

2006

Judges as "Amateur Scientists"

David L. Faigman

UC Hastings College of the Law, faigmand@uchastings.edu

Follow this and additional works at: http://repository.uchastings.edu/faculty_scholarship



Part of the [Judges Commons](#), and the [Science and Technology Law Commons](#)

Recommended Citation

David L. Faigman, *Judges as "Amateur Scientists"*, 86 *B.U. L. Rev.* 1207 (2006).

Available at: http://repository.uchastings.edu/faculty_scholarship/153

This Article is brought to you for free and open access by UC Hastings Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of UC Hastings Scholarship Repository. For more information, please contact marcus@uchastings.edu.

Faculty Publications

UC Hastings College of the Law Library

Author: David L. Faigman

Source: Boston University Law Review

Citation: 86 B.U. L. REV. 1207 (2006).

Title: *Judges as "Amateur Scientists"*

Originally published in BOSTON UNIVERSITY LAW REVIEW. This article is reprinted with permission from BOSTON UNIVERSITY LAW REVIEW and Boston University.

JUDGES AS “AMATEUR SCIENTISTS”

DAVID L. FAIGMAN*

INTRODUCTION	1207
I. JUDGES (A.K.A. LAWYERS) AND THEIR COMPREHENSION OF SCIENCE.....	1209
II. THE SCIENCE EMBEDDED IN THE LAW	1211
A. <i>Error Rates Are Public Policy</i>	1211
B. <i>The Methods That Underlie the Statistics</i>	1216
C. <i>Bringing Scientific Research to Legal Doctrine</i>	1218
D. <i>Bringing the General Down to the Specific</i>	1220
CONCLUSION.....	1225

INTRODUCTION

The role of the judge in the twenty-first century cannot be understood without due consideration of the place of science and technology. Of particular concern must be judges’ lack of preparation for the times ahead. Science’s centrality to society’s welfare marked the twentieth century, both in terms of posing dire threats and promising salvation. While the importance of science to society is likely to expand geometrically in the century ahead, judges, on the whole, have little training in, knowledge of, or inclination to learn science. Scientifically illiterate judges pose a grave threat to the judiciary’s power and legitimacy. Like all ignorance, scientific illiteracy casts knowledge into the shadows, where only forms can be made out and detail is impossible to discern. Scientifically illiterate judges abdicate power and shun responsibility. In the twenty-first century, no judge will deserve the title if he or she does not know science.

The imperative for judges to know science can be reduced to a simple syllogism:

Applied science is almost invariably probabilistic and so cannot be used adequately without knowledge of probabilities and statistics;

Judges regularly rely on applied science as an integral part of lawmaking; therefore, it is incumbent on judges to understand probabilities and statistics.

* Distinguished Professor of Law, University of California, Hastings College of the Law. This Essay is adapted from remarks delivered on April 22, 2006, for a panel on “Judges and Social Science,” at a symposium sponsored by the Boston University School of Law on “The Role of the Judge in the Twenty-First Century.”

Although this syllogism is fairly obvious, the consequences that flow from it have been largely ignored and are immensely controversial. If the science the law uses is invariably uncertain – that is, statistical and methodological uncertainty is inevitably part of the legal calculus – then all legal decisions involving scientific evidence require management of the costs of error. To adequately manage these costs of error, which are policy judgments, judges must first understand the basic scientific methods and statistics used to generate the error rates.

Judges use science in a wide variety of legal contexts, both as a procedural matter in the areas of civil procedure and evidence, and as a substantive matter involving virtually all areas of the law, including criminal,¹ civil,² administrative,³ and constitutional.⁴ Perhaps the most explicit attempt to reckon with the realities of the interface between law and science came in the evidentiary context with *Daubert v. Merrell Dow Pharmaceuticals, Inc.*⁵ In *Daubert*, the Court held that judges are gatekeepers who must evaluate the methodological bases of proffered scientific evidence.⁶ In a subsequent case, *Kumho Tire Co. v. Carmichael*,⁷ the Court extended this injunction to all expert opinion, whether it be from rocket scientists or real estate agents.⁸ This mandate requires judges to have some understanding of research design and statistics, since they are required to examine the methods and principles underlying the expert's opinion. But this new responsibility, Chief Justice

¹ Perhaps the most abundant examples in the criminal law involve forensic science, see generally 4 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (2005-2006 ed.), and psychological syndromes, see generally 2 *id.*

² In civil cases, expert evidence is often presented, on topics ranging from damage assessment in personal injury cases to mass toxic torts. See Samuel R. Gross, *Expert Evidence*, 1991 WIS. L. REV. 1113, 1119-20.

³ See Wendy E. Wagner, *Importing Daubert to Administrative Agencies Through the Information Quality Act*, 12 J.L. & POL'Y 589, 591-92 (2004).

⁴ See I. BERNARD COHEN, SCIENCE AND THE FOUNDING FATHERS: SCIENCE IN THE POLITICAL THOUGHT OF JEFFERSON, FRANKLIN, ADAMS, AND MADISON 237 (1995); David L. Faigman, "Normative Constitutional Fact-Finding": Exploring the Empirical Component of Constitutional Interpretation, 139 U. PA. L. REV. 541, 545 (1991); Kenneth L. Karst, *Legislative Facts in Constitutional Litigation*, 1960 SUP. CT. REV. 75, 105; Laurence H. Tribe, *Seven Deadly Sins of Straining the Constitution Through a Pseudo-Scientific Sieve*, 36 HASTINGS L.J. 155, 156 (1984); cf. Henry P. Monaghan, *Constitutional Fact Review*, 85 COLUM. L. REV. 229, 264-65 (1985) (discussing appellate review of constitutional facts). See generally DAVID L. FAIGMAN, LABORATORY OF JUSTICE: THE SUPREME COURT'S 200-YEAR STRUGGLE TO INTEGRATE SCIENCE AND THE LAW (2004) [hereinafter FAIGMAN, LABORATORY OF JUSTICE].

⁵ 509 U.S. 579 (1993).

⁶ *Id.* at 592-93.

⁷ 526 U.S. 137 (1999).

⁸ *Id.* at 141.

Rehnquist complained, would require federal judges "to become amateur scientists."⁹ And indeed they must.

In the twenty-first century – and the sooner the better – judges have no choice but to become amateur scientists. The job requires it. This is true well beyond the narrow region of admissibility rules for expert evidence and includes all contexts in which empirical research is relevant to legal decision making. Legal decision makers simply cannot properly use scientific knowledge if they do not understand the premises of research methods and statistics. In this brief Essay, I hope to make two basic points. The first is to register concern regarding the current state of scientific comprehension within the legal community. The second is to illustrate how science and statistics are endemic to the judge's job and to outline this complex task judges face. Although the demands of the twenty-first century require judges to be amateur scientists, they are not well prepared to assume this role, nor will it be easily achieved. The question is no longer whether a judge should be an amateur scientist, but how he or she will become one.

I. JUDGES (A.K.A. LAWYERS) AND THEIR COMPREHENSION OF SCIENCE

By its nature, law requires judges to be generalists. While law is a distinct institution with its own goals and objectives, it constantly interacts with the world and institutions around it. The law is at bottom an empirical and practical profession. It receives input from a variety of sources, digests it through the legal process, and applies the output with the expectation of effecting some result. These steps require judges to have extraordinarily broad understanding of an assortment of professional disciplines. In constitutional law, for instance, history is essential, since original intent is a key authority for determining the Constitution's meaning. A court considering the original intent of the Second Amendment, for example, would have to wade through volumes of historical documents and debate.¹⁰ And unless a true consensus among historians existed, the court could not defer to experts on this matter. The judges would, in effect, be operating as "amateur historians." This would simply be a requirement of the job and no one would seriously doubt its necessity. Moreover, it would be extraordinarily disconcerting if any judge decried the prospect of being an amateur historian or professed ignorance of the subject. Similarly, judges must sometimes be amateur political theorists, economists, linguists, and sociologists – all without complaint.

Yet when it comes to science, and particularly statistics, judges pause and sputter, wondering whether it is truly part of their responsibility to know the details of scientific methods. In addition to Chief Justice Rehnquist's swipe at "amateur scientists," many judges have raised concerns about their collective

⁹ *Daubert*, 509 U.S. at 601 (Rehnquist, C.J., concurring in part and dissenting in part).

¹⁰ See, e.g., Sanford Levinson, *The Embarrassing Second Amendment*, 99 YALE L.J. 637, 646 (1989).

ability or desire to learn science.¹¹ Indeed, judges sometimes proudly declare their ignorance of the subject, cavalierly stating that knowledge of science and statistics is not necessary to legal analysis. In *Craig v. Boren*,¹² the Court applied intermediate scrutiny to strike down an Oklahoma law that prohibited men under twenty-one years of age from purchasing “nonintoxicating” 3.2% beer while permitting women over eighteen years of age to buy it.¹³ Oklahoma had justified the discrimination on the basis of statistical studies indicating that young men account for a disproportionate share of drivers arrested for driving while intoxicated.¹⁴ Justice Brennan initially dismissed the studies as methodologically weak and of little use.¹⁵ But rather than rely on the cogency of his statistical critique, Brennan added an apologia:

There is no reason to belabor this line of analysis. It is unrealistic to expect either members of the judiciary or state officials to be well versed in the rigors of experimental or statistical technique. But this merely illustrates that proving broad sociological propositions by statistics is a dubious business, and one that inevitably is in tension with the normative philosophy that underlies the Equal Protection Clause.¹⁶

This is a remarkable statement in so many ways. Imagine substituting “historical” for “experimental or statistical.” Would it be “unrealistic to expect the judiciary to be well versed in the rigors of historical technique”? Would not such a proclamation of judicial ignorance be front page news? Moreover, the Court applied intermediate scrutiny, which required state officials to prove that the liquor law was “substantially related to achievement of [important governmental] objectives.”¹⁷ The statistical studies provided, at least in part, this proof. In effect, the Supreme Court struck down a law legitimately enacted by Oklahoma on the basis that the State provided insufficient justification for the law, though the Court eschewed knowing the experimental or statistical bases for that justification. This display of scientific disinterest

¹¹ See, e.g., *Daubert v. Merrell Dow Pharms., Inc.*, 43 F.3d 1311, 1316 (9th Cir. 1995) (“As we read the Supreme Court’s teaching in *Daubert*, therefore, though we are largely untrained in science and certainly no match for any of the witnesses whose testimony we are reviewing, it is our responsibility to determine whether those experts’ proposed testimony amounts to ‘scientific knowledge,’ constitutes ‘good science,’ and was ‘derived by the scientific method.’”); *United States v. Cline*, 188 F. Supp. 2d 1287, 1294 (D. Kan. 2002) (“Those of a ‘scientific’ bent certainly can take issue with whether the judges and lawyers have the education or training to engage in ‘scientific’ testing . . .”).

¹² 429 U.S. 190 (1976).

¹³ *Id.* at 210.

¹⁴ *Id.* at 200-01.

¹⁵ *Id.* at 201-03.

¹⁶ *Id.* at 204.

¹⁷ *Id.* at 197.

gives new meaning to the "counter-majoritarian difficulty."¹⁸ If the Court's power depends on its judgment rather than control of the purse or sword, this failure to provide an explanation for invalidating a duly enacted state law undermines the Court's legitimacy.

An assortment of reasons probably explains the judiciary's general ignorance of science and continued reluctance to learn much about it. The primary reason, however, appears to be fairly simple. Lawyers, of which judges are merely a subset, generally lack good training in the methods of science. Most lawyers do not speak the language of science. Lawyers and scientists come from different worlds of education and experience. Indeed, the sorting of professionals into highly compartmentalized categories begins as early as elementary school and is largely complete by college. Students with aptitude for and interest in math and science gravitate toward careers in medicine, engineering, physics, biology, statistics, and the like. Students not so inclined can avoid real science classes almost entirely or slip through with "artsy" versions of science courses. Many who have spent much of their educational life avoiding math and science become lawyers.

In fact, in my experience the typical lawyer is not merely ignorant of science, but rather has an affirmative aversion to it. Nothing will put a class of law students to sleep faster than putting numbers on the chalkboard. A bell curve makes their eyes glaze over. A minor equation or two, or calculating a standard deviation, renders law students unconscious; and a more complicated regression analysis induces a deep coma. The average law student's attitude toward mathematics is the same as Huckleberry Finn's:

I had been to school most all the time, and could spell, and read, and write just a little, and could say the multiplication table up to six times seven is thirty-five, and I don't reckon I could ever get any further than that if I was to live forever. I don't take no stock in mathematics, anyway.¹⁹

Judges, however, no longer have the luxury – if they ever did – of ignoring the imperatives of science. They must begin to take "stock in mathematics." As the next section illustrates, the basic judicial task demands that judges know the rigors of experimental and statistical techniques. This will be no easy task.

II. THE SCIENCE EMBEDDED IN THE LAW

A. *Error Rates Are Public Policy*

Those lacking scientific training make the critical mistake of thinking of scientific knowledge as categorical or certain. But brief reflection by even the most scientifically naive should dispel this notion. Applied science, in

¹⁸ See ALEXANDER M. BICKEL, *THE LEAST DANGEROUS BRANCH: THE SUPREME COURT AT THE BAR OF POLITICS* 16-23 (1962) (coining "counter-majoritarian difficulty" as a term for the problem of reconciling judicial review with democratic principles).

¹⁹ MARK TWAIN, *ADVENTURES OF HUCKLEBERRY FINN* 21 (Random House 1996) (1885).

particular, is almost invariably probabilistic in nature. Anyone who has checked a weather forecast knows this basic lesson. Yet courts regularly ignore this component of scientific information. The Supreme Court is particularly guilty of adopting this myth of scientific certitude. In *Roe v. Wade*,²⁰ for instance, the Court held that a state's interest in the potential life of the fetus became "compelling" – and thus sufficient to ban abortions – at "viability."²¹ At viability, the Court explained, a fetus "has the capability of meaningful life outside the mother's womb."²² But viability is, in fact, a statistical prediction of survivability that varies widely over many weeks during the late second and early third trimesters.²³ The Court never so much as mentioned the statistics associated with this new bedrock of constitutional law. Ignoring the statistical bases for empirical statements, however, does not make them any less probabilistic. Instead, it buries the policy choices inherent in the shifting probabilities and allows the Court to shun responsibility for deciding the tough cases around the margins. The empirical uncertainties of factual statements are as important as the statements themselves and should be part of the legal calculus.

Suppose forecasters at the National Weather Service tell the Governor of Florida that computer models predict with 95% confidence that there is a 35% likelihood of a Category 5 hurricane hitting Miami. Should the governor order a mandatory evacuation of the city? Or suppose that the defendant in a civil commitment hearing has a 35% likelihood of committing a sexually violent act in the future, and that the actuarial model used to make this prediction is statistically significant at the 95% level of confidence. Is this probability of future violence constitutionally sufficient to deprive a defendant – who is currently accused of no wrong – of his liberty?²⁴ Naturally, a governor or judge making these decisions will wish to consider the consequences of making a mistake in light of the likelihood of such error, and balance this calculus with the probability of and the benefits from making the correct decision. Even if the basic statistics are accepted as a given, this is not a simple exercise.

In these two examples, and countless others that policymakers and judges confront every day, error can be readily divided into two types: false positive and false negative. A false positive error in the first scenario would result in the mistaken evacuation of Miami, and a false negative would result in the

²⁰ 410 U.S. 113 (1973).

²¹ *Id.* at 163.

²² *Id.*

²³ See FAIGMAN, LABORATORY OF JUSTICE, *supra* note 4, at 220.

²⁴ Under applicable constitutional law, a defendant cannot be civilly committed unless found to be both (1) mentally abnormal and (2) dangerous. *Kansas v. Crane*, 534 U.S. 407, 409-14 (2002); *Kansas v. Hendricks*, 521 U.S. 346, 358 (1997). Unfortunately, neither of these terms have been well defined by the Court. See 2 FAIGMAN ET AL., *supra* note 1, § 13:6.

failure to evacuate Miami before a major hurricane hit. Similarly, the mistaken commitment of someone who would not have been violent is a false positive, and the failure to commit someone who will be violent is a false negative. Even brief reflection reveals the very different consequences that flow from each kind of mistake. Moreover, there are two possible correct decisions, true positives and true negatives. These also present very different benefits to the policymaker or judge. Table 1 depicts the basic table that results from these alternatives.

Table 1

Ground Truth/Actual Result

		Ground Truth/Actual Result	
		Yes	No
Legal/Policy Decision	Yes	True Positive	False Positive
	No	False Negative	True Negative

The Yes/No outcomes refer to the answers provided to the empirical question of interest. (For example, "will a Category 5 hurricane hit Miami?"; "will the defendant commit future acts of violence?")

As noted, the four possible outcomes of a single decision, two correct and two incorrect, present widely varying consequences that must be evaluated in light of the statistical likelihood that each outcome will occur. Table 2 provides a glimpse of the difficulty of the decision in the context of civil commitments of sexually violent predators ("SVPs"). In order to keep the example simple, the table illustrates the consequences of using a violence-screening test with what is today an unrealistically high accuracy rate of 0.90 and a relatively high base-rate of 0.50, in order to illustrate what might be the best argument for using such a test.²⁵ The sensitivity of the test is set at 0.80, resulting in the pass/fail threshold identifying 80% of those who would actually be violent.²⁶ Making the test more sensitive, of course, brings with it the result that more false positives will occur. In addition, the cells list some of

²⁵ See generally 2 FAIGMAN ET AL., *supra* note 1, §§ 13:20-51.

²⁶ Sensitivity and selectivity are associated with the concepts of true and false positives. Sensitivity refers to the chance of testing positive among those who are positive, and selectivity (also referred to as specificity) refers to the chance of testing negative among those who are negative.

the consequences that likely follow the respective outcome and thus are subjects of consideration for the decision maker.²⁷

Table 2
Predictions of Violence in SVP Cases

Expected Results of a Violence-Screening Test with an Accuracy Index of 0.90 in a Hypothetical Population of 1000 Examinees with a Base-Rate for Violence of 0.50 (Sensitivity Set at 0.80).

		Ground Truth/Actual Outcome	
		Violent	Not Violent
Legal Judgment	Violent	True Positive (400) <ul style="list-style-type: none"> • Avoid harm to third person • Incarcerate violent person • Possibly provide treatment to defendant • Costs of incarceration 	False Positive (80) <ul style="list-style-type: none"> • Deprivation of liberty • Costs of incarceration • Potential loss of productive member of society
	Not Violent	False Negative (100) <ul style="list-style-type: none"> • Allow violent person to go free • Violent acts committed in community • Avoid costs of incarceration • Negative publicity possibly resulting in loss of elected judgeship 	True Negative (420) <ul style="list-style-type: none"> • Give liberty to non-violent person • Avoid costs of incarceration • Avoid loss of productive member of society

Given the assumptions underlying Table 2, a judge would have to weigh the costs and benefits of the decision whether or not to incarcerate the defendant in light of the consequences that flow from each of the four possible outcomes. Complicating matters further, each “consequence” identified in Table 2 has likelihood statistics associated with it. Nonetheless, the basic task established in the SVP context is clear. Given the values inherent in the Constitution (i.e., due process, ex post facto, and double jeopardy), is it acceptable to incarcerate someone given these statistics? It does not matter that the numbers selected in Table 2 are speculative. In fact, if anything, the numbers used in this example strongly overstate the statistical case for predictions of violence, since accuracy rates are well below 0.90 in practice. Indeed, one court has accepted accuracy

²⁷ For a good overview of the many factors that contribute to policy decisions in a situation analogous to predicting violence, see COMM. TO REVIEW THE SCIENTIFIC EVIDENCE ON THE POLYGRAPH, NAT’L RESEARCH COUNCIL, THE POLYGRAPH AND LIE DETECTION 29-61 (2003).

rates below 0.50.²⁸ The reader is invited to run any set of numbers he or she prefers. In fact, that is the point. The question presented in SVP cases is what ratios would be minimally adequate to guarantee defendants in these cases their basic rights under the Constitution. The Court did not even mention this analysis in two major decisions on the subject.²⁹

Yet, as a general matter, judges and lawyers are well acquainted with the basic task of allocating costs of error. Procedural mechanisms such as burdens of production and proof are directed at managing the costs of error in different substantive legal contexts. In criminal cases, for example, in which false positives pose the greatest risks, the "beyond a reasonable doubt" standard applies.³⁰ This stringent standard generally reflects the well-known colloquialism that it is better to let ten guilty men go free than to convict one innocent man.³¹ This colloquialism simply states the legally acceptable ratio between false negatives and false positives. Implicit in this statement is that the ratio selected between these two errors inevitably affects the power of the criminal trial process to identify true positives and true negatives. In particular, all things being equal, reducing the number of false positives will also reduce identification of true positives – a real and substantial cost to society.

These kinds of statistical statements, however, do not correspond neatly to ordinary conceptions of the burden of proof for at least a couple of reasons. First of all, the analogy itself may not be apt. No clear relationship exists between burdens of proof and probability estimates. We can say generally that the "preponderance of the evidence" standard is akin to a probability estimate greater than 50%, but such a description is ambiguous and misleading. Both the probability estimate and the burden of proof are subtle and complex statements, culturally tied to statistics and law, respectively, with only some overlap in meaning. Burdens of proof in law are not quantified and, at best, reflect an intuitive judgment regarding the degree of proof needed in light of the gravity of the decision to be made. They operate as rough and ready

²⁸ *People v. Ghilotti*, 44 P.3d 949, 973 (Cal. 2002).

²⁹ The Court did not discuss the inherent uncertainties of predictions of violence in either *Hendricks* or *Crane*. State courts, however, have considered the subject in some detail in a series of cases interpreting statutes requiring that a person only be civilly confined if he is "likely" to engage in sexual violence. See, e.g., *State v. Ehrlich (In re Leon G.)*, 59 P.3d 779, 787 (Ariz. 2002) (holding that "likely" means "highly probable"); *Ghilotti*, 44 P.3d at 968 (holding that "likely" means "a substantial danger – that is, a serious, well-founded risk," but at the same time "does not mean the risk of reoffense must be higher than 50 percent"); *Commonwealth v. Reese*, No. 00-0181-B, 2001 WL 359954, at *15 (Mass. Super. Ct. Apr. 5, 2001) (defining "likely" to mean "at least more likely than not"), *vacated*, 781 N.E.2d 1225 (Mass. 2003); *In re Linehan*, 557 N.W.2d 171, 180 (Minn. 1996) (holding that "likely" means "highly likely").

³⁰ See Scott E. Sundby, *The Reasonable Doubt Rule and the Meaning of Innocence*, 40 HASTINGS L.J. 457, 458 (1989).

³¹ *Id.* at 460.

guidelines and are not intended to have true quantitative correlates. In science, by comparison, probability estimates are objective statements, albeit packed with a wide assortment of explicit and implicit assumptions. If the underlying assumptions hold, the probability estimate is set forth as an accurate statement about some specifically defined empirical proposition.

In addition, the burden of proof operates on the ultimate question of fact, whereas scientific evidence tends to be relevant to one or more individual component facts of the ultimate decision. In SVP cases, for instance, two empirically based determinations are necessary for commitment under the Constitution: (1) mental abnormality and (2) likelihood of future violence.³² I have so far only discussed the complexity of the second, but the first factor similarly presents profound empirical challenges. The Court defined mental abnormality as “serious difficulty in controlling behavior,”³³ a neuro-psycho-physiological fact of some difficulty. In many cases, moreover, the scientific research does not speak specifically to the legal issue in dispute. For example, DNA profiling evidence provides a probability statement regarding the likelihood that another match would occur in a random sample of the population.³⁴ It does not provide the probability of guilt, which is the principal concern of the burden of proof. DNA profiling, like other scientific evidence, must be integrated into the other evidence available and, in combination, can be said to support or not support the applicable burden of proof. Rarely, if ever, will the probabilities of empirical research directly correspond to legal burdens of proof.

B. *The Methods That Underlie the Statistics*

Complicating matters further, the complex statistical statements discussed above are only the tip of the empirical iceberg. In practice, statistics are only as good as the research methods used to generate them. Junk research methods produce junk statistics. Virtually every context in which scientific research is employed in the law presents issues involving the quality and quantity of the underlying research. Consider, for example, the recently completed Illinois study comparing the traditional simultaneous lineup (sometimes referred to as a “six-pack,” since six “suspects” are displayed to the witness at one time) to an alternative procedure whereby witnesses view one suspect (photograph or person) at a time, known as a “sequential lineup.”³⁵ Most of the research conducted in this area has been done in the laboratory and has involved

³² See *supra* note 24.

³³ *Kansas v. Crane*, 534 U.S. 407, 413 (2002).

³⁴ See 4 FAIGMAN ET AL., *supra* note 1, § 32:49.

³⁵ See generally SHERI H. MECKLENBURG, ILL. STATE POLICE, REPORT TO THE LEGISLATURE OF THE STATE OF ILLINOIS: THE ILLINOIS PILOT PROGRAM ON SEQUENTIAL DOUBLE-BLIND IDENTIFICATION PROCEDURES (2006) [hereinafter ILLINOIS REPORT], available at <http://www.chicagopolice.org/IL%20Pilot%20on%20Eyewitness%20ID.pdf>.

contrived circumstances and college-age subjects.³⁶ The Illinois research was a "field study," in which the two lineup procedures were compared in real cases.³⁷ Field studies have the advantage of real world verisimilitude, but suffer the messiness and potential confounds of actual practice.

Laboratory research in this area had seemed to settle the question regarding the advantages and disadvantages of the two procedures. This research generally indicated that sequential lineups were less sensitive than simultaneous lineups, meaning that they resulted in fewer accurate identifications (i.e., fewer true positives), but also fewer misidentifications of subjects (i.e., fewer false positives).³⁸ One theoretical explanation for this was that witnesses might be inclined to pick "the best suspect" from a comparative analysis of a simultaneous lineup, whereas sequential lineups avoided such comparative judgments by requiring a yes or no decision with each picture or person shown.³⁹

The Illinois field study, however, did not replicate the findings from the laboratory studies. Indeed, the Illinois research found that sequential procedures "resulted in an overall *higher* rate of known false identifications than did the simultaneous lineups."⁴⁰ Sequential lineups resulted in a 9.2% rate of false identifications, while simultaneous lineups had a 2.8% false positive rate.⁴¹ Additionally, the simultaneous procedures resulted in a higher rate of true identifications than did the sequential lineups, which was consistent with laboratory findings. Witnesses who viewed simultaneous lineups identified the suspect 59.9% of the time, whereas those who viewed sequential lineups identified the suspect 45% of the time.⁴² Hence, in a test of the hypothesis in the field, simultaneous lineups appeared to both maximize the identification of perpetrators ("true positives") and minimize the misidentification of innocents ("false positives").⁴³ What should one make of this research?

³⁶ *Id.* at 4 n.5.

³⁷ *Id.* at ii.

³⁸ Gary L. Wells & Elizabeth A. Olson, *Eyewitness Testimony*, 54 ANN. REV. PSYCHOL. 277, 288 (2003). For a meta-analytic comparison of sequential and simultaneous line-ups, see generally Nancy Steblay et al., *Eyewitness Accuracy Rates in Sequential and Simultaneous Lineup Presentations: A Meta-Analytic Comparison*, 25 LAW & HUM. BEHAV. 459 (2001).

³⁹ Wells & Olson, *supra* note 38, at 288. Another possible explanation for differences between lineup procedures is that subjects use a different selection criterion for sequential lineups than they use for simultaneous lineups. See Christian A. Meissner et al., *Eyewitness Decisions in Simultaneous and Sequential Lineups: A Dual-Process Signal Detection Theory Analysis*, 33 MEMORY & COGNITION 783, 790 (2005).

⁴⁰ ILLINOIS REPORT, *supra* note 35, at iv.

⁴¹ *Id.* at 38 tbl.3.a.

⁴² *Id.*

⁴³ *Id.* at 61.

As noted, the weight of the statistics depends on the strength of the research methods used. The fact that the Illinois study was a field study gives it greater power in some respects while undermining it in others. The concerns with the Illinois study, however, are not specifically associated with it being field research. Although many complaints might be made, as is true with all empirical research, two in particular undermine the value of any lessons that might be drawn from the statistics Illinois obtained. First, the measure of success for identifications was whether the *suspect* was correctly identified, not whether the perpetrator was correctly identified.⁴⁴ The study made no attempt to establish the veracity of the suspects as the perpetrators, even in a subset of the sample where DNA or other definitive evidence might have been available.⁴⁵ This is especially problematic because a high positive identification rate should be found if simultaneous lineups lead to comparative selections among the lineup participants, as critics contend.⁴⁶ However, this defect would not explain why simultaneous lineups also had *lower* false positive rates. A second defect in the research method might explain this latter finding. The Illinois study compared *blind* sequential lineup procedures to *non-blind* simultaneous lineups.⁴⁷ Many researchers believe that when administrators of lineups know the suspect's identity, there is a risk that they will provide subtle (or not so subtle) clues to the witness as to which one is the "correct" choice.⁴⁸ Failure to blind the administrator in the simultaneous lineups while blinding administrators of the sequential lineups is clearly a huge confound that might explain the lack of correspondence between the field research and the laboratory studies.

C. *Bringing Scientific Research to Legal Doctrine*

As the Illinois example makes clear, statistics cannot be viewed independently from the research methods used to generate them. Even when adequate research methods have generated robust statistics, this empirical work must be applied to some policy decision. What are the policy implications – either for a state legislature or a court considering the due process implications of lineup procedures – of research comparing simultaneous and sequential

⁴⁴ See *id.* at iii.

⁴⁵ Although there was no concerted effort to establish that the suspects were in fact the perpetrators, the researchers did report that "many suspect identifications recorded in the Illinois Pilot Program were corroborated by independent evidence." *Id.*

⁴⁶ See, e.g., Wells & Olson, *supra* note 38, at 288.

⁴⁷ See ILLINOIS REPORT, *supra* note 35, at v.

⁴⁸ Although it is widely believed that lineup administrators sometimes give implicit or explicit clues to witnesses regarding the "correct" choice, research has yet to fully demonstrate this hypothesis. See Ryann M. Haw & Ronald P. Fisher, *Effects of Administrator-Witness Contact on Eyewitness Identification Accuracy*, 89 J. APPLIED PSYCHOL. 1106, 1109-10 (2004).

lineups? Even if the research studies were relatively clear, the policy choices are not.

Assume that the Illinois study's results are an artifact of the confounding variables and that the laboratory work best describes the policy choice presented by sequential and simultaneous lineups. Suppose, in particular, that simultaneous lineups are more sensitive than sequential lineups, thus producing more true positive identifications, but also more false positives. This might lead some to advocate sequential lineups on the basis that, on balance, it is much worse to convict the innocent than to free the guilty. Others might advocate simultaneous lineups, arguing that they are a more powerful tool for law enforcement and that other evidence or cross-examination at trial can discern incorrect identifications.

A third possibility exists which does not require choosing one procedure over the other.⁴⁹ The data in this example support the proposition that simultaneous lineups produce more positive identifications, of both the true and false varieties. But in some cases false positives may be less worrisome, and the more powerful (albeit less discerning) test might be preferable. For instance, in sexual assault cases in which forensic DNA evidence is available, we should prefer a lineup procedure that would maximize positive identifications because subsequent DNA testing will clear anyone who is wrongly accused. There are certainly significant costs associated with wrongful accusations that result in arrest and DNA testing, but these costs are relatively minor compared to wrongful convictions (i.e., false positives) and the failure to apprehend the perpetrators (i.e., false negatives). In contrast, in cases in which eyewitness identification is likely to be the best or only substantial evidence available, sequential lineups might be a better policy choice. In those cases, reduction of false identifications might very well be of paramount importance.

In many respects, the legal issues presented by broad policy determinations, like the choice between simultaneous and sequential lineups, are the simplest uses of scientific research. Of greater prevalence in the law are the conceptually more difficult problems associated with applying general scientific research to particular cases. As difficult as it is to say whether, on balance, simultaneous or sequential lineups are to be preferred, determining from the data whether a particular lineup resulted in a true or false positive is a greater problem by several orders of magnitude. It is these sorts of judgments that are made daily in civil and criminal trials, often based on, or informed by, scientific research very much like that involved in the debate over lineups.

⁴⁹ Still another possibility, always available, is to maintain the status quo until more data are collected. See Amina Memon & Fiona Gabbert, *Unravelling the Effects of Sequential Presentation in Culprit-Present Lineups*, 17 *APPLIED COGNITIVE PSYCHOL.* 703, 712-13 (2003).

D. *Bringing the General Down to the Specific*

Virtually all scientific research is done at the population level, yet it is often used to make statements about particular cases. As I have put it elsewhere, “[w]hile science attempts to discover the universals hiding among the particulars, trial courts attempt to discover the particulars hiding among the universals.”⁵⁰ Scientists usually study variables at the population level and design most of their methodological and statistical tools for this kind of work. The trial process, in contrast, usually concerns whether a particular case is an instance of the general phenomenon.

Consider the hypothesis that secondhand smoke causes lung cancer. Research provides general probability statements regarding whether secondhand smoke causes lung cancer and the strength of any such relationship.⁵¹ As noted above, these probability statements ordinarily are based upon research methods of varying kinds and of various quality. In the case of secondhand smoke, the underlying research foundation could be made up of an assortment of methods, including clinical anecdotes, toxicological experiments (*in vitro* and *in vivo*), and epidemiological studies. Each of these research paradigms also can vary dramatically in terms of the quality of the methods used. Different methods present diverse opportunities to commit mistakes. Error in science could be a product of statistical variation (i.e., chance differences) or a consequence of dozens of possible methodological errors, such as coding errors, hypothesis guessing by subjects, recall bias, experimental error, and even scientific fraud. And most of these errors, unlike statistical variation, cannot be quantified.

Courts have recognized some of the difficulties inherent in employing general scientific data to reach conclusions about specific cases.⁵² This recognition has largely occurred in medical causation cases in which courts routinely distinguish between “general causation” and “specific causation.”⁵³ Not all science is engaged in describing cause and effect relationships, however, so “general causation” and “specific causation” are subcategories of what might be better termed “general propositions” and “specific applications.” Sometimes general scientific propositions will be stated in terms of causation, but very often they will be associational, technical, or descriptive. Specific application refers to the determination of whether a particular case is an instance, use, or example of general propositions that are supported by adequate research. Because ordinary science operates at the general level of descriptive and inferential statistics, it readily can be employed to determine general propositions. In the example of secondhand smoke, one

⁵⁰ DAVID L. FAIGMAN, *LEGAL ALCHEMY: THE USE AND MISUSE OF SCIENCE IN THE LAW* 69 (1999).

⁵¹ 3 FAIGMAN ET AL., *supra* note 1, § 27:57.

⁵² *See id.* § 23:1.

⁵³ *See id.* § 23:2.

would expect, and indeed would find, considerable research on this general question.⁵⁴ In the courtroom, however, the ultimate question is whether secondhand exposure to cigarettes caused the plaintiff’s lung cancer. This requires proof not only that secondhand smoke *could cause* lung cancer but also that *some other cause was not responsible* for the plaintiff’s lung cancer. Although courts have recognized the challenge presented by applying general science to specific cases, they have yet to fully accept the complexity of the task. In fact, scientists themselves have largely failed to explore the many dimensions of this matter.

As an initial matter, it is a relatively straightforward exercise to assess the validity of general propositions, that is to say whether various kinds of research converge sufficiently to give scientists enough confidence to say that exposure to *X* causes (or is associated with) condition *Y*. Indeed, this sort of exercise is a staple of training in science. Ideally, a community of researchers study hypotheses using a wide range of methodological and statistical tools. Researchers who study factors that interfere with eyewitness reliability, for example, employ a wide assortment of methodologies, including anecdotal reports, data from DNA exoneration cases, field studies, and laboratory experiments. If these differing methods all point in the same direction, then some general conclusions might be made regarding the phenomenon of interest. When they do not point in the same direction, however, the task is complicated greatly, if not made impossible, until more research is completed. Even when the body of empirical work is robust, conclusions are likely to be tentative and described in probabilistic terms.

In the courtroom, however, research on general propositions presents merely the threshold question, since the ultimate issue is whether a particular case is an instance of the general phenomenon. In the eyewitness reliability example, the issue is whether the witness misidentified the defendant; in a tort case, the issue might involve whether the defendant’s product proximately caused the plaintiff’s illness. This issue of specific application poses a complex and difficult cognitive exercise.

The principal tool used to move from general research findings to statements about individual cases is “differential etiology,” sometimes misleadingly referred to as “differential diagnosis.”⁵⁵ Properly understood, differential *diagnosis* refers to the identification of the illness or behavioral condition that a person is experiencing.⁵⁶ Differential *etiology* refers to the cause or causes of that condition.⁵⁷ Hence, in the context of psychological practice, the determination that a person suffers from “dissociative amnesia” and not

⁵⁴ See *id.* § 27:14.

⁵⁵ See Edward J. Imwinkelried, *The Admissibility and Legal Sufficiency of Testimony About Differential Diagnosis (Etiology): Of Under – and Over – Estimations*, 56 BAYLOR L. REV. 391, 392 (2004).

⁵⁶ See *id.*

⁵⁷ See *id.*

“dissociative fugue” is a diagnostic issue.⁵⁸ The determination that a sexual assault at age ten (and not a medical condition or physical trauma) caused the diagnosed dissociative amnesia is an etiological matter. Very different skill sets are usually involved in these two determinations. Indeed, the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR) explicitly eschews any claim of etiological verity of its diagnostic categories.⁵⁹ It is also worth emphasizing that the validity of the diagnosis of dissociative amnesia is a matter of general research. Hence, the entire process of differential diagnosis and differential etiology assumes that the designated general category has adequate empirical support in the first place.

In ordinary clinical psychology, as is true in much of clinical medicine, the primary concern is diagnosis and not etiology. An oncologist might be curious about what caused his or her patient’s leukemia, but the doctor’s first task is to diagnose and treat the condition, not to determine whether trichloroethylene, electromagnetic fields, a genetic disorder, or something else caused it. Similarly, a psychologist treating a person thought to suffer from either Posttraumatic Stress Disorder (PTSD) or Adjustment Disorder is primarily concerned with identifying and treating the condition, not determining the true causes of that condition. In the ordinary practice of clinical medicine and clinical psychology, treatment and therapy are the principal objectives, not assessing cause. A person presenting symptoms associated with PTSD, therefore, may claim that the traumatic event was a sexual assault committed by her uncle. From the therapeutic standpoint, at least at the start, the important factor is that the patient honestly believes that a traumatic event occurred. Whether the patient’s uncle caused the trauma (or that a traumatic event even occurred) need not be specifically resolved for diagnostic and therapeutic purposes. In the law, of course, *who* caused the traumatic event is the crux of the matter.

In the courtroom, therefore, differential etiology is the operative issue. Moreover, the same basic principle is implicated if the expert opinion comes from research-based science or clinical practice. Whether researchers or clinicians have the ability to assist triers of fact in applying general research propositions to specific cases is a threshold legal matter that should depend on the reliability and validity of the differential etiology done in the respective case. Differential etiology, however, is anything but a straightforward affair, and most areas of science give it little or no attention.

Differential etiology is a reasoning process that involves a multitude of factors, few of which are easily quantified. The first task is to demonstrate that the substance could have caused the ailment, and the second is to show that other substances probably did not cause it. In the simplest situation, general research indicates that the substance causes an ailment that is uniquely

⁵⁸ AM. PSYCHIATRIC ASS’N, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 523-26 (4th ed., text rev. 2000) [hereinafter DSM-IV-TR].

⁵⁹ *Id.* at xxxiii.

associated with that substance. For instance, exposure to asbestos causes most mesotheliomas in the United States.⁶⁰ Since mesothelioma is a “signature disease,” the only question concerns the circumstances of the individual’s exposure to asbestos (i.e., whether the defendant was responsible), not whether exposure caused the condition. In contrast, while secondhand smoke has been linked to lung cancer, many other substances are known or suspected causes. Hence, in regard to identifying the cause of a person’s lung cancer, an expert must not only rule in smoking as a possible cause, but also rule out other possible causes.

An expert offering an opinion regarding a specific case must first consider the strength of the evidence for the general proposition being applied in the case. If substance *X* is claimed to have caused plaintiff’s condition *Y*, the initial inquiry must concern the strength of the relationship between *X* and *Y* as a general proposition. For example, both secondhand smoke and firsthand smoke are associated with lung cancer, but the strength of the relationship generally is much stronger for the latter than it is for the former. The inquiry regarding the strength of relationship will depend on many factors, including the statistical strength of any claims and the quality of the methods used in the research. Additionally, the general model must consider the strength of the evidence for alternative possible causes of *Y*, the strength of their respective relationships, and possible interactions with other factors. Again, the quality of the research and the different methodologies employed will make comparisons difficult. The myriad possible causes that have been inadequately studied further complicate matters in identifying potential causes of condition *Y*. Hence, determining the contours of the general model is a dicey affair in itself, as it requires combining disparate research results and discounting those results by an unknown factor associated with additional variables not yet studied. This is just the first part of the necessary analysis if the expert wants to give an opinion about an individual case.

The second part of the analysis – specific application of general propositions that are themselves supported by adequate research – requires two abilities, neither of which is clearly within most scientists’ skill sets. The first, and perhaps less problematic, is that of forensic investigator. Exposure or dosage levels will be relevant, regardless of the empirical relationship, to both medical and psychological diagnosis. The first principle of toxicology is that “the dose makes the poison,” since any substance in sufficient quantities could injure or kill someone.⁶¹ Similarly, in a wide variety of psychological contexts, the exposure or dose will be the poison. For instance, degree of trauma affects diagnostic category between PTSD and Adjustment Disorder,⁶² level of

⁶⁰ 3 FAIGMAN ET AL., *supra* note 1, § 28:20.

⁶¹ *Id.* § 24:17.

⁶² See DSM-IV-TR, *supra* note 58, at 682.

anxiety affects eyewitness identifications,⁶³ and extent of sleep deprivation affects false confession rates.⁶⁴ The expert testifying to specific causation must determine exposure and dosage levels for the suspected cause as well as for all other known or possible causes. This task is difficult enough alone, but is enormously complicated by the significant potential for recall bias, since what is recalled will profoundly affect the litigation.

The second skill set that is needed has not yet been invented or even described with precision. Somehow, the diagnostician must combine the surfeit of information concerning the multitude of factors that make up the general model, combine it with the case history information known or suspected about the individual, and offer an opinion with some level of confidence that substance or experience *X* likely caused condition *Y*. In practice, this opinion is usually stated as follows: "Within a reasonable degree of medical/psychological certainty, it is my opinion that *X* caused [a particular case of] *Y*." But this expression has no empirical meaning and is simply a mantra repeated by experts for purposes of legal decision makers who similarly have no idea what it means. Case-specific conclusions, in fact, appear to be based on an admixture of knowledge of the subject, experience over the years, commitment to the client or cause, intuition, and blind faith. Science it is not.

Finally, it should be mentioned that differential etiology is implicitly at the center of another area of expert evidence that is possibly the biggest embarrassment to the legal profession at this time. Many areas of forensic identification science operate on an etiological model, in that forensic experts offer opinions regarding particular cases. A firearms expert, for example, testifies that the bullet that killed the victim came from the defendant's gun to the exclusion of all guns in the world.⁶⁵ Similar kinds of testimony are heard from experts in the areas of fingerprints, handwriting, tool marks, bite marks, non-DNA hair analysis, and many others.⁶⁶ Unlike scientists who often make inferential leaps from general research to particular cases, forensic experts generally do not have any general data at all. These experts offer testimony that a particular case is or is not unique without data by which they may evaluate their assertion. These forensic identification specialists are essentially technicians who apply a technology built upon general statistical models that do not exist.

⁶³ See Thomas H. Kramer et al., *Weapon Focus, Arousal, and Eyewitness Memory*, 14 *LAW & HUM. BEHAV.* 167, 182-83 (1990).

⁶⁴ See Richard J. Ofshe & Richard A. Leo, *The Decision To Confess Falsely: Rational Choice and Irrational Action*, 74 *DENV. U. L. REV.* 979, 998 (1997).

⁶⁵ See generally Adina Schwartz, *A Systemic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification*, 6 *COLUM. SCI. & TECH. L. REV.* 2 (2005), <http://www.stlr.org/cite.cgi?volume=6&article=2> (arguing against "the admissibility of firearms and toolmark identification").

⁶⁶ See Michael J. Saks, *Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science*, 49 *HASTINGS L.J.* 1069, 1071 & n.8 (1998).

CONCLUSION

In *Daubert*, Chief Justice Rehnquist lamented that making judges gatekeepers for scientific evidence would require them to become "amateur scientists." In addition, he expressed doubt that judges were adequately trained to complete the task. He was correct on both counts. Given the integral role science plays in all areas of the law, not simply in matters of admissibility, it is long past time that judges became "amateur scientists."

Judges are generalists, and in order to decide cases they must sometimes consider history, economics, political theory, linguistics, and other areas. In effect, therefore, judges often find themselves in the role of amateur historians, amateur economists, or amateurs of some other sort. Science is pervasive in the law and there is no reason why it should be treated differently. Moreover, by its very nature, applied science cannot be effectively employed to shape legal doctrine if the statistics and research methods upon which it is built are not truly understood.

Currently, however, lawyers and judges are not well trained in, or favorably inclined to learn, the nuts and bolts of scientific inquiry. This is particularly problematic because scientific knowledge is intrinsically uncertain. Applied science is ordinarily expressed in probabilistic terms and the research methods on which it is based inevitably possess limitations and flaws. These uncertainties – the error rates built into the scientific premises – must be taken into account in the process of interpreting and applying the law. Judges' illiteracy in science means that they are ignorant regarding certain premises that are essential to modern judicial discourse. Judges no longer have any choice: their failure to become amateur scientists means their failure as professional judges.

BOSTON UNIVERSITY LAW REVIEW