SYMPOSIUM

A Cross-Disciplinary Look at Scientific Truth:
What’s the Law to Do?

THE UNEASY RELATIONSHIP BETWEEN SCIENCE AND LAW: AN ESSAY AND INTRODUCTION

Margaret A. Berger† & Lawrence M. Solan††

It would seem to be a match made in heaven. Trials attempt to seek the truth about contested events. Science attempts to seek the truth about observable phenomena. When the events that are the subject of legal disputes can be determined, at least in part, by virtue of scientific discovery, we might expect the law to embrace science as a means for ensuring that legal procedures get it right. Richard Katskee, an attorney who, on behalf of Americans United for Separation of Church and State, successfully challenged the inclusion of intelligent design in the biology curriculum of a Pennsylvania school district,1 puts it this way:

Scientific evidence has special value in legal proceedings because science confers intersubjective validity that other categories of truth-claims often lack. It offers factfinders and concerned observers a common yardstick against which to measure the validity and explanatory power of proffered evidence. So opinions grounded in science carry their own tests for reliability and usefulness, thus inspiring special confidence in judgments based on them. And by fostering greater public trust in legal rulings, judgments premised on scientific evidence reinforce the legal system’s ability to resolve

† Suzanne J. and William Miles Professor of Law, Brooklyn Law School.
‡ Don Forchelli Professor of Law, Director of the Center for the Study of Law, Language and Cognition, and Associate Dean for Academic Affairs, Brooklyn Law School.
disputes that might otherwise threaten a peaceful, well-ordered society.²

He later elaborates:

A conclusion based on evidence derived from research properly employing the scientific method inspires confidence because everyone can evaluate it using common and relatively easily applied criteria (namely, those that a scientific discipline sets for itself to test and potentially falsify hypotheses). And hence, there is never any need to take it on faith that an opinion or assertion is reliable.³

Yet science and law do not enjoy such a comfortable relationship, and the tension is nothing new. The replacement in the eighteenth century of court-appointed scientific experts by experts called to testify by parties within the adversarial system, combined with changes in the nature of scientific inquiry and the proliferation of legal cases that raise questions of science and technology, have led to shifting roles for scientists in the courtroom. In his insightful essay, “Revisiting the History of Scientific Expert Testimony,” historian Tal Golan introduces many of these developments, and the legal system’s reactions to them.⁴ If anything, things are better now than they have been. Golan concludes:

Far from being a late twentieth-century pathology, the putative problem of scientific expert testimony has been chronic for over two centuries. Moreover, during the twentieth century, the courts were able to take advantage of the professionalization of science and the standardization of the market of expertise and actually improved their ability to control the performance of science in the courtroom.⁵

In recent years, as the Supreme Court has formulated a new test for the admissibility of expert testimony to be administered by the trial judge as the “gatekeeper,”⁶ the legal system has been experiencing difficulty in determining just how it

---

³ Id. at 869.
⁵ Id. at 881.
should properly control this performance. It is in that context that we explore the question of truth in science and law.

The successful employment of science in the courtroom is most likely to happen when natural phenomena upon which the scientific community has reached consensus just happen to be in dispute in a legal case. For example, scientists might be able to assist the legal system in determining whether a particular toxin was emitted by a factory by scraping the factory’s smokestacks and analyzing the residue using well-accepted methods for the detection of chemical substances. When the scientific evidence relevant to a legal dispute is a matter of observable phenomena that have been studied and recorded scientifically, it is easy enough for the legal system to absorb this knowledge into its factfinding mission.

But both sides are likely to disappoint each other. The problems are most salient when scientists are called upon to offer opinions on causation. Although the title of this symposium contains the word “truth” and does not contain the word “causation,” it should be no surprise that many of the articles herein deal directly with the question of causation. Using as an example the well-accepted hypothesis that a particular virus causes a certain cancer, epidemiologist Douglas Weed explains why causation is such a problem for the legal system:

[T]he causal claim itself—that this type of virus caused that sort of cancer—does not have this same sort of connection back to some unique event that can be documented, verified, and directly observed. The causal claim is a scientific hypothesis and we cannot ever know if it is true in the same sense as the existence of the virus, the cancer, and its author. The hypothesis can be well supported or not by the available evidence. It can be more or less certain, more or less proven, but it cannot ever be true. The reason is remarkably straightforward. Causation cannot be seen. Causation cannot be proven. And the evidence for causation always underdetermines our capacity to choose between the causal hypothesis of interest and its various alternatives.7

The problem that underlies the indeterminacy of causation, philosopher Richard Scheines explains,8 is that determining causation necessarily requires that we think counterfactually, and drawing inferences from what has never occurred can be a tricky business. If we want to know whether

the emission of a chemical by a factory has led to the increase in endocrine disease in the immediate area, solid proof of causation can come only from comparing the actual world in which a population has experienced an increase in the disease with an imaginary world in which the very same population has had precisely the same experiences except for exposure to the chemical. If the occurrence of disease does not increase in this possible world, then we can conclude that the chemical has caused the disease, since exposure to the chemical is the only difference between the real world and the possible world.9

Of course, such experimentation is impossible, both for ethical and practical reasons. As a result, we must compare the population with the increase in disease with itself before exposure to the chemical, or with other populations assumed to be similar in all relevant respects. At this stage, doubts arise: Doesn’t this generation eat more fast food than earlier generations? Isn’t the base rate of pollution different than it was in the past? Are there other respects in which either an earlier generation or a neighboring population are not really the same?

So scientists do the best they can to tell what they consider the most reasonable story given what they know. The business of science is to investigate the range of possible variables and to select those most likely to produce a correct diagnosis. Scientists seem to be comfortable with this degree of uncertainty and with their quest for more certainty. They have chosen to make their livings that way. Susan Haack, in her essay, “Of Truth, in Science and Law,” puts it this way:

Whether or not they articulate it explicitly, most serious scientists have a firm-enough grasp of the complexities of evidence; this is why, wary of claiming to have found the truth, they prefer to say, “this seems like a promising idea,” “this model seems to fit what we know so far,” “probably the value of c is approximately n,” “perhaps the explanation might be this,” “possibly, it’s this way,” and such.10

But these scientists are not as certain as the lawyers would like them to be. Although there are many scientific truths accepted in both the scientific and lay communities, much of contemporary science involves researchers hypothesizing about

---

9 Philosophers use the expression “possible worlds” to describe this process. See David Lewis, Counterfactuals 84-90 (1973). Professor Scheines’ essay incorporates this terminology.

natural phenomena and offering tentative explanations that become the subject of further research, which results in both refinements and broad challenges. Moreover, there is often legitimate disagreement among scientists about the mechanisms that cause disease.

What should courts do when the proof of causation involves calculations demonstrating that the alleged offending substance could have caused an increase in illness, that the best account is that it did cause the increase in illness, but that the scientist, in all candor, can only make an educated guess? Permitting such proof may result in shifting the costs of illness to a defendant corporation that, the scientific community willingly admits, might not have caused the illness. Barring such proof, however, will almost certainly result in a grant of summary judgment in favor of the defendant for lack of proof of causation. As a result, there is no recovery for seriously ill or injured people whose plights, scientists believe, were caused by the defendant, but which they cannot prove to the judicial system’s satisfaction.

Often enough, the legal system’s answer to this question is that the evidence should be excluded. In many of these cases, plaintiffs fail to establish general causation, never mind specific causation. Daubert was such a case, although in that case, the weight of scientific opinion on the question at hand—whether Bendectin caused birth defects in children, and in particular, whether it had caused birth defects in Mrs. Daubert’s child—did not weigh in favor of the plaintiffs. One of us (Berger) has written critically of this regime, suggesting that when proof of causation would seem to fail because adequate research into the dangers of an alleged toxin has not been conducted, the plaintiff’s burden to prove causation should be relaxed, placing the onus of learning about the safety of chemicals on the companies that manufacture them.


Moreover, a closer look at the different goals of science and law can explain why judges appear to demand more of science than science demands of itself.\textsuperscript{14} By convention, scientists assume that they have not proven a relationship (say, between ingestion of a drug and an increase in heart attack) if they cannot reject the hypothesis that there is no such relationship with ninety-five percent certainty. Even then, they recognize that confounding information might lead them to change their minds later. The legal system, in contrast, wants to know what happened. When a scientist testifies that she has not proven a particular relationship, a judge may not believe that he has any choice but to reject the scientist’s opinion concerning the relationship. This has the effect of increasing the burden of proof in scientific cases from a preponderance of the evidence to near certainty.

The legal system’s hostility to uncertainty brings with it some ironic results. Consider the following situation: rather than being in relative consensus, albeit without clear proof, the scientific community can often be in vigorous disagreement. Just as historians might argue about the relative importance of, say, the various events that led up to the American Civil War, scientists argue about the best explanations for natural phenomena. At any scientific conference, researchers will present papers that attempt to explain—better than the current literature—the phenomena that they have devoted their lives to investigating, which may include anything from crystal formation to the mechanisms that lead to various kinds of liver disease. The researchers will use methods that are, at least as a general matter, accepted as good science, but they will reach different conclusions. What is ironic is that the legal system is far more welcoming of dueling experts who reach opposite conclusions than it is of consensus without certainty.

As Jennifer Mnookin points out in her essay, “Expert Evidence, Partisanship, and Epistemic Competence,”\textsuperscript{15} there are two dangers in this situation. The first is that experts working within the adversarial system are prone to become partisan. The second, which is somewhat in tension with the first, is that legal decision-makers, whether judges or jurors,

\textsuperscript{14} For fuller discussion, see Margaret A. Berger, \textit{Upsetting the Balance Between Adverse Interests: The Impact of the Supreme Court’s Trilogy on Expert Testimony in Toxic Tort Litigation} 64 \textit{Law & Contemp. Probs.} 289 (2001).

are not likely to have the knowledge or expertise to evaluate the relative scientific merits of the competing positions, a point echoed by Frank Keil’s essay, “Getting to the Truth: Grounding Incomplete Knowledge.”\textsuperscript{16} Returning to the use of court-appointed experts cannot provide a satisfactory answer, Mnookin explains, in part because these neutral experts might still testify based upon flawed reasoning, and jurors and judges are no better at understanding the scientific explanations of a neutral expert than a partisan one.

In combination, these concerns present serious problems for a legal system bent on discovering the truth through adversarial proceedings. Mnookin puts it this way:

What this means is that those experts who succeed in the marketplace for experts within our adversarial processes will often not be those with the most knowledge or actual expertise in a particular area, but rather those whom parties believe will succeed in persuading the factfinder. The confluence of adversarialism with the need for expert information also has permitted the creation of a class of “professional” expert witnesses, those for whom expert witnessing is no longer a sideline, a once-in-a-while add-on to their primary work as a physician, economist, epidemiologist, statistician, or whatnot, but rather is a significant, or even primary, source of income.\textsuperscript{17}

The result is that parties attempt to put on the witness stand individuals who are charismatic and whose past experience will impress jurors (or judges) regardless of the relative merits. The lawyer’s job is to win cases, and it would violate the duty of representing their clients vigorously to do otherwise.

An additional, complementary problem enters the picture: the legal system must maintain a healthy skepticism about claims of scientific rigor in order to shield itself from being duped by those who practice “junk science” or “pseudoscience.” As Frank Keil points out in his essay, reinforcing Mnookin’s concerns, people are not very good at recognizing the limits of their own understanding of complex systems: “Most people are quite inept at estimating how well they understand various everyday phenomena, showing a strong tendency to assume they understand how the world works in far more detail than they really do.”\textsuperscript{18} Thus, Keil observes, people may be susceptible, for example, to giving special credence to techno-

\textsuperscript{17} Mnookin, \textit{supra} note 15, at 1011-12.
\textsuperscript{18} Keil, \textit{supra} note 16, at 1038.
logical explanations using, say, fMRI information, assuming that they understand the process well enough to make a valid judgment. Sometimes such evidence can be important. But it can also create the illusion of scientific certainty. Dennis Patterson’s essay provides an argument for a case in point:19 recent theorists have claimed that our brains contain universal and innate moral judgments characterized as a set of rules that we follow unselfconsciously and, more or less, automatically.20 By challenging the conceptual foundations of a set of claims receiving considerable attention in the legal academic literature, Patterson points out that our ability to judge the legitimacy of scientific claims is surely subject to question.

This is not to say that courts should hold science in low esteem, even the so-called “soft” sciences. David Faigman, in his essay, “Scientific Realism in Constitutional Law,”21 points out that the Supreme Court routinely relies upon tacit theories of sociology and of folk psychology. Social scientists, in this context, may indeed have a role to play. Faigman illustrates his point with, among other examples, Witherspoon v. Illinois, in which the Court permitted prosecutors to exclude as jurors in the guilt phase of capital trials those who have moral scruples against capital punishment or who are opposed to the practice altogether.22 The Court found that the data adduced by the defendant indicating a bias favoring guilt among a jury so selected was not worthy of consideration, but acknowledged that well-conducted studies might be relevant, at least in principle. However, once such studies were conducted by social scientists in response to the challenge in Witherspoon, the Court rejected the relevance of this sort of data altogether.23 Faigman, whose work has been highly critical of the introduction of junk science into the courts,24 nonetheless disagrees with

---

23 Lockhart v. McCree, 476 U.S. 162, 168-69, 171-72 (1986). The Court further endorsed a unitary jury deciding both the guilt and penalty phases of a capital trial, thus permitting prosecutors to reject potential jurors who are opposed to the death penalty to decide whether a defendant should be put to death. Id. at 182.
24 See, e.g., David L. Faigman et al., Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence, 15 CARDozo L. REV. 1799 (1994).
this position. While social science research may lack the crispness of the hard sciences, it can be conducted responsibly and should certainly not be ignored when its findings are relevant to the concerns of the legal system, even constitutional ones.

Sam Glucksberg’s commentary begins, “Truth is hard to come by, even in optimal circumstances where the criteria are explicit and clear, and where it can be objectively established whether those criteria have been met (at least in principle).” Courts are often confronted with circumstances that are not optimal. It is even more difficult to assess the reliability of a scientist’s candid statement that there is a “best story” to tell, although it cannot yet be proven than it is to come by truth in Glucksberg’s ideal situation. Nonetheless, the articles in this volume lead to an important conclusion: There is an important difference between junk science on the one hand and incomplete knowledge on the other. Junk science makes its way into the courtroom when experts offer opinions based merely on intuition, without evidence that their intuitions are any better than those of lay people. Michael Risinger, among others, has written about the historical admissibility of handwriting expertise as an example of this phenomenon, which similarly afflicts many other so-called forensic identification sciences.

Junk science has no place in the legal system. But when a scientist says, “I’m not sure, but the data are suggestive,” the scientist’s words are not necessarily a tell-tale sign of junk science. On the contrary, it may be a sign that real science is occurring. The Supreme Court’s Daubert trilogy appears to have allowed the continuation of junk science, while denying individuals their day in court when their proof includes real science at a state of incomplete knowledge. The excellent essays in this volume tell us how this has come to happen. The solution, we believe, must lie in the legal system, judges and lawyers alike, recognizing what it means to be a gatekeeper with respect to scientific truth. And that, in turn, requires the legal system to come to understand just what it is that scientists do.

27 See id. at 135-43 (discussing courts’ willingness to accept expert identification testimony based on bite marks and handwriting without requiring proof that those fields are grounded in scientific research).
Science, Intersubjective Validity, and Judicial Legitimacy

Richard B. Katskee

The problems associated with discovering truth in the courtroom are well known. Eyewitness testimony is notoriously unreliable. Statistics are almost endlessly manipulable. Paid experts slant their findings or, less disreputably, perhaps, but equally effectively, adjust the questions that they ask in order to yield findings supporting the party who hired them. But in litigating challenges to the incorporation of religious beliefs packaged as science into public-school curricula, my concern with the relationship between expert testimony and scientific truth has less to do with the mechanics of weighing possibly conflicting expert opinions than with the deference so often afforded to those who don the trappings of science, whether they engage in bona fide scientific research or merely peddle nonscientific truth-claims masquerading as science.

Although much of this symposium has focused, in one way or another, on whether science offers a window on the truth commensurate with the pride of place that scientific evidence receives in legal factfinding, that question may be too narrow to acknowledge the full value of scientific evidence in judicial proceedings. If scientific research offers access to truth that other forms of evidence do not, affording it extra deference makes perfect sense. But whether ultimate or objective truth even exists, and, if so, whether we as humans have epistemic access to it (through scientific inquiry or otherwise), are metaphysical puzzles that have plagued philosophers and

† Assistant Legal Director, Americans United for Separation of Church and State.

1 The author was one of the principal attorneys for the plaintiffs in the 2005 case successfully challenging the inclusion of intelligent-design creationism in the biology curriculum at a public high school in Dover, Pennsylvania. See Kitzmiller v. Dover Area Sch. Dist., 400 F. Supp. 2d 707 (M.D. Pa. 2005). The author has since served as lead counsel in other cases involving the teaching of intelligent design, creation science, and other religiously based attacks on the scientific theory of evolution.
theologians since at least Plato's day; and it is not clear that we are any closer to solving them than the Ancient Greeks were. Courts must, of course, set aside those thorny issues in performing their daily tasks, acting instead as though ultimate truth exists and is reasonably accessible to anyone who in good faith looks for it. Otherwise, metaphysical and epistemological anxieties would overwhelm courts' ability ever to provide definitive judgments. But the need for courts to act as though they can discover ultimate truth through ordinary legal proceedings does not resolve the deeper questions. Nor, therefore, does it justify viewing science in the courtroom as useful for getting to the truth of the matter being adjudicated.

Irrespective of the metaphysical status of ultimate truth, or of science's relation to it, a better question may be whether anything about the nature of science (whether or not it relates to the capacity to reveal truth) warrants setting scientific evidence above other categories of truth-claims as a grounding for legal judgments. As background for considering that question, Part I provides a more detailed statement of the concern about courts' institutional competence to deal with scientific evidence, and Part II describes courts' institutional aims and the value of publicly justified judicial rulings for achieving those aims. Part III explains why science as a discipline has special power to promote shared understanding. And Part IV seeks to defend the respect that courts show to scientific evidence in light of the public confidence in the legal system that follows from judgments' being rooted in shared understandings of the sort that science provides.

The point is, in the end, a simple one: Scientific evidence has special value in legal proceedings because science confers intersubjective validity that other categories of truth-claims often lack. It offers factfinders and concerned observers a common yardstick against which to measure the validity and explanatory power of proffered evidence. So opinions grounded in science carry their own tests for reliability and usefulness, thus inspiring special confidence in judgments based on them. And by fostering greater public trust in legal rulings, judgments premised on scientific evidence reinforce the legal system's ability to resolve disputes that might otherwise threaten a peaceful, well-ordered society.
I. THE INSTITUTIONAL-COMPETENCE PROBLEM

In defending, at least in general (if not necessarily in specific cases), the role that science plays in legal proceedings, I am not deaf to the frequently voiced complaint that juries and judges are ill suited to the task of evaluating competing scientific claims. The complaint is one about institutional competence: Factfinders are not trained as scientists, so they are unlikely to possess the substantive knowledge and analytical skills required to parse scientific claims; and far from making up for those deficiencies, the adversary system compounds the problem by driving parties to present skewed accounts and to gloss over weaknesses and inconsistencies in the methodologies employed, the data collected, and the conclusions drawn by experts testifying on their behalf. But while institutional-competence concerns are certainly important, complaints about courts’ inability to understand science are not terribly informative when divorced from consideration of institutional objectives. And when courts’ institutional purpose and social role are taken into account, criticisms of factfinders’ supposedly dismal performance in evaluating scientific evidence may be overblown.

Consider a run-of-the-mine tort case in which the factfinder (whether judge or jury) must decide which party is providing the best account of how some injury occurred. Irrespective of what that injury is, or what the legal claims are, or what proof each party offers, it is natural to wonder whether the factfinder will have knowledge and experience of the relevant sorts, and in sufficient measure, to weigh the evidence, evaluate the competing arguments, and come to the correct conclusion—in other words, to figure out the truth of the matter. And when apparently conflicting scientific evidence is a trial’s centerpiece, the anxiety deepens: If we doubt judges’ and juries’ ability to evaluate witnesses’ credibility, for example, even though everyone enters the courtroom with at least some independent experience distinguishing truthfulness from deceit, must we not be even more skeptical of factfinders’ capacity to weigh competing scientific claims, when so few among us possess even rudimentary training in that enterprise? After all, most judges and jurors are unlikely to have ever before tried to make sense of the sorts of data being put before them in a trial; they may be swayed by flash rather than substance; and even when they would otherwise possess sufficient acumen to separate scientific wheat from junk-
science chaff, their efforts may be stymied by the parties’ attempts to sweep under the rug the limitations of an expert witness’s research program and the qualifiers that would inevitably accompany the findings if they were presented in even the most slipshod research paper.

But if factfinders’ evaluation of expert testimony is good enough to allow courts to fulfill their designated social role, then courts are sufficiently competent as factfinders to consider the evidence for purposes of deciding cases. And judicial decisions will be acceptably reliable, even if judges’ and juries’ evaluations of competing scientific claims might sometimes be unsophisticated enough to bring tears to the eyes of any high-school biology or physics teacher—much less to a qualified researcher working in the relevant field of study. So determining institutional competence to evaluate scientific evidence requires looking, in the first instance, not at factfinders’ scientific acumen, but at courts’ designated social role.

II. LEGAL JUDGMENTS AND INTERSUBJECTIVE VALIDITY

As John Locke explained, law courts are a prerequisite to social stability: Impartial judges empowered to make definitive legal rulings provide a mechanism for disputants to resolve their disagreements without having to resort to physical violence or other self-help remedies. 2 We set up judges and courts as higher authorities with the power to declare who wins and who loses, who receives compensation and who pays, who exacts retribution and who suffers punishment, not because we think that they will always get things exactly right, but just because deferring to a neutral arbiter helps ensure that we don’t end up in blood feuds to resolve every petty grievance. But a legal system prevents interpersonal violence and quells broader social strife only if both the parties to particular controversies and the public in general are in the end willing to accept and obey legal judgments. For if not, invoking the judicial process might delay, but will not prevent, resorts to bloodshed. It thus turns out that disputes over how to assess the reliability of various categories of evidence or how to deal with factfinders’ possibly inept weighing of truth-claims implicate deeper issues of political legitimacy: If a court is to

---

resolve disputes definitively and finally, and hence to justify invoking political authority to enforce its orders, it must persuade those subject to its rulings to trust it to decide who should bear the costs of an accident or who should receive punishment for a crime—in other words, to make conclusive determinations regarding litigants’ fates.

In light of that socially important, but in a metaphysical sense fairly modest, institutional aim, courts need not be infallible expositors of truth. In most cases, it is enough that a court comes to some decision—any decision—that the parties will accept as dispositive. For courts don’t just attempt to ascertain truth; they define it. Justice Jackson’s famous aphorism, “We are not final because we are infallible, but we are infallible only because we are final,” provides a pointed reminder that a judicial ruling makes true as a legal matter the findings that it encompasses, irrespective of whether the judge, jury, or appellate tribunal got everything exactly right. Judicial fiat is not, of course, a terribly satisfying basis for dispute resolution; and hence, it alone is not a secure grounding for a legal system. But if the parties and the public have the considered conviction that courts do a pretty good job finding facts, deciding cases, and resolving disagreements most of the time—in other words, that the legal system on the whole serves us better than trying to settle the score ourselves whenever we feel wronged—then we will generally be willing to let the courts act as final arbiter. And we will usually accept even unfavorable judgments from them.

Thus, what matters for a legal system’s legitimacy is, in the first instance, that losers in legal actions believe that the treatment they received from the courts was at least minimally fair and respectful of them as the prevailing parties’ political equals. A losing party must be able to see that the court has listened to its arguments, weighed all the relevant evidence, and explained the ruling in a way that the party can recognize as a valid—albeit unfavorable and perhaps imperfect—exercise of judicial authority. As I have suggested elsewhere, the losers’ view matters most because winners do not need to have their victories justified to them. Unless a prevailing party expects to be a repeat player in litigation, fighting the same battle against future opponents in other cases, it is unlikely to care why it

won: Knowing *that* it won is quite satisfying enough. But to the unsuccessful litigant, quite a lot depends on believing that the arbitrator was impartial, that the hearing was fair, and that the final decision fell within at least a broad range of outcomes warranted by the evidence when evaluated using permissible forms of analysis. A litigant’s views about a judgment on those dimensions may make the difference between accepting an unfavorable ruling and keeping the fight going outside the courts.

Stepping back from the parties’ views of any single case to an aggregate social perspective, the procedures that a legal system employs and the judgments that it produces, when taken all together, have to be fair enough, both on average and in the most significant cases, to encourage members of the society to keep reverting to the courts as the principal mechanism for resolving disputes. Courts’ legitimacy depends, in other words, not just on individual losing parties’ walking away with the conviction that a courtroom was the proper venue to resolve grievances (however upset the parties might have been about the final verdict), but on the public’s having faith that the legal process will afford a fair hearing and generally fair treatment to those who invoke it—and that the courts will give careful, respectful consideration even to nonparties’ interests when they are implicated in lawsuits.

If courts are to provide what most of us will regard as a fair trial in those senses, and hence to instill confidence that the legal system is the right place to turn to resolve disagreements, they must base their rulings on publicly accessible facts, which they must then weigh, measure, and test using publicly accessible forms of legal reasoning. In other words, courts have to limit themselves to types of evidence and modes of analysis that are equally intelligible to everyone, at least in principle, and to employ relatively transparent procedures, so that the resulting judgments are defensible as something more than mere caprice. If a court can explain its reasons for reaching some conclusion, and those reasons are generally accepted as valid—that is to say, commonly recognized as appropriate to a judicial decision-maker given the applicable legal rules, and taking into account the relative seriousness of the case—the accuracy or truth of the judgment in a metaphysical sense may be rather beside the point. For if a court relies on commonly shared premises and modes of reasoning that litigants and the public as a whole can recognize as suitable, and if it uses those analytical tools to draw what
the parties and the public can regard as reasonable inferences from facts that are themselves intelligible to, and in some sense verifiable by, everyone, then the resulting judgment will have intersubjective validity—in other words, the basis for shared understanding of its fitness. And if a judgment is grounded in that sort of shared understanding, it cannot be written off as an exercise of raw power; it must be regarded, even by those who dislike the result, as deserving of respect, and ultimately, of obedience.

To be sure, the higher the stakes are, or the more in the public eye the controversy is, the more accuracy that both litigants and the public as a whole will demand. If a limb from my neighbor’s tree falls onto my fence, we might disagree heatedly about which of us should bear the cost of the repairs and in what amount. But it does not matter very much for practical purposes whether a court hearing the dispute orders my neighbor to pay me enough to build a new, stronger fence; or awards just enough to enable me to patch the broken boards; or denies my claim altogether. Pretty much any definitive answer will do: As long as there is some mechanism to assign fault and assess compensation, I will not need to try to break even by stealing my neighbor’s lawnmower—or to get even by poisoning his azaleas. But if, for example, the question is whether my physician should compensate me for a lifetime’s lost wages because of a misdiagnosis, and whether she should in the process suffer a substantial blemish on her professional record, the stakes are higher; so a judgment bearing greater indicia of reliability is warranted. And when the stakes are at their highest—as when the question is whether a court should impose a severe criminal penalty or announce a new legal rule that would affect large numbers of people—the parties and the public reasonably demand a judgment with the highest degree of accuracy possible. The systematic failure to live up to that heightened expectation, especially in well-publicized cases, would substantially erode public confidence in the legal system as the appropriate means to resolve disputes, thus encouraging people to forsake the courts entirely and fall back on self-help remedies. If the most significant controversies are seen not to be amenable to resolution through the legal system, in other words, the fact that the courts handle mundane grievances tolerably well is hardly a source of long-term social stability.

But to say that we expect greater accuracy or certainty in the cases where the stakes are high is nothing more (and nothing less) than to acknowledge that we demand more robust
public justifications for courts' decisions. Definitiveness alone is insufficient: We insist on exceptionally clear, detailed explanations for the rulings—explanations grounded in facts and legal principles that are both widely intelligible and generally accepted as reasonable.

In demanding more careful, complete justifications in important cases, we also shrink, to some extent, the pool of potential judgments that might be considered legally warranted. For in the first instance, the losing party when the stakes are extraordinarily high will likely be unwilling to accept an adverse judgment absent strong reasons to think that the court (a) considered all the relevant evidence and arguments with particular thoroughness and care; (b) applied only proper forms of legal reasoning; and (c) drew inferences and reached conclusions that are intelligible, reasonable, and suited to the seriousness of the case. The more conscientious the court has been in rendering its decision, however, the more limited the set of possible outcomes will be. Whereas in a minor dispute, a judgment might be perfectly good if it does nothing more than give the parties bare direction in how to act toward one another so that they can get on with their business—and hence virtually any decision will do (so long as it can practicably be implemented)—in important cases, the greater need for publicly justified rulings means that only more rigorous and more rigorously fair decision-making will pass muster, with the result that more possible outcomes will be rejected as illogical, irreconcilable with the evidence, fundamentally unfair, or otherwise legally untenable.

In the end, most of us will only rarely be aware of mundane legal disputes that do not involve us; we need only believe that, at the most general level, the courts provide a reasonably efficient mechanism to resolve those everyday controversies. But when a case’s outcome will affect large numbers of people (as in toxic-tort suits or suits involving corporate malfeasance injuring employees, shareholders, and customers alike), or when it implicates our most basic notions of liberty (as in criminal prosecutions where the accused faces severe penalties) or equality (as in civil-rights actions challenging official discrimination), we demand that the courts provide especially strong public justifications for their rulings. For what confers legitimacy on a legal system (thus making peaceable resort to the courts the accepted means to resolve the disputes that would otherwise have the greatest potential to create social divisions and civil unrest) is the ability of
everyone, at least in principle, to understand important judicial decisions and to regard them as worthy of respect. To be sure, intersubjective validity of that sort may not be easy to achieve, especially in the most important, most public, and most difficult cases—which are, of course, the ones where it is most needed. But ensuring intersubjective validity is far more modest an institutional aim than discovering ultimate truth. So the criticism that courts are not infallible expositors of that truth—that they sometimes get things wrong—need not be viewed as a fatal blow to a legal system's legitimacy.

III. **Methodological Naturalism and Intersubjective Validity**

Just as judicial opinions become more persuasive the more that judges take care to apply rules and invoke principles that are intelligible to and generally accepted by the parties and the public, so too do factual findings provide more substantial support for judgments when the evidence supporting them can be measured against generally accepted standards. When parties and interested observers can weigh the evidence for themselves and conclude that the factfinder has drawn reasonable inferences and reached warranted conclusions, public confidence in authoritative rulings will be at its zenith. Considered on that dimension, scientific evidence deserves to be regarded as especially useful in court: Science as a discipline imposes strict limitations on what can count as a scientific truth-claim; so long as those limitations are respected, the standards that the scientific community imposes for evaluating data and the inferences drawn therefrom will also provide intersubjectively intelligible measures of the validity and strength of the particular truth-claim at issue. Put more simply, science provides accepted tests for whether and to what extent opinions deserve to be respected as scientific conclusions; and legal judgments that employ those tests thus partake of the authority that the scientific method confers on robust research results.

To start with science's ground rules—which philosophers of science call methodological naturalism, but those in the lab just call the scientific method—doing science means committing oneself to the search for natural explanations for natural phenomena. As the National Academy of Sciences has put it:
In science, explanations are restricted to those that can be inferred from the confirmable data—the results obtained through observations and experiments that can be substantiated by other scientists. Anything that can be observed or measured is amenable to scientific investigation. Explanations that cannot be based on empirical evidence are not a part of science.5

Science thus differs from other disciplines not in its areas of interest—which are almost infinitely varied, and which overlap with matters addressed by philosophers, historians, theologians, and scholars in many other fields—but in its modes of analysis. Only observation and inference confirmable using sense data are regarded as permissible forms of scientific inquiry. If we can’t see, feel, hear, smell, or taste something, we can’t study it using science (though there may be many nonscientific ways to analyze it profitably) because only if conclusions are based wholly on empirical observation can others repeat, test, and potentially falsify the results. We might hypothesize an unseen force—gravity, for example—to explain the otherwise unexplainable—why the apple falls from the tree. We cannot call gravity a scientific theory, however, until we have ascertained that our hypothesis holds up under rigorous empirical testing and retains its explanatory power in a wide array of conditions. And if the hypothesis cannot be tested and potentially falsified, it is not amenable to scientific inquiry.

Methodological naturalism is a pragmatic rule, not a deeper philosophical commitment to materialism. The reason that science as a discipline limits itself to natural explanations is that what is observable, repeatable, testable, and falsifiable provides the basis for making useful predictions. If, for example, we can explain drug-resistant illnesses in terms of evolutionary theory—natural selection acting on populations of bacteria undergoing random genetic mutations—we can predict how new antibiotics or healthcare practices might minimize the growth of drug-resistant strains.

In making the choice to seek causal explanations that allow for prediction, we formally reject nonnatural explanations—divine intervention, spirit forces, or anything

5 Working Group on Teaching Evolution, National Academy of Sciences, Teaching About Evolution and the Nature of Science 27 (1998); see also National Academy of Sciences & Institute of Medicine, Science, Evolution, and Creationism 10 (2008) (defining science as “[t]he use of evidence to construct testable explanations and predictions of natural phenomena, as well as the knowledge generated through this process”).
else partaking of the supernatural—as standing outside science’s ken. We do so not because those explanations are necessarily wrong, but because, by definition, supernatural actors or forces exist and function outside the laws of nature, doing whatever they please, whenever they wish, and however they choose. So the willingness to entertain the idea of nonnatural causation as part of one’s scientific research program would mean throwing in the towel on the possibility of doing science at all: Because one can never prove or disprove supernatural explanations—because, in short, “God did it” explains everything and nothing—one can never reach conclusions about when, where, how, or even whether a nonnatural cause will next occur. Each nonnatural cause—each act of divine intervention—is sui generis. So we can never confirm or reject a supernatural explanation, even in principle; a leap of faith is always required. And hence, unless nonnatural causes are formally excluded, one can never have the slightest confidence in any prediction; supernatural acts can always break any causal chain. Rather than being the product of natural selection acting on random genetic mutations in bacteria, perhaps drug-resistant diseases might be the work of some unseen, undiscoverable being or beings who occasionally start epidemics in order to keep the human population in check, or to test our faith with suffering (as the Catholic Church in the Middle Ages viewed infectious diseases now routinely cured with antibiotics), or to punish sinners (as some fundamentalist-Christian leaders today explain the AIDS virus and, for that matter, hurricane Katrina and September 11). Why bother attempting to develop new antibiotics or to formulate better healthcare practices when, from our limited point of view, there’s no rhyme or reason to how God, the devil, the ghost in the machine, or a merry band of cosmic pranksters will act next—unless, of course, we have had the grand plan handed to us as revealed truth?6

6 I do not mean to suggest that pure reason, empirical observation, and religious belief or other invocations of nonnatural causes are inherently irreconcilable. They may or may not be in tension, depending on one’s theological commitments. But while for many if not most people, all three cohere and are mutually reinforcing—a philosophical and theological view with a distinguished Enlightenment pedigree but roots extending much further back in Western intellectual and religious history—one still must necessarily take a leap of faith from science’s empirical observations and conclusions to the idea that a supernatural force is working behind the scenes.

As Catholic theologian John Haught explained in testifying for the plaintiffs in the Kitzmiller intelligent-design case, science and faith need not be at loggerheads because science has more modest aims than religion does: It restricts itself
In formally excluding from scientific inquiry those sorts of explanations for natural phenomena, while still acknowledging, of course, that they might be true, and that science might be inadequate to tell us everything about everything, science as a discipline puts a premium on continuing the search for causal explanations that can potentially serve as the basis for prediction and responsive action. If no natural explanations work, there will be time enough at the end of the day to step outside the realm of scientific inquiry and invoke nonnatural explanations. The alternative—accepting at the outset that the causal mechanisms producing observable phenomena are unknown and inherently unknowable—would be a recipe for stagnation, encouraging would-be scientists to throw up their hands in despair of ever adding to the store of human knowledge or solving any practical problem.

IV. SCIENTIFIC AUTHORITY AND LEGAL JUDGMENTS

The methodological restriction on the sorts of explanations that can count as scientific turns out to have a salutary effect for legal proceedings as well as for science itself: Limiting research to what is repeatable, falsifiable, and testable means drawing conclusions that are empirically verifiable using sense data to which everyone has epistemic access. If one maintains a commitment to methodological naturalism (and therefore refuses, for purposes of conducting research, to consider any explanations other than those that are falsifiable through empirical observation and testing), then what is left as the basis for permissible inferences is equally

to what Haught termed “how” questions (“How do systems function?”), whereas religion transcends the material to ask “why” questions (“Why are we here?”). See Kitzmiller, 400 F. Supp. 2d at 735.

Methodological naturalism functions as a bar to scientific exploration of the “why” questions, even though scientific discoveries very often have consequences for how we think about those bigger metaphysical issues. So while research scientists may choose to comment on what they view as the philosophical or theological implications of their work, they do so not as scientists but as educated laity. Although they may not always be punctilious about informing their audience when they are doffing their lab coats to confront questions of faith and spirituality that they regard as partially illuminated by discoveries made in their professional capacities, science as a discipline requires them to draw a careful line between conclusions within the scope of scientific inquiry and those outside it. For an example of appropriate rigor in distinguishing between scientific findings and metaphysical views arising in part from nonscientific interpretations of those findings, see KENNETH R. MILLER, FINDING DARWIN’S GOD: A SCIENTIST’S SEARCH FOR COMMON GROUND BETWEEN GOD AND EVOLUTION (1999).
intelligible to everyone, at least in principle: We all have roughly the same capacity to observe, measure, and test, so we each, independently, have the ability to assess the validity and strength of any inferences drawn from the data. The upshot is that science has at its core a strong form of intersubjective validity: A conclusion based on evidence derived from research properly employing the scientific method inspires confidence because everyone can evaluate it using common and relatively easily applied criteria (namely, those that a scientific discipline sets for itself to test and potentially falsify hypotheses). And hence, there is never any need to take it on faith that an opinion or assertion is reliable.

Although science is enormously useful in its own right, and has the potential to inspire great confidence in decisions or courses of action grounded in it, I certainly do not mean to suggest that all invocations of science are inherently trustworthy, or that scientific fraud never occurs. Nor are courtroom misrepresentations (whether intentional or resulting from lawyers’ or experts’ ignorance or lack of preparation) less harmful when made by those claiming the mantle of scientific authority than when made by anyone else. Quite the contrary. If a legal system’s legitimacy turns on the ability of stakeholders and others to evaluate evidence independently and satisfy themselves that judgments are fair, careful, and warranted in light of the evidence, the operative legal rules, and permissible forms of legal reasoning, then misrepresentations that play on judges’ and jurors’ lack of scientific acumen will distort factfinding and cast a pall over judgments, leaving parties and observers to worry that the legal system is inadequate to resolve disputes. Obtaining acceptably reliable judgments grounded in scientific-expert testimony thus turns out to depend, after all, on ascertaining the amount of respect that an expert witness’s truth-claims ought to receive. But that calculus depends on the degree to which the evidence demonstrably bears the indicia of genuine science, and thereby takes advantage of the intersubjective validity that the scientific method provides—whatever one may think of science’s ability to reveal truth in a metaphysical sense.

In his classic experiments on obedience to authority, psychologist Stanley Milgram showed how dressing an actor in a lab coat and introducing him as a research scientist could cause people from all walks of life, with all levels of education, to set aside their own judgment and afford deference—indeed, blind obedience—far beyond what reason, or even simple
decency, should have permitted.\(^7\) Confronted with the apparent authority and imprimatur of the scientific community, Milgram’s test subjects would obey instructions to deliver harmful electrical shocks to another person.\(^8\) Believing that they were assisting in the conduct of an experiment, rather than that they themselves were the subjects of study, they would suspend compassion and all sense of moral responsibility, delivering shock after shock, in what appeared to be increasingly dangerous voltages, just because the fellow in the lab coat said that the scientific enterprise required their continuing participation.\(^9\)

In the courtroom, this white-coat effect leads not to blind obedience, but to blind acceptance, though the effect is every bit as real—and in high-stakes cases, potentially every bit as pernicious. Technospeak alone may be quite enough to fool even well-educated nonscientists into thinking that there must be merit to what a witness is saying, even if the testimony is, in actuality, entirely vacuous. Judges and juries may suspend their reason, judgment, and basic common sense, accepting arrant nonsense as true just because the witness who delivers it looks and sounds like a bona fide scientist. Conversely, they may reject science of the highest caliber if the witness presenting it does not wear the mantle of scientific authority well: A genuine expert who tries too hard to speak plainly, or who simply does not cut an impressive figure in the witness box, may be unfairly dismissed as a hack or a fraud. What could cast greater doubt on the reliability and integrity of legal judgments than the realization that ordinarily skeptical people, empowered to make decisions about others’ fates and fortunes, may base their decisions on unreflective acceptance of charlatanism supported by showmanship, ignoring solid research backed by training and experience?

Thus, the problem of institutional competence resurfaces, albeit at a different level of abstraction: What matters is not so much whether judges and juries can independently evaluate the data put before them, but whether they can truly grasp the nature of science itself in order to ascertain whether a statement of opinion should receive the respect that intersubjectively valid scientific findings do as foundations for

\(^{7}\) See generally Stanley Milgram, Obedience to Authority (1974).
\(^{8}\) See id. at 3-4, 16-19.
\(^{9}\) See id. at 3-4, 16-23.
authoritative judgments. The worry, then, is that if factfinders are not equipped to distinguish good chemistry or physics or genetics from bad, how much more difficult must it be for them to look beneath the particulars of slickly presented expert opinions in order to determine not just which side boasts the more convincing findings, but whether and to what extent each proffered expert is engaging in the scientific enterprise in the first place?

But limitations on judges’ and juries’ ability to parse and weigh scientific evidence notwithstanding, my own view is that the institutional-competence problem is not nearly so hopeless as we often suppose—in part because it turns out that courts are better at answering the philosophically rich question, “Is it science?” than at resolving the more pedestrian question, “Is it good science?”

One reason why is that the Daubert test offers a readily administrable formula for ascertaining whether an opinion counts as scientific, though not necessarily for determining the strength of particular scientific inferences and conclusions. Although courts applying the Daubert criteria routinely speak in terms of distinguishing good science from junk science (a term that encompasses both poor science and pseudoscience), what they are really doing, and what, as I read Daubert, the Supreme Court meant for them to be doing, is distinguishing between science qua science (empirical inquiry employing the scientific method), on the one hand, and nonscience (inferences and conclusions not derived from testable, repeatable, falsifiable observations), on the other. Daubert adopts and employs the scientific community’s self-understanding about what science is, requiring the party proffering the expert witness to demonstrate that the data and opinions offered satisfy science’s ground rules. In the process, the Daubert test also brings to light the specific features of the evidence that the scientific community itself would use to assess the validity, reliability, and explanatory power of any purported findings; but those added benefits are in a sense only incidental to the Daubert inquiry.

Although most commentators treat the Daubert test as though it was designed to address those second-order

---

11 Id. at 590.
12 Id. at 590-95.
questions, that conventional view is inconsistent with both the test’s specific features and the analytical approach that the Supreme Court took in formulating it. The Court did not, after all, try to define “good science” or “good enough science” according to some qualitative measure. It built the Daubert standard on a definition of “scientific knowledge”—the operative term in Federal Rule of Evidence 702—that the Court drew not from background legal principles or established court practices, but from the scientific community’s own definition of science as methodological naturalism. Thus, the Daubert factors are, as formally stated, far better suited to discovering whether expert testimony is grounded in the scientific method than to deciding how much explanatory power an expert’s findings actually have. Whether an expert’s procedures or conclusions can be and have been tested and whether they have known or potential error rates are, for example, straightforward tests for determining whether an opinion is falsifiable—not whether it has been falsified. Had the Supreme Court meant to focus on the strength of the evidence rather than on its classification as scientific or nonscientific, the Daubert test surely would have required parties to show that their experts’ proffered findings have withstood actual testing and that they provide substantial explanatory power with demonstrably low error rates. That testing is possible and that error rates can be calculated are certainly prerequisites to assessing the strength of the evidence the way that practitioners in the relevant scholarly field would; but in making the Daubert inquiry turn on the threshold question of testability rather than on the results of actual testing, the Supreme Court effectively instructed the lower

13 Relying principally on an amicus brief filed by scientific organizations, the Supreme Court breathed life into Rule 702’s language by explaining:

The adjective “scientific” implies a grounding in the methods and procedures of science. Similarly, the word “knowledge” connotes more than subjective belief or unsupported speculation. . . . Of course, it would be unreasonable to conclude that the subject of scientific testimony must be “known” to a certainty; arguably, there are no certainties in science. . . . But, in order to qualify as “scientific knowledge,” an inference or assertion must be derived by the scientific method.

Id. at 590 (“Science is not an encyclopedic body of knowledge about the universe. Instead, it represents a process for proposing and refining theoretical explanations about the world that are subject to further testing and refinement,” (quoting Brief for the American Association for the Advancement of Science and the National Academy of Sciences as Amici Curiae in Support of Respondent, at 7-8, Daubert, 509 U.S. 579 (No. 92-102), 1993 WL 13006281)).
courts to ask “Is it science?” and not “Is it good science?” Similarly, publication in peer-reviewed journals and general acceptance in the scientific community demonstrate that the evidence is sound science that can be taken seriously; they do not measure the strength of, or confidence in, any reported effect. Insofar as either party goes ahead and provides test results, error rates, and the like, that information will, of course, be highly relevant to the admissibility determination under Rule 702 (or else to judgments about the evidence’s weight in the merits determination, if it is admitted). But those assessments of explanatory power speak only to whether and to what degree the evidence will be helpful to the trier of fact—a distinct consideration under Rule 702—and not to whether the evidence counts as scientific knowledge in the abstract.

Courts as institutions are, as it turns out, quite well suited to conducting the “Is it science?” inquiry that Daubert mandates. The weaknesses ascribed to their handling of scientific evidence become relevant, if at all, only at the second-stage assessment—whether the evidence is sufficiently strong to be useful to the factfinder if admitted—or in weighing the evidence at trial. In part, courts do a decent job performing the threshold inquiry because the standards that Daubert imposes do not require understanding all the ins and outs of any particular scientific claim. But beyond that, adversary court proceedings allow for, and indeed encourage, litigants to get at the heart of what science is—to educate the factfinder about the scientific method—in order to show whether the evidence being offered passes muster as science. Although litigants who take the additional pains to show how the scientific community would evaluate the strength of their experts’ findings may well invite more refined analysis of the probative value of their evidence, the fact that courts may often accept those invitations should not blind us to how much intersubjective validity the threshold inquiry embodied in Daubert alone provides. In all events, the basic determination that some opinion is scientific is a prerequisite to any second-order tests for greater intersubjective validity that a court may choose to employ.

To illustrate the point, let me draw on my experience litigating the intelligent-design case—Kitzmiller v. Dover Area

---

School District”—as the Kitzmiller trial and the decision that came out of it bring into sharp relief the difference between the “Is it science?” and “Is it good science?” inquiries. Kitzmiller involved a battle of experts over whether intelligent design is a scientific theory or a nonscientific, religious view—though with a substantial measure of the good-versus-bad-science debate thrown in. The Supreme Court had long since held that the Establishment Clause prohibits teaching creationism (a religious view) in public schools, even if that view is dressed up to look like science. So one of the two central questions in Kitzmiller was whether intelligent design is simply a repackaging of creationism to look like a scientific claim. The Dover school district tried to recast the controversy, however, as an internecine dispute among scientists about whether intelligent design is good or bad science, without worrying overmuch about how anyone might think that supposed debate would turn out. Indeed, the school district’s expert witnesses themselves contended only that intelligent design is science in its infancy, with one characterizing it as no better than “fringe science.” As a litigation strategy, that approach was entirely rational. The school district did not need to make any more robust claims about intelligent design because while the Establishment Clause forbids teaching religion in public schools, it places no restriction on teaching science (no matter how poor or misguided or flaky that science might be). So the claim that intelligent design is science would, if true, have gone a long way toward insulating the school district’s actions from constitutional challenge.

16 See id. at 716-23, 735-46.
18 Kitzmiller, 400 F. Supp. 2d at 711-12. The other issue central to the case was whether the Dover school board had a religious objective when it added intelligent design to the school district’s biology curriculum. Id. at 762-63. Both questions mattered because, under the Establishment Clause, governmental action is unconstitutional if either its primary purpose or its primary effect is to advance religion. Edwards, 482 U.S. at 582-83 (citing Lemon v. Kurtzman, 403 U.S. 602, 612-13 (1971)).
19 Kitzmiller, 400 F. Supp. 2d at 738.
20 It would not have done so absolutely, however, because a decision to teach or to refrain from teaching even genuine science would violate the Establishment Clause if the reason for the choice was school officials’ desire to tailor the science curriculum “to the principles or prohibitions of any religious sect or dogma.” Epperson v. Arkansas, 393 U.S. 97, 106 (1968).
In a thorough, carefully reasoned opinion that has earned praise from both the scientific and legal communities, Judge John E. Jones III of the United States District Court for the Middle District of Pennsylvania found that intelligent design does not obey science’s methodological ground rules, but instead invokes divine causation: It explains the diversity of life on earth by saying, in essence, that “God did it.” Hence, intelligent design is not science, good or bad, mainstream or fringe; it is, the court concluded, a religious view.

What Kitzmiller shows in spades is that the adversary system and the rules of evidence serve incredibly well to allow lawyers and expert witnesses to inform judges and juries about what science is and how it works, so that the factfinders, in turn, can make at least gross distinctions to screen out nonscientific beliefs dressed up to look like science. As Margaret Talbot explained in an account of the Kitzmiller trial published in the New Yorker:

You sometimes hear it said that a courtroom is not a proper venue for debating science. In this case, it proved to be an ideal forum. . . .

21 See, e.g., Kitzmiller, 400 F. Supp. 2d at 718-22.
22 To be sure, Judge Jones’s opinion in Kitzmiller did not just make gross distinctions. Quite the contrary; I am told that Kitzmiller is now required reading in many undergraduate- and graduate-level science courses, not just because of its clear explanation of the basics of evolutionary biology, but also because of its precise exposition of the nature of science itself.

Courts need not, however, speak in every case involving scientific-expert testimony (or its pseudoscience counterpart) with the precision and detail that the Kitzmiller court did. As argued above (see supra Part II), the degree to which courts must explain and publicly justify their judgments turns, at least in part, on the importance and the publicity of the case. The stakes in Kitzmiller were extraordinarily high: Against a national movement to market intelligent-design creationism to public schools, to state boards of education, and to legislatures, there stood eleven parents who were not only vindicating their own right to decide what religious education their children would receive, but also acting as proxies for concerned parents and defenders of science everywhere. With so much on the line for so many, on both sides of the issue, the court’s legitimacy to decide the case, and thereby to take a step toward quelling the larger culture clash over science education and religious control, depended on Judge Jones’s issuing an opinion that would provide an almost unprecedented level of public justification for the ruling. See Katskee, supra note 4, at 158-61.

The trial . . . allowed the lawyers to act as proxies for the rest of us, and ask of scientists questions that we’d probably be too embarrassed to ask ourselves. In a courtroom, you must lay an intellectual foundation in order to earn a line of questioning—and so the lawyers stripped matters neatly back to the first principles of science.23

What Talbot found so compelling was that, far from misleading the court about science, the parties’ presentations of expert testimony in an adversary proceeding informed by the principles in Daubert revealed the nature and core characteristics of science and scientific truth, thus allowing the court (and everyone else) to assess whether the intelligent-design movement could legitimately claim to be a scientific enterprise.24

The fact that the legal proceedings worked so well to expose intelligent design as a nonscientific, religious view resting on belief in a supernatural creator is an especially strong indicator of what courts and the adversary system can accomplish. After all, the intelligent-design movement’s grand strategic plan—its raison d’être—is to recast religious belief as a simulacrum of science in the hope that judges will have neither the skill nor the patience to look behind the façade.25 If an adversary proceeding applying Daubert-type criteria could prove so effective for stripping away that façade, built up over

24 I should note that there were no Daubert hearings in Kitzmiller. The court’s first opportunity to evaluate the reliability of the expert testimony came during the six-week bench trial. But while neither party formally invoked Daubert, the Daubert factors were the subtext of both sides’ expert cases because Daubert so effectively encapsulates the scientific community’s own understanding of what science is.

The school district did challenge the testimony of Barbara Forrest, a philosopher and social historian who gave expert evidence for the plaintiff-parents about the history and character of the intelligent-design movement. The school district’s counsel sought, unsuccessfully, to discredit Dr. Forrest in order to try to persuade the court to exclude her testimony about the movement’s inherently religious aims and strategic plan. See Defendant’s Brief in Support of Motion in Limine to Exclude the Testimony of Barbara Forrest, Ph.D., at 1, Kitzmiller, 400 F. Supp. 2d 707 (No. 04-CV-2688), 2005 WL 3628818. But neither party made any attempt to exclude any scientific—or putatively scientific—testimony.

25 As the Kitzmiller court found, and as many of us have explained elsewhere, intelligent-design creationism is a repackaging of so-called creation science (itself a repackaging of straightforward biblical creationism) to try to overcome the Supreme Court’s ruling in Edwards, which prohibited the teaching of creation science in public schools. See Kitzmiller, 400 F. Supp. 2d at 722; BARBARA FORREST & PAUL R. GROSS, CREATIONISM’S TROJAN HORSE: THE WEDGE OF INTELLIGENT DESIGN (2004); Matthew J. Brauer, Barbara Forrest & Steven G. Gey, Is It Science Yet? Intelligent Design Creationism and the Constitution, 83 WASH. U. L.Q. 1, 22-23 (2005); Katskee, supra note 4, at 141-50.
decades by a dedicated and well-funded cadre whose sole objective was to deceive courts, legislatures, and the public into accepting a nonscientific, religious belief as a scientific truth-claim (so that religion could then be slipped into public-school curricula), surely it is not too much to expect courts in general to uncover the far less sophisticated attempts to portray nonscientific views as science that parties might cook up on the fly and try to spoon-feed to a judge or jury in a single case.

There is also a simpler reason why the institutional-competence problem is not so serious as many suppose—one that brings me back to my initial focus on courts' institutional objective. Even if courts may not infallibly identify the best science when they are confronted with genuine disputes in which both sides properly invoke the scientific method and competing scientific truth-claims may appear to be in something close to equipoise, that limitation does not preclude courts' fulfilling their social function. If, as argued above, the point of a legal system is to resolve disagreements sufficiently fairly and reliably to ensure that people continue resorting to lawsuits rather than guns, even when they find themselves embroiled in disputes serious enough to threaten social stability, entirely correct decisions are not required in every instance. Nor, of course, can they reasonably be expected: Although ensuring perfect judgments all the time is an admirable aspirational goal, we all recognize that perfection of that sort is not possible for any human institution. Courts must do a decent job screening out pseudoscience—that is to say, nonscience posing as science—because it provides no basis for the shared understanding that leads to trust in judgments. But as long as courts achieve the far more modest aim to foster that trust, the rulings that they issue, after weighing legitimate but competing scientific claims and applying recognized modes of legal reasoning, will have sufficient intersubjective validity to make the judgments acceptable to most people most of the time. In that case, courts will be able to fulfill their basic social function to manage conflict, irrespective of whether they parse scientific claims exactly right in any particular case—much less whether they invariably and infallibly discover ultimate truth.
Revisiting the History of Scientific Expert Testimony

Tal Golan†

INTRODUCTION

This Article provides a broad historical narrative of scientific expert testimony in the adversarial courtroom, from the late eighteenth century to the present. No such narrative exists in the current literature and it would good to have one, considering the growing attention to the topic in the last two decades. This Article identifies four large-scale developments that were involved in the shaping of modern scientific expert testimony as we have come to know it today: the place of the expert in the courtroom; the nature of the expertise deployed; the amount and importance of expert testimony; and the treatment of expert testimony in the courtroom.

The first development was part of the fundamental transformation of the English legal system that took place during the later part of the eighteenth century and came to be known as the Adversarial Revolution. This development changed the position of the expert in the courtroom, shifting it from a neutral court-appointed position to that of a partisan witness, chosen and paid by the parties. The second development also started during the second part of the eighteenth century. It involved changes in the scientific community that began to narrow its focus to the inanimate world and provided first indications of practical utility. Legal historians have paid little attention to this important development, which redefined the legal range of recognized expertise and introduced into the courtroom a new figure—the proto-scientist, who functioned like a skilled professional but cogitated like a natural philosopher. The third development was associated with the Industrial Revolution and with the further development of science. Originating in late eighteenth-

† History Department, University of California, San Diego.
century England and extending in America well into the late nineteenth century, the Industrial Revolution brought to the courts a rising tide of cases involving technological and scientific argumentation. This development established the newly defined scientific expert witness as a pivotal but also highly problematic figure in the modern courtroom. The fourth development took place during the twentieth century and was driven by the professionalization of science and by society’s growing dependence on science. This growing dependence forced the courts to take an active role in managing the deployment of science in the courts, and the professionalization and standardization of twentieth century science offered the courts new means to do so. Consequently, the twentieth-century trial judge turned into an active gatekeeper, charged with the responsibility of screening unreliable scientific evidence away from the jury.

This Article’s first five parts are structured around these developments. Part I discusses the emerging role of the partisan expert witness during the Adversarial Revolution. The main argument in this Part is that the newly defined expert witness was not conceived as a premeditated judicial solution to the problems of deploying expertise in the new adversarial courtroom. Instead, the partisan expert witness emerged as a necessary exception, the only source of information the new system could not rationalize under its evolving doctrines. And, as such, it would stay—an incompatible yet indispensable figure in the modern adversarial courtroom.

Part II discusses the changing nature of expert testimony during the late eighteenth century. To that end, it focuses on the 1782 case *Folkes v. Chadd*, a legal episode that became iconic in the legal literature as the origin story for modern expert testimony. A close look demonstrates that *Folkes v. Chadd* showed little judicial concern with the problems awaiting partisan expertise in the modern adversarial courtroom. Instead, the case revolved around a different important problem—the legal status of a nascent species of experts, the proto-scientists, or Newtonian philosophers as they were then styled, who propounded in court theories whose station on the legal continuum between fact and speculation were yet to be settled.

Parts III and IV discuss the nineteenth-century developments of scientific expert testimony in England and America, respectively. They demonstrate that while the volume of scientific expert testimony constantly increased during the
nineteenth century, the respect paid to it by the courts and the public constantly diminished. The main argument here is that by the late nineteenth century the putative problem of scientific expert testimony had already acquired most of the features that today are blithely assumed to be new. These two Parts further suggest that, although first raising its head in English courts, it was in America that the problem of expert testimony reached its fullest expression.

Part V discusses the twentieth-century attempts to control the problem of scientific expert testimony. Much of the current scholarship portrays the controversies surrounding scientific expert testimony as a late twentieth-century development, the result of the difficulties of the courts and the lay jury in handling the growing volume and complexity of modern science. The discussion in Part VI paints a more nuanced picture. Far from being a late twentieth-century pathology, the putative problem of scientific expert testimony has been chronic for over two centuries. Moreover, during the twentieth century, the courts were able to take advantage of the professionalization of science and the standardization of the market of expertise and actually improved their ability to control the performance of science in the courtroom.

I. THE EMERGING ROLE OF THE PARTISAN EXPERT WITNESS

Common law has long acknowledged the importance of scientific advice in cases where the disputed facts were such that the courts lacked sufficient knowledge to draw from them an informed decision. In 1554, an English judge declared:

> [If] matters arise in our law which concern other sciences or faculties, we commonly apply for the aid of that science or faculty which it concerns. Which is an honourable and commendable thing in our law. For thereby it appears that we do not despise all other sciences but our own, but we approve of them and encourage them as things worthy of commendation.¹

Over the centuries the legal system had developed two procedural options to deploy such sciences in the courtroom by experts who, from their special training and experience, could

instruct the court and the jury in regard to the disputed facts. The first option was to call them as jurors. The second was for the court to nominate them as consultants whose advice the court or the jury could adopt as they pleased.

There was also a third option, which was for the parties to call experts as witnesses testifying on their behalf. However, unlike court experts and expert juries, there was no special procedure that would define such witnesses as experts. Thus, physicians and surgeons testified in criminal, insurance, and will cases; surveyors testified in property cases; linguists testified concerning the meaning of Latin phrases used in contracts; merchants concerning the particular customs and norms of trade; tradesmen concerning the quality of particular goods; ship builders concerning the state and construction of vessels; other artisans concerning their respective subjects of mechanical skill, and so on and so forth. However, these testifying experts were not clearly differentiated from all other lay witnesses, who often were also allowed to testify to their opinions, especially if it was based on their direct knowledge of the facts of the case. Thus, in the absence of a procedure that would define witnesses as experts or a theory that would restrict lay witnesses from testifying to their opinions, the testifying experts were regarded and treated merely as witnesses.

The theories and practices that differentiated the testifying experts from all other lay witnesses evolved only late in the eighteenth century, as part of a larger transformation of the English legal system that came to be known as the Adversarial Revolution. This so-called revolution was primarily associated with the expanding presence of lawyers in criminal proceedings. Until the 1700s, lawyers were generally kept out

---


4 Beuscher, *supra* note 2, at 1109-10.


6 *Id.*


of criminal trials.\textsuperscript{9} The judges dominated the proceedings, and evidence was mostly adduced by direct in-court altercation between the accuser, the accused, and the witnesses, and by the judge, who examined the parties and the witnesses himself.\textsuperscript{10} In this environment, which kept criminal proceedings quick and simple, testimonial constraints had little, if any, meaning.\textsuperscript{11} Thus, although the common law requirement that regular testimony be limited to personal knowledge based on experience was old, lay witnesses were nevertheless allowed to testify to their opinion or present hearsay evidence. And if an objection was raised, the courts were content with allowing it to go to the weight of the evidence rather than to its admissibility.\textsuperscript{12}

By the 1730s, defense counsel began to participate in regular criminal proceedings. Not yet allowed to directly address the jury, they were permitted to gather and adduce evidence and examine and cross-examine witnesses. The reasons for this “most remarkable change” to ever take place in English criminal law (as the noted Victorian judge and legal historian Sir James Fitzjames Stephen called it\textsuperscript{13}) are not fully clear. It has been suggested that the appearance of the defense lawyer was, at least partially, a response to the expanding prosecutorial capacities of the Crown.\textsuperscript{14} Whatever the reasons were, by the second part of the eighteenth century, according to Stephen, “A practice sprung up . . . by which counsel were allowed to do everything for prisoners accused of felony except addressing the jury.”\textsuperscript{15} The active participation of the lawyers slowly reshaped the processes of criminal litigation. Outside the courtroom the lawyers gave new significance to pretrial activities such as preparing records and seeking out witnesses.

\textsuperscript{9} J.M. Beattie, Crime and the Courts in England 352-56 (1986); Thomas Andrew Green, Verdict According to Conscience: Perspectives on the English Criminal Jury 1200-1800, 135-36 (1988); Langbein, supra note 8, at 310.
\textsuperscript{11} Cockburn reports that one judge heard as many as fifty Crown cases a day. Id. at 109.
\textsuperscript{15} Stephen, supra note 13, at 424.
Inside the courtroom they increasingly took over the examination of witnesses, developed the techniques of cross-examination to perfection, and even established the right to argue points of law.\(^{16}\)

Looking back, William Best, a leading mid-nineteenth-century authority on legal evidence, wrote that “the necessary consequence of [allowing defense counsel in criminal cases] was that objections to the admissibility of evidence were much more frequently taken, the attention of the judges was more directed to the subject of evidence, their judgments were better considered, and their decisions better remembered.”\(^{17}\) Indeed, still prohibited from speaking directly to the jury, the lawyers mostly fought their battles over the content and the presentation of the evidence before the jury in terms of evidentiary objections.\(^{18}\) By the end of the eighteenth century these evidentiary battles had produced two powerful legal doctrines: the hearsay doctrine, which attempted to limit testimony to information based solely on personal observation, and the opinion doctrine, which sought to control the form in which witnesses communicated their perceptions to the jury, requiring them not to use inferences where the subject matter is susceptible to factual statements.\(^{19}\)

These two powerful evidentiary doctrines rendered the expert witness into a distinct legal entity. However, they did so indirectly, by curtailing the privileges of all other testimonial sources until only the testifying expert was left as the last but necessary exception to the rules—a witness who did not have to observe the facts of the case personally but nevertheless was allowed to pronounce an opinion on them in court. The expert


\(^{19}\) See Landsman, supra note 14, at 572; Gallanis, supra note 17, at 530-37; see also Hand, supra note 2, at 44-45; John H. Wigmore, The History of the Hearsay Rule, 17 HARV. L. REV. 437, 448 (1904).
witness was not conceived therefore as a deliberate judicial solution to the problems of deploying expertise in the new adversarial courtroom. Instead, the expert remained as the only source of information the new system could not rationalize under its evolving doctrines—a freak, if you will, in the new adversarial courtroom.

The new adversarial system not only redefined the role of the expert as a partisan witness but also had a dramatic effect on the deployment of expert testimony in the courtroom. Traditionally, experts, whether as part of the jury or as court advisors, were summoned and controlled by the court, which conferred on these experts a large degree of impartiality. But during the late eighteenth century, as the court gradually assumed a neutral position, as the litigants assumed responsibility for developing their own proof in court and summoned their own experts to represent them before the jury, and as adversarial ideology was given free reign, a demanding problem seemed to emerge: how to ensure that in this adversarial environment the lay jury would still have access to reliable expert guidance when the jury needed it. Surprisingly, though, an analysis of major late eighteenth and early nineteenth-century rulings finds little judicial awareness of this problem. While the judges were certainly busy with the delicate act of balancing the demands of the increasingly defined rules of evidence with the growing supply of expert testimony, they seemed far less concerned that the practice itself of calling experts as partisan witnesses was expanding.

The absence of judicial anxiety about expert testimony is all the more remarkable if we take into account that there was ample judicial dismay about lay witnesses for hire. The late eighteenth century was a period in which the slightest interest in the result of the trial rendered the witness unreliable. Persons were not allowed to testify in cases in which they had financial interest. Husbands and wives were forbidden from testifying for or against each other. Even the parties to the lawsuit themselves, by the same reasoning, were not allowed to testify. Why then the expert witness for hire?

---

21 Id. at 106-16.
22 Id. at 119-21.
23 Id. at 116-19.
We will return to this question toward the end of the next section.

II. THE CHANGING NATURE OF EXPERTISE

The practice of deploying expertise in the new adversarial courtroom was not the only thing that was changing during the late eighteenth century. A second change, equally important, was in the nature of the expertise deployed in the courtroom. For centuries, experts summoned to give their opinions in the courtroom were men of large and tested experience, who, from their special training and experience, could instruct the court and the jury in regard to the disputed facts. The specialized knowledge these experts brought to the courts was expected to be based on personal and empirical observations, readily traceable to the specific training and experience of the particular expert pronouncing it. Distilled through ages of legal experience and immortalized in the early seventeenth-century writings of Lord Chancellor Francis Bacon, this legal epistemology disdained abstract explanations and stressed the necessity of direct observational data in processes of proof.24

By the late eighteenth century, however, a new culture of expertise began its rise to dominance, which defied this epistemology. This was the culture of science, confident in its ability to discern the hidden laws of nature, however subtle their workings were. At the start of the eighteenth century, this was still a bookish culture that studied nature in general, but by the end of the century it had narrowed its focus to the inanimate world, supplemented books with experiments, borrowed some mathematics, and gave indications of practical utility. Its practitioners, who styled themselves men of science (the word “scientist” was yet to be invented) or Newtonian philosophers (after their great leader, Isaac Newton), may have still theorized like natural philosophers, but they increasingly acted like skilled professionals. They reasoned from first principles but concerned themselves with the observable, the measurable, and the practical. By the end of the eighteenth century, they became central to Britain’s booming economy—revolutionizing agriculture; inventing and improving engines, pumps, and other machinery; designing and overseeing the

construction of waterways, bridges, and harbors. It was not long before the courts were asked to consider the status of these new experts in the courtroom.

The matter was laid squarely before the courts in the 1782 civil case of Folkes v. Chadd. In this celebrated legal episode, several experts were summoned by the litigants to the courtroom to testify before the jury to what in their opinion caused the decay of a certain harbor on the Norfolk coast of England. The testimony of one of those experts, a prominent Newtonian philosopher named John Smeaton, was excluded by the trial judge on the ground that his theoretical explanations were “matter of opinion, which could be no foundation for the verdict of the jury.” On appeal, