculating Probabilities: A Response to Professor Kaye*, 73 CORNELL L. REV. 78 (1987) [hereinafter *Conceptualizing Proof*].

known as the United Nations Convention on the Assignment of Receivables in International Trade). More recently, I have been a delegate to UNCITRAL's effort to develop an international legislative guide for secured transactions.

I mention these international activities not to recite my resume but, rather, because they relate in unexpected (to me, at least) ways to the topic at hand in this symposium. In many of the international negotiations in which I have been engaged, the most difficult paths to resolution were not those that involved doctrines with respect to which legal systems obviously differed. In those cases, representatives of the relevant states were prepared for the advocacy and give-and-take that are part of the resolution of the differences in an international forum. Rather, the greatest challenges to resolving differences arose where legal systems used the same or similar words to describe legal doctrines based on very different concepts. In these areas, the delegates from the affected states had significant difficulty understanding the fact that delegates from other states could be using the same words to describe very different concepts. The result, in many cases, was that there was often a greater appearance of agreement than there was agreement itself. Moreover, when delegates finally discovered their disagreements, solutions were more difficult to reach because the initial reactions to discovery of the nature of the disagreement tended to be on the order of "[t]hey don't know what they're talking about—that word just doesn't mean what they're using it to mean!"

As a result, similarities in nomenclature first obscured differences in substantive norms and then made these differences more difficult to resolve. The result of such apparent agreement, and more difficult resolution of disagreement, created by inconsistent nomenclature can easily lead situations in which those versed in different systems talk at each other, each convinced of the inherent virtue of the concepts of his or her system, rather than learn from each other.

At this point, a reader might ask how these difficulties in international substantive law negotiations relate to the role of experts in civil litigation. The answer, I believe, is that the relationship is quite strong. When experts from other disciplines testify in a legal proceeding, particularly when the proceeding is a fact-finding exercise in civil litigation governed by standard burdens of persuasion, rather than a criminal trial with its higher burden on the prosecution, there is significant risk that the use, by law and the other disciplines, of similar vocabulary to abbreviate materially different concepts will result in mirages that can easily lead to inappropriate

evidentiary and substantive decisions.

In this article, I focus on civil litigation, rather than on the criminal process, because the legal system is much more open about the meaning of the fact-finding process in civil litigation. While no two people seem to agree on the meaning of the “beyond a reasonable doubt” standard of proof in a criminal case,² this is not the case for civil litigation. In the civil context, the range of views as to the meaning of the burden of persuasion is much narrower. In particular, it seems well-accepted that the standard burden of persuasion in civil cases—the “preponderance of the evidence” standard—can be expressed probabilistically: the plaintiff (or other party bearing the burden of persuasion) must demonstrate that the probability of the facts supporting its case exceeds 0.5.³ This ability to express the burden of persuasion probabilistically has several implications. For one thing, it promises (but does not always deliver) a conceptual link between legal proof standards and the standards of scientific and technical disciplines that rely on probabilistic reasoning. Second, and perhaps more important, it masks the important differences in the value systems that govern standards of legal proof and parallel standards of scientific and technical inquiry. In particular, the standard of proof in civil litigation, and the value system that lies beneath it, weigh the cost of errors so differently than most scientific disciplines that the fact that a “mainstream” scientist would not testify as to a particular conclusion does not necessarily mean that the same conclusion is valueless or “junk science” for the purposes of law.

I write this essay because I am quite concerned that the use of rules generated by *Daubert v. Merrell Dow Pharmaceuticals*⁴ and its progeny to determine whether to admit scientific and technical expert testimony will fall prey, and perhaps already has fallen prey, to the “similar-words-but-differing-meaning” phenomenon that I have described above. Federal Rule of Evidence 702,⁵ with its emphasis on scientific *knowledge*, and *Daubert*, with its emphasis on acceptance in

² See, e.g., *United States v. Fatico*, 458 F. Supp. 388, 409-11 (E.D.N.Y. 1978) (Weinstein, J.) (surveying judges of the United States District Court for the Eastern District of New York), *aff'd on other grounds*, 603 F.2d 1053 (2d Cir.1979).

³ Of course, this is an oversimplification in many respects. Moreover, the consensus I describe is subject to some important conceptual differences from a simple probabilistic formulation. See Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357 (1985); see also *Confidence in Probability*, *supra* note 1.

⁴ 509 U.S. 579 (1993).

⁵ Fed. R. Evid. 702.

the scientific community, create a great risk that evidence that is relevant to the demonstration of facts at issue in a civil proceeding will be excluded or substantively minimized by misunderstandings about how data (i.e., evidence) is utilized to develop statements of “knowledge” in scientific communities consistent with the norms of those communities—norms that differ fundamentally from the norms of civil litigation.⁶ Because plaintiffs typically bear the burden of persuasion in civil litigation, plaintiffs will likely bear the costs of this misunderstanding. Even if one does not shed a tear for plaintiffs as a general matter, however, I suggest that one should nonetheless be concerned about the fact that such misunderstandings can increase the amount and cost of errors in the civil litigation system.

Before continuing, I should point out at this juncture that my lengthy sojourn from immersion in the fields of probability and proof, combined with the fact that I am not, and do not purport to be, an expert in the law of evidence generally, leads me to paint with a somewhat broad brush. I am also likely to fall prey to the same phenomenon that I described in the context of international litigation; namely, that my use of various terms in this article may not always be precisely consistent with their meaning as terms of art in the world of evidence or in the world of probability and statistics. Moreover, I have striven to use as little technical vocabulary as possible in this essay, so as not to obscure my areas of concern. Indeed, the legal system’s attempts to piggy-back on the language used in scientific reasoning, rather than on the underlying reasoning itself, may be part of the problem about which I am concerned.

Part I of this essay compares and contrasts the process by which courts in civil litigation and scientists, particularly those in highly empirical disciplines such as epidemiology (a frequent source of scientific expertise in civil litigation), utilize the knowledge of discrete and incomplete data about the possible existence of an ultimate fact to form a conclusion as to whether that fact exists. Part II of this essay identifies concerns emanating from the actual and potential misunderstanding of these differences by courts. Finally, Part III of this essay sketches out the contours of a theory explaining why the *Daubert* model of gatekeeping might be appropriate to filter out “junk,” but inappropriate when used to filter out legitimate analysis that is sensitive to the needs of civil litigation.

⁶ See Lucinda M. Finley, *Guarding the Gate to the Courthouse: How Trial Judges Are Using Their Evidentiary Screening Role to Remake Tort Causation Rules*, 49 DEPAUL L. REV. 335 (1999).

I. FACT-FINDING IN LAW AND SCIENCE

While some aspects of science are highly theoretical and deductive, many other aspects are inductive. In this context, I use “inductive” not so much as a description of the reasoning process as a description of the knowledge development process. Although the process of generating theories or models from higher (or earlier established) principles already known or believed to be true is deductive, the process of testing these theories—examining the factual evidence to see if the theories work as promised—is inductive. Most scientific endeavors contain this inductive, or fact-inferential, aspect. Some disciplines, such as epidemiology, are largely dominated by it.

Similarly, while much of legal reasoning is deductive (minor rules that follow from the existence of a major rule are typically determined by deductive reasoning), much of the legal system follows an inductive path. The legal realists certainly realized this. Moreover, some of the most renowned examples of modern lawmaking, such as Llewellyn’s UCC, overtly used inductive reasoning to find governing legal principles from the realities of business practice, rather than the other way around. Yet, the inductive nature of much of the legal system is not limited to rule-creation. In particular, the process of finding facts at trial is inductive.

Not only do law and science share this inductive approach to reaching factual conclusions, they both use words like “prove” (or similar words such as “demonstrate”) to describe what they do. Yet, while people often speak of legal “proof” or scientific “proof” as though they are akin to mathematical or logical proof, this is certainly not the case. Mathematical proofs are either valid or flawed depending on the logical rigor by which one moves from the initial premise to the conclusion. Legal proof and scientific proof are quite different from mathematical proof; so much so that it is in many ways remarkable that they share the same key word—“proof”—with mathematical proof. Legal and scientific proof would perhaps more accurately be described as “fact finding” or, even better, “fact inferring.” These disciplines do not traverse a logical route from premise to conclusion but, rather, traverse a route from individual data points of information—fact evidence—to inferences about ultimate facts at issue. Both law and science are willing to draw such inferences, and both have standards to determine when such inferences may be drawn. The standards in the legal world are more formal, even if often stated casually or incompletely. Conversely, although the standards governing inference from data in scientific

disciplines are more a product of custom and informal consensus than of formal normative mechanisms, they are typically expressed more formally than the norms of the legal system.

Thus, in neither empirical science nor in litigation are facts “proven” in any rigorous sense of the word. Rather, facts relevant to the proposition sought to be proved or disproved (sometimes called the *probandum*) are considered and, if those facts lead the decision maker to conclude that the probability that the *probandum* is accurate is sufficiently high, the matter is pronounced proven (or demonstrated).⁷ This is an important point. In epidemiology, for example, the scientist typically does not profess knowledge of the underlying biological mechanism causing a particular medical problem. Instead, the epidemiologist utilizes various probabilistic techniques to examine the degree of association between the suspected “cause” and the undesired “effect.” Similarly, in litigation, unless the facts sought to be “proven” consist of an event that takes place in front of jury, the facts are not really proven in the sense of either actual knowledge or logical proof. Instead, other facts, such as the testimony of fact witnesses, are brought to the attention of the factfinder, and the factfinder must decide whether these other facts lead the factfinder to assess the probability of the *probandum* as being sufficiently high. Even in the case of eyewitness testimony, the process is ultimately probabilistic—the factfinder must assess the likelihood that the witness’s perceptions were accurate and that he or she is testifying truthfully about those perceptions.⁸ In other words, while those who report the results of epidemiological or legal fact-finding typically speak of “proof,” neither of those disciplines really engage in proof in the more formal sense of the term. Rather, they engage in something very different—the drawing of inferences from incomplete information. This observation is not a criticism of either law or science; rather, it is simply a description of what they do. It is this “proof” process, however, in which the role of experts is debated.

II. MISUNDERSTANDING THE DIFFERENCES BETWEEN LAW AND SCIENCE

Given the similarities between legal fact-finding and scientific fact-finding, it may appear that both disciplines are engaging in the same endeavor. Thus, scientists arguably should be able to speak easily to a legal audience about facts that have been found or demonstrated in science, with the concept of such demonstration neither gaining nor losing meaning. In my opinion, though, the

⁷ See, e.g., *Confidence in Probability*, *supra* note 1.

⁸ See, e.g., *Costs of Acceptability*, *supra* note 1, at 566.

apparent synonymy between legal proof and scientific proof is chimerical. Although the two types of “proof” share both superficially similar vocabulary and the basic goal of drawing inference from incomplete data, they approach these endeavors in fundamentally different ways.

Associate Justice Harry Blackmun seemed to realize this ten years ago in *Daubert*, when he wrote:

[T]here are important differences between the quest for truth in the courtroom and the quest for truth in the laboratory. Scientific conclusions are subject to perpetual revision. Law, on the other hand, must resolve disputes finally and quickly. The scientific project is advanced by broad and wide-ranging consideration of a multitude of hypotheses, for those that are incorrect will eventually be shown to be so, and that in itself is an advance. Conjectures that are probably wrong are of little use, however, in the project of reaching a quick, final, and binding legal judgment—often of great consequence—about a particular set of events in the past. We recognize that, in practice, a gatekeeping role for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations.⁹

In my view, Justice Blackmun got this exactly right, and then proceeded to get it exactly wrong. I believe that Justice Blackmun was correct as to the existence of “important differences between the quest for truth in the courtroom and the quest for truth in the laboratory.”¹⁰ More importantly, though, I think that his statement in *Daubert* turned the implications of those important differences on their heads. Justice Blackmun seemed to think that science sometimes relies on a fact-finding thought process that is insufficiently rigorous and inappropriate for the legal system and, thus, should be filtered out. I would argue that the opposite is true. Science, particularly empirical science that relies on statistical or other probabilistic methods, routinely uses filters that prevent its experts from reaching exactly the sort of opinions as to the truth of ultimate facts that should be utilized in a civil trial governed by the preponderance of the evidence rule.

In particular, there are three main differences between the process of formulating scientific conclusions and the process of formulating factual conclusions. First, the two systems have very different explicit and implicit value judgments concerning the

⁹ 509 U.S. at 596-97.

¹⁰ *Id.*

relative costs of errors. Standards of factual inference used in the world of science tend to assume that the costs associated with inaccurately determining that a proposition is proven or demonstrated—the costs of inaccurately discarding what many call the “null hypothesis”—are much greater than the costs of inaccurately declining to conclude that the proposition has been proven or demonstrated even though the proposition is true.¹¹ One of the premises of civil litigation, on the other hand, is that the societal cost of errors favoring the plaintiff is the same as the societal cost of errors favoring the defendant.¹² To put it another way, the preponderance of the evidence standard suggests that the civil litigation system ascribes essentially equal costs to inaccurately proclaiming a proposition to be demonstrated and to inaccurately declining to proclaim that the proposition has been demonstrated.

Second, science is comfortable with a fact inference process that creates a broad category of “suggested but not proven” facts. That is, scientific analysis of empirical data frequently leads to indeterminate conclusions that, if fully expressed, could be summarized as:

The information collected in this study is consistent with the proposition that A is correlated with B (and, indeed, does not support the proposition that A is not correlated with B) but, because the amount of information is relatively small, we cannot comfortably rule out the possibility that our observations are a product of chance rather than true correlation, and, therefore, we will not conclude at this time that the correlation exists. We must emphasize, though, that this does not mean that we conclude that the correlation does *not* exist.

Historically, the civil legal system has not utilized this sort of non-finding. In a trial, the facts are either proven or not—either the plaintiff wins or the defendant wins. While Scotland distinguishes between verdicts of “innocent” and “not proven” in criminal trials,¹³ there is no analog in American civil litigation. While the concept of equipoise exists in civil litigation, it is usually conceptualized as a narrow, knife-edged tie—a finding of a 50-50 balance in the probability assessment by the factfinder.¹⁴ Thus, in a sense, scientific

¹¹ See *Confidence in Probability*, *supra* note 1, at 412-18.

¹² See *generally Confidence in Probability*, *supra* note 1.

¹³ See www.scotland.gov.uk/faq/notprov.asp (last visited May 18, 2003).

¹⁴ I have written elsewhere that we should recognize that concepts of legal proof incorporate the idea of “confidence intervals” which would, in their operation, create a small category of “not proven,” even in civil litigation. See *supra* note 1. I note, however, that the costs of pro-plaintiff errors and pro-defendant errors in civil litigation should be equal and, thus, the legal “confidence intervals” would be narrower than those typically used in science. As a result, application of my theories

factfinders have available three possible answers to the question of whether A is associated with B: (i) no, (ii) yes, and (iii) “the evidence suggests yes, but we are not yet ready to proclaim that the answer is yes because the evidence could be an artifact of chance.” On the other hand, legal factfinders in the context of civil litigation have only two answers truly available: yes and no. The legal system, faced with the scientific trichotomy, typically, yet mistakenly, translates the third scientific answer into a simple “no.”

Third, the process of finding facts in the context of civil litigation is a “one-shot deal.” Either the plaintiff has proven her facts, or she has not, but she has only one chance to make her case—now and forever. This is assuredly not the approach of the world of science. In that world, a conclusion of “suggested but not proven” simply means that there is more work to be done (and often is used as the argument for funding that work). In other words, the scientific verdict of “not proven” can be translated to “stay tuned.” Law, however, does not stay tuned. After one trial, it tunes out.

In many ways, the points I have summarized above with respect to science have their strongest resonance in scientific disciplines such as epidemiology, in which the inferential reasoning is most overtly statistical and as to which the admissibility of expert testimony appears to be most frequently challenged. If the question at hand is whether A causes B (or, to avoid the various conundra associated with causation, whether A is associated with B), an epidemiologist does not examine the physical mechanism by which A and B might be connected (such as, for example, the cell biology of why exposure to benzene may be associated with increases in certain cancers). Rather, an epidemiologist examines raw data (whether collected by the epidemiologist or others) and performs statistical analysis. Typically, the epidemiologist will utilize both descriptive and inferential statistics with the output being a descriptive statement about the observed relationship between A and B in the data set being examined and an inferential statement that considers the possibility that the described facts could have occurred simply by chance in the absence of the relationship suggested by the observation. Most commonly, the epidemiologist will decline to characterize the data as showing a particular relationship between A and B unless the probability that the relationship could have occurred by chance even in the absence of the observed relationship is quite low. The most

would result in a somewhat broader concept of equipoise than the classical view, but it would still leave that category much narrower than the scientific “suggested but not demonstrated” category.

common threshold for such a probability is 5%. Sometimes, an epidemiologist or similar researcher will provide the descriptive statistics along with an analysis stating at what significance level (i.e., the probability that the results would occur by chance in the absence of the observed relationship) the descriptive statistics are “significant.” In such cases, the epidemiologist, or the user of his or her studies, will typically not place weight on observed results that are not significant at the 5%, or 0.05, level.¹⁵

Let us go through this a bit more carefully and specifically. Epidemiologists and similar scientific analysts of empirical data often describe their hypothesis testing process as determining whether they have disproved a “null hypothesis” of the absence of the point sought to be proved.¹⁶ In the case of a possible toxic substance, for example, the null hypothesis is that the substance is not toxic. After examining the evidence, an epidemiologist may pronounce the null hypothesis as disproven, or not. If it is disproven, the null hypothesis of no relationship is rejected. If it is not pronounced disproven, whether because the evidence supports the null hypothesis or because the evidence does not support it but the possibility that the evidence was created by chance rather than by the falsity of the null hypothesis is too high, null hypothesis is not rejected.

In other words, data can fail to “disprove” the null hypothesis for two very different reasons: (i) the evidence is consistent with null hypothesis, (ii) the evidence is inconsistent with the null hypothesis but, given the amount and nature of the data and of the null and alternative hypotheses, the epidemiologist or similar scientist cannot sufficiently rule out the possibility that the apparent inconsistency results solely from the luck of the draw.

Take a simple example—a public opinion poll. If 100 people are interviewed and 98 say that they will vote for the Democratic presidential candidate, that would certainly be sufficient to disprove the null hypothesis of a Republican victory.¹⁷ Similarly, if 98 of the interviewees indicate that they will vote for the Republican, that would clearly be consistent with the hypothesis of Republican victory. If, however, 52 of the 100 state that they will vote for the Democrat

¹⁵ See generally T. WONNACOTT & R. WONNACOTT, *INTRODUCTORY STATISTICS FOR BUSINESS AND ECONOMICS* § 19-1, at 539-43 (2d ed. 1977).

¹⁶ See H. BLALOCK, *SOCIAL STATISTICS* 113-14 (2d ed. 1972).

¹⁷ In this oversimplified example, I am, of course, painting with an even broader brush than normal. Obviously, a laundry list of assumptions—such as, for example, that the interviewees truly represent a random sample and that they will vote for the same candidate whom they indicated to the pollster they would support—should qualify this statement.

and 48 state that they will vote for the Republican, the pollster does not state that a Democratic victory is imminent; rather, the pollster will state that the race is “too close to call.” Thus, in both the second and third situations, the pollster will not reject the null hypothesis of a Republican victory. Yet, if one asked the pollster in the second and third cases what election result she would predict if she *had* to make a prediction based solely on the polling data, it seems obvious that those predictions would differ.

These two decisions to decline to reject the null hypothesis are very different. In the first case, the facts are consistent with the null hypothesis. If anything, the evidence makes us believe in the null hypothesis even more. In the second case, though, the facts are not what one would expect if the null hypothesis were true. Indeed, the facts should make one have some additional doubts about the null hypothesis. An epidemiologist declines to rule out the null hypothesis in the second case not because the facts support the null hypothesis—they do not—but because the inconsistent facts are of questionable robustness. The probability that they could occur just by coincidence if the null hypothesis is true is too high for comfort.

The first type of declining to reject the null hypothesis is akin to a verdict of innocent—the factfinder is confident of the truth of the null hypothesis. The second situation, though, is more akin to the Scottish verdict of not proven. That verdict does not involve a statement as to the ultimate truth or falsity of the *probandum* or the null hypothesis. It is, as Charles Nesson might say, a statement about the evidence, rather than the event.¹⁸ It is a statement that the evidence, in light of formal and informal burdens of persuasion, is insufficient to establish a particular hypothesis, even though the facts suggest the truth of that hypothesis.

Nonetheless, in either case, epidemiology pronounces that the null hypothesis has not been disproven. It is important to note that this is not very different from adjudication in our criminal justice system, in which the conclusions of “not proven” and “innocent” are combined in the single verdict of “not guilty.” There is one key difference, of course. A legal verdict of not guilty ends the matter once and for all.¹⁹ A scientific verdict of not proven is merely one step in a continuing effort to ascertain how the world works.

Again, the epidemiological verdict declining to rule out the null hypothesis does not mean that the null hypothesis is, indeed, true, or that the epidemiologist believes it to be true. It may be false, and the

¹⁸ See Nesson, *supra* note 3.

¹⁹ See U.S. CONST. amend. V.

epidemiologist may have a personal belief that it is (probably) false. The “not proven” statement means only that the epidemiologist declines to proclaim it false because, *within the scientific system in which the epidemiologist operates*, the cost of a wrongful proclamation that null hypothesis is disproven is higher than the cost of inaccurate silence as to the disproof of the null hypothesis when that hypothesis is, indeed, disproven.

Thus, there are situations in which, if a wealth maximizing epidemiologist were forced to place a wager at even odds as to whether the null hypothesis is true or false, without the option of declining to bet, she would place her bet on the proposition that the hypothesis is false, even though, in the system in which she operates, the conventional balancing of costs of errors leads to a statement that the invalidity of the null hypothesis is “not proven.”

To put this another way, epidemiologists and similar scientists are given three choices to describe the results of their inquiries: (i) the evidence is consistent with the null hypothesis; (ii) the evidence is inconsistent such that we are sufficiently confident the evidence did not result from chance; and (iii) evidence is inconsistent with the null hypothesis but too close to the line for the investigator to go out on a limb with an unnecessary proclamation.

If the scientific conclusion about the evidence is such that an assessment of the truth proposition under consideration is “too close to call,” science doesn’t stop. Rather, one more item is added to the pile of things about which we have not reached a firm conclusion, but to which we may return at some future date.

Let us now return to fact-finding in the context of civil litigation. In civil litigation, the vocabulary is similar, but the analysis is quite different. As in science, the rules of civil litigation do not demand that plaintiff “prove” its case in any real sense. Logical proof is not required. Neither is proof by the exclusion of all other possibilities, nor by the demonstration that the probability of the *probandum* equals or closely approaches 1.0. Rather, all that is required is that the plaintiff demonstrate that the *probandum* of his or her case is more likely than not to be true. In other words, all that is required is a probability greater than or equal to 0.5.

This preponderance of the evidence standard means that the legal system is willing to tolerate errors in either direction in the adjudication of civil cases. This is different from criminal prosecution, where the “beyond reasonable doubt” standard explicitly indicates the existence of a strong negative societal value for false convictions by indicating that society does not want to convict

the defendant unless the factfinders are virtually sure of the defendant's guilt and, concomitantly, society's willingness to acquit those who are "probably" guilty but not indubitably so. Thus, the implicit "error economics" of criminal prosecution are such that the cost of a false conviction is much higher than the cost of an inaccurate acquittal.²⁰

The error weighting of civil litigation is very different from this criminal justice proof model. By providing that the plaintiff wins if the probability of the *probandum* is even slightly greater than 0.5, and that the plaintiff loses even if that probability is slightly less than 0.5, the legal system has implicitly told us that the cost of a factually inaccurate verdict for the plaintiff is essentially the same as the cost of a factually inaccurate verdict for the defendant. After all, the system is telling us both that the plaintiff should prevail if the probability of the facts at issue is 0.51 (and, thus, there is a forty-nine percent chance that the verdict is factually inaccurate and that the cost of this error will be borne by the defendant) and the defendant should prevail even if the probability of the plaintiff's facts is 0.49 (and, thus, there is a forty-nine percent chance that the verdict is factually inaccurate and that the cost of this error will be borne by the plaintiff).

The civil litigation proof model is also very different from the implicit rules in scientific decision making described above. As noted, the process of scientific inquiry often results in evidence that does not support the null hypothesis, but which does not differ from that hypothesis to an extent that survives significance testing at the selected threshold—usually the 0.05 standard. In the world of science, this failure of the evidence countering the null hypothesis to survive significance testing does not disqualify the subject-matter from further inquiry, resulting in a permanent scientific decision in favor of the null hypothesis. Rather, it is an invitation of sorts for further research (and grant applications) to gather sufficient data to determine whether this will yield an inference against the null hypothesis that is significant.

In civil litigation, on the other hand, the verdicts and consequences, associated with the factfinders' conclusion that either

²⁰ See *Utah v. Weldon*, 314 P.2d 353 (Utah 1957) ("It is better that ten guilty go free, than that one innocent be punished."); see also *In re Winship*, 397 U.S. 358, 372 (1970) ("We do not view the social disutility of convicting an innocent man as the equivalent to the disutility of acquitting someone who is guilty."); D. Michael Risinger, *John Henry Wigmore, Johnny Lynn Old Chief, and "Legitimate Moral Force"—Keeping the Courtroom Safe for Heartstrings and Gore*, 49 HASTINGS L.J. 403, 442-45 (1998); Alexander Volokh, *Aside: Guilty Men*, 146 U. PENN. L. REV. 173 (1997).

the null hypothesis (absence of facts resulting in liability) is supported by the evidence or that matters are “too close to call”—i.e., in equipoise—are identical: the plaintiff loses and cannot relitigate the case. There is no opportunity to put off a final verdict until more data can, perhaps, be gathered in the future and no opportunity to relitigate if and when such data are available.

In light of the finality (against the plaintiff) associated with a conclusion that matters are too close to call, it is not surprising, then, that the “too close to call” region—i.e., equipoise—is traditionally conceptualized as razor thin in models of proof subject to the preponderance of the evidence standard. Most models hypothesize a very simple proof mechanism, in which the probability of the facts at issue is ascertained (whether with rigor or by seat-of-the-pants calculus), and if the probability of the *probandum* is less than 0.5 or exactly equal to 0.5 the plaintiff loses. If the plaintiff demonstrates a probability even a hair’s breadth greater than 0.5, however, the plaintiff prevails.²¹

To put this another way, by analogy to statistical terms, the civil litigation system tends to act as though the only relevant piece of probabilistic information that need be considered in determining whether the evidence supports a verdict for the plaintiff is the “point estimate” of the probability of the plaintiff’s facts. The concept of a “confidence interval” surrounding the point estimate, the choice of a significance level utilized to construct that interval, the size of the resulting interval, and whether that interval straddles the 0.5 threshold, are not considered in the current model.

As noted previously, several years ago I suggested a more nuanced model of civil proof, utilizing the statistical concept of confidence intervals to explain a somewhat broader area of equipoise in which the plaintiff does not prevail. I suggested that

this new model clarifies the practical content within the concept of equipoise. Situations in equipoise—in which neither party can meet the burden of persuasion—would include not only situations where the true probability is exactly 0.5 but also all situations in which the interval estimate of the probability of the facts supporting liability straddles 0.5. In any such case, the evidence provided by the parties would be insufficient to allow the factfinder to state with sufficient confidence that the probability that the facts support either party’s position exceeds 0.5.

²¹ See, e.g., John Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065 (1968); Lozowick et al., *Law and Quantitative Multivariate Analysis: An Encounter*, 66 MICH. L. REV. 1641 (1968); Zeisel, *Statistics as Legal Evidence*, 15 INT’L ENCYCLOPEDIA SOC. SCI. 246 (D. Sills ed. 1968).

Accordingly, in these cases it is important to determine which party will suffer for that mutual inability. Under the reformulated definition of the quantum of the burden of persuasion, allocation of the burden determines who will lose when the factfinder cannot determine with the requisite amount of confidence on which side of 0.5 the true probability lies.²²

Even in my model, though, the nether region of indeterminacy would be narrower than the corresponding region in the sciences. I argue strongly that the significance level that determines the size of confidence interval for legal proof should be determined by the values of civil litigation, in which pro-plaintiff errors and pro-defendant errors are assumed to carry equal societal costs, not the values of epidemiology and similar sciences, where inaccurate proclamations of fact are more disfavored than silence when the facts are true.²³

The upshot of all this is that, whether one follows the traditional model of civil proof or my alternative model, the process and structure of making legal fact-finding decisions from available evidence is quite different from the process and structure of making fact-finding decisions from available evidence in much of the scientific world. As noted earlier, not only do the two systems place different relative values on the different types of fact-finding errors, but they differ as to the number of opportunities for the proponent of an idea to make his or her case, and the consequences of indeterminate evidence. The legal system gives each plaintiff only one chance to prove his or her case, with a limited role for the concept of equipoise. Epidemiological and other scientific fact-finding, on the other hand, do not limit the opportunity to reach a conclusion to a single investigation, and build into the decision process a serious role for equipoise. Even apart from the other differences, this broad concept of equipoise, in which the null hypothesis is deemed not to have been disproven, casts a pro-null hypothesis bias into the assessment of whether any particular set of data leads to the conclusion that the probability of the null hypothesis is less than 0.5. Translating this bias to civil litigation would transform it into a pro-defendant bias.

It may be easier to see the difference between the two systems by describing them in terms of wagers. Models of scientific proof could

²² See *Confidence in Probability*, *supra* note 1, at 419.

²³ Compare *Conceptualizing Proof*, *supra* note 1, with David H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 CORNELL L. REV. 54 (1987).

be described as a wager in which the scientist/gambler has three choices: he or she may bet on the truth of the null hypothesis, bet on the falsity of the null hypothesis, or decline to place a bet at this time because the facts currently available render the matter too close to call with a high degree of confidence. Moreover, if the scientist/gambler opts for the too close to call option, he or she may place a bet at a later time if additional facts are developed that lead the scientist/gambler to a different conclusion. Legal proof governed by the preponderance of the evidence standard, on the other hand, as noted earlier, allows the legal factfinder to choose from only two possible wagers—the factfinder must bet, at even odds, either for or against the plaintiff's facts. Quite obviously, as factfinders forced to wager our own money, many of us would decline to bet if the evidence in a particular case left matters so close that it was difficult to separate meaningful inferences from the role of chance,²⁴ but the choice of betting is not allowed legal factfinders. Moreover, a legal factfinder cannot return and cast a different bet if additional evidence is developed in the future.

I have explored these differences in legal decision making and scientific decision making—differences that are profound despite the similarities of language used to describe them—because I fear that these differences are not recognized by courts that must make decisions about scientific evidence, particularly evidence that the scientific world examines through a probabilistic lens.

Indeed, some post-*Daubert* cases suggest that the deference to the methods of the world of scientific decision making is already beginning to have the undesirable effect of confusing scientific and legal values, and the norms that follow from them. An example is provided by a decision of the Supreme Court of Texas in one of the Bendectin cases, declining to give credence to epidemiological testimony utilizing a significance standard less rigorous than the traditional 0.05 standard. According to that court:

We think it unwise to depart from the methodology that is at present generally accepted among epidemiologists. *See generally* Bert Black, *The Supreme Court's View of Science: Has Daubert Exorcised the Certainty Demon?*, 15 *Cardozo L. Rev.* 2129, 2135 (1994) (stating that “[a]lmost all thoughtful scientists would agree . . . that [a significance level of five percent] is a reasonable general standard” (quoting Amicus Curiae Brief of Professor Alvan R. Feinstein in Support of Respondent at 16, *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 113 S. Ct. 2786, 125 L. Ed.

²⁴ See *Conceptualizing Proof*, *supra* note 1.

2d 469 (1993) (No. 92-102)). Accordingly, we should not widen the boundaries at which courts will acknowledge a statistically significant association beyond the 95% level to 90% or lower values.²⁵

The *Havner* decision was rendered in the context of sufficiency of evidence rather than admissibility of evidence. Nonetheless, the reasoning of the Supreme Court of Texas does not appear to be limited to that context. Indeed, the journey from a conclusion that use of a significance level other than 0.05 renders epidemiological inference insufficient on which to base fact-finding to a conclusion that the use of such a significance also renders it inadmissible is minuscule.²⁶

Moreover, it should be noted that the *Havner* court's analysis was not reached casually and without serious thought. The *Havner* opinion quotes the assertion of one of the plaintiff's experts that a significance level of 0.10 should be used²⁷ and the explanation of another expert that "[you] don't ever see [confidence intervals of 50% or 60%] in a scientific study because that means we're going to miss it a lot of times and [scientists] are not willing to take that risk."²⁸ In addition, the court noted that a pre-*Daubert* federal district court had concluded "that the scientific standard for determining causation is much stricter than the standard employed by the court and that confidence levels of 95%, 90%, or even 80% should not be required."²⁹

Thus, the *Havner* court was well-aware that legal decision making need not follow norms of scientific decision making, yet it decided nonetheless to yoke the two together. Accordingly, the spectre of barring the testimony of scientific experts because their threshold standard for willingness to conclude that a particular set of facts clearly exists. This is a part of a general failure of the legal system to understand the implications of the methods of scientific decision making and their effect on legal decision making.

Thus, I believe that there is reason to be seriously concerned that unthinking conformity to scientific decision making standards suggested by *Daubert* will have a serious substantive impact on civil litigation—especially litigation about facts that are incompletely

²⁵ Merrell Dow Pharmaceuticals, Inc. v. Havner, 953 S.W.2d 706, 724 (Texas 1997).

²⁶ See also Finley, *supra* note 6.

²⁷ 953 S.W.2d at 717-18.

²⁸ 953 S.W.2d at 724.

²⁹ *Id.* (quoting Longmore v. Merrell Dow Pharms., Inc., 737 F. Supp. 1117, 1119-20 (D. Idaho 1990)).

understood. Of course, admitting the testimony of scientific experts whose conclusions would not be reached by applying scientific standards of decision making under uncertainty will result in some errors. That goes without saying—in fact-finding under conditions of uncertainty, of course mistakes will be made. But civil litigation does not presume the same obsession with avoiding the error of premature conclusion as does science, which can always re-examine matters again in the future. The tragedy of equating legal fact-finding governed by the preponderance standard with scientific fact-finding is that the cost of the errors caused by this obsession—costs that are borne primarily by plaintiffs—will be discounted or ignored entirely. The result would be a *sub silentio* imposition of the values inherent in scientific decision making on the very different world of civil litigation.

Let us return to Justice Blackmun in *Daubert*. Justice Blackmun recognized these differences, but suggested that the implication is that law should filter out what science considers. As I see it, it is just the opposite—law should allow in conclusions that science filters out.

III. BLACK BOXES AND CLEAR BOXES

I think that the problem about which I am concerned may be caused in part by an implicit assumption by *Daubert* as to the nature of scientific and similar expert testimony. *Daubert* seems to assume a “black box” model of scientific and other expert testimony in which data is given to the expert (or developed by the expert himself or herself), the expert puts the data into a “black box” in which his or her analysis takes place out of sight of the factfinders, and out pops a conclusion. An example of that model might be provided by a proffered handwriting expert, who is given samples of the defendant’s handwriting and the document in question, analyzes them, and pronounces the document to have been written (or not to have been written) by the defendant. While the expert might recite the factors that lead him or her to the conclusion, the process by which those factors are weighed and balanced, as well as the justification for using those factors and not others, takes place in the expert’s mind.

Close examination of the *Daubert* criteria—testing to see if the expert’s theory can be falsified, whether the theory or technique has been subjected to peer review and publication, the known or potential rate of error of the scientific technique in question, and general acceptance within the scientific community—reveals that they are attuned to determining whether such a black box should be

trusted.

In addition, the *Daubert* criteria, and the gatekeeping function that they animate, is based on an assumption that judges will be better than jurors at identifying and excluding untrustworthy black boxes, and the experts who propose to use them. This paternalistic view of the court, as Joseph Sanders calls it,³⁰ would seem to depend for its validity on the accuracy of that assumption. Are judges indeed better than jurors in choosing when to disregard the “scientific” claims of charlatans and hired guns? If the answer to this question is no, the gatekeeper function would seem to rest on dubious assumptions. Moreover, if judges filter out more invalid black boxes, but at the cost of also filtering out more valid black boxes that are inaccurately seen as insufficient for admissibility under the *Daubert* criteria, one cannot say that the paternalistic role is justified without examining the cost of the factually mistaken verdicts that result from inaccurate exclusion of testimony. One cannot count only the positive value of the tuna captured in a efficient net; one must also count the cost of the innocent dolphins inadvertently destroyed by the same net.

Yet, as discussed in this essay, the sort of expertise that appears often in civil litigation is not “black box” expertise such as the testimony of handwriting experts. Rather, the testimony of epidemiologists and scientists using similar methods to extract empirical judgments from individualistic data, is based on expertise that takes place in a “clear box” in which the entire thought process of the expert can be monitored and assessed.

Imagine an epidemiologist who offers to testify as follows (transformed into an imaginary version both wordier than one would expect to see in actual testimony and somewhat oversimplified for ease of discussion) in a lawsuit claiming that a drug manufacturer should be liable for health problems related to high blood pressure allegedly caused by allergy medication sold by that manufacturer:

In a study of two groups of 100 otherwise similar women, one group taking the particular allergy medication and the other group taking no allergy medication, two of the women in the unmedicated group developed high blood pressure, while six of the women in the medicated group developed high blood pressure. This difference (2% versus 6%) suggests that more than half of the cases of high blood pressure in the medicated group resulted from the medication. Of course, though, this observed

³⁰ Joseph Sanders, *The Merits of the Paternalistic Justification for Restrictions on the Admissibility of Expert Evidence*, 33 SETON HALL L. REV. 881 (2003).

difference may not reflect a real difference resulting from the medication; even if the two groups are composed of otherwise equivalent women, it is possible that the medication has no effect on high blood pressure rates and the observed difference is simply a result of random fluctuations. As an epidemiologist, I am familiar with probability theory. Applying standard probabilistic models of binomial distributions, I can tell you that, if the medication had absolutely no effect on high blood pressure rates, we could expect to see the observed data (that are suggestive of such a relationship) approximately X% of the time. I would not proclaim in an academic paper the existence of a link between the medication and high blood pressure because X% is greater than 5% (a common epidemiological standard for making such a proclamation); rather, I would write that the link is suggested by the data but does not meet stringent scientific standards designed to minimize accidental proclamation of inaccurate findings—even at the cost of failing to proclaim accurate findings—and, therefore, additional studies should be done to determine if this suggested connection actually exists. But you have not asked me to present an academic paper; you have asked me to help you conclude, in a setting in which you have only one chance to get it right, and in which mistakes in either direction are deemed to be equally costly, whether there is such a link. In light of those two factors, which are not present in my academic milieu, I would not use the 5% significance level to filter out suggestions that emerge from the data. In this setting, I would set the threshold somewhat lower, which I believe more accurately reflects the balance of considerations in this setting. The observed link between the medication and high blood pressure would occur less than Y% of the time if there were no actual link. Thus, the key question is whether significance at the Y% level is sufficient to consider the observed relationship. In light of the purposes of civil litigation as I understand them, Y% would appear to be sufficient but, of course, I understand that this choice does not belong solely to me. Thus, while I would conclude, for the purposes of this factfinding endeavor, that the medication appears to be linked to high blood pressure, I urge you to decide the relative risks and costs of errors that flow from inaccurately finding a link and from inaccurately finding that there is no link, and determine a significance level appropriate for that task and compare that significance level to Y% in order to decide whether to accept my testimony as evidence of a link.

There is no “black box” in this testimony. All of the proffered testimony is transparent—the box is clear.³¹ The expertise consists of

³¹ Actually, I suppose some of the calculations, such as the binomial probabilities

a recital of data points already found or assumed, followed by a summary of what the data show and a step-by-step explication of what probability theory tells us about the likelihood that a different state of affairs. It concludes with an overt statement of the expert's opinion, accompanied by directions on how to utilize the data based on the factfinder's choice of significance level.

Examining the *Daubert* criteria in the context of scientific expertise of this sort reveals their utter inapplicability to it. The expert is proffering no theory that is subject to falsification. The techniques of inferential statistics have obviously been subject to peer review, but this is not necessarily the case with respect to the expert's exposition concerning the use of significance levels. The scientific technique of evaluating the evidence has no error rate—it is, itself, merely a statement about error rates! Finally, the testimony discloses on its face that the method the expert suggests for resolving the fact-finding endeavor in a legal setting is not generally accepted for use in the scientific community. I suggest that the *Daubert* criteria provide little or no guidance for this sort of testimony. I would further suggest that in the litigation described above, the proffered evidence should be admitted.

One might respond that this case is too easy. Of course it is. I constructed it to make a point. If, however, we make the case a bit more realistic by deleting much of the explanation from the expert's testimony, we would be left with the same core of fact-finding (“I would conclude, for the purposes of this fact-finding endeavor, that the medication appears to be linked to high blood pressure . . .”), reached by a scientifically nonstandard method of reasoning that is hard to justify under the *Daubert* tests. *Daubert* gatekeeping creates too great a risk that this sort of testimony—accurate as to facts, helpful as to probabilistic analysis, and with implicit value judgments in the choice of significance levels so easily illuminated by cross-examination that there is no need to protect the factfinders by exclusion of the testimony—will nonetheless be excluded.

A believer in the paternalistic model might still justify the gatekeeper role with the assertion that judges can better evaluate analyses of this sort than can juries, but I doubt that a strong case has been made for the truth of this assertion. Comparative studies of the likelihood of judges and juries falling for common statistical fallacies tell only part of the story. We must also consider the possibility of

referred to in the proffered testimony, emerge from a black box. These too, however, could be explained, albeit at greater length, so that all of the expert's reasoning would be transparent.

unwarranted hostility to statistical and probabilistic proof on the part of the judiciary. After all, the judge as gatekeeper in civil litigation is justifiable only if the cost of errors from wrongful judicial filtering-out of “good” testimony is less than cost of errors from the jury wrongfully believing “junk.” Since, in civil litigation, the societal cost of the two errors is presumed to be equal, this means that role of judges as filters enhances accuracy in fact-finding only if there will be fewer errors from wrongful judicial filtering than from wrongful jury gullibility.

A casual perusal of court decisions should rob anyone of the delusion that black robes provide immunity from that sort of error. More common are whoppers like that of Justice Powell, casually dismissing statistical evidence of racial disparities in death penalty sentencing in *McCleskey v. Kemp*.³²

Even a sophisticated multiple-regression analysis such as the Baldus study can only demonstrate a *risk* that the factor of race entered into some capital sentencing decisions and a necessarily lesser risk that race entered into any particular sentencing decision.”³²

Somehow, Justice Powell did not choose to point out that even the fact that an eyewitness testifies that she saw a particular event occur also only demonstrates that there is a risk (i.e., a probability) that the event has occurred. The factfinder must take into account (explicitly or implicitly) such factors as whether the witness had a motive to lie, the possibility that the witness, whether or not she had a motive to lie did, in fact, make statements at variance with her memory of the event, whether the witness sincerely misremembers events, and whether the witness’s eyesight or hearing led the witness to believe she saw or heard something that did not happen, all leading ultimately to a probabilistic judgment as to whether the event occurred. In other words, eyewitness testimony “can only demonstrate a *risk* that” the allegedly witnessed event occurred. Somehow, though, it seems unlikely that Justice Powell would have so glibly dismissed a case based on eyewitness testimony on the ground that the testimony established only a risk. Rather, it appears that he simply was reluctant to utilize overtly probabilistic testimony to make an important decision. Perhaps important societal values, not identified by Justice Powell, justify this reluctance³³. Minimizing the cost of errors (or at least optimizing them)—the touchstone of the preponderance standard in civil litigation—does not appear to justify

³² 481 U.S. 279, 292 n.7 (1987).

³³ See, e.g., Nesson, *supra* note 3.

it, though.

Accordingly, it is not clear to me that we are better off with judges as gatekeepers of “clear box” scientific testimony in civil cases. The *Daubert* criteria do not apply well, and the virtue of the judiciary as a filter of otherwise-relevant evidence in order to enhance the accuracy of verdicts is far from clear.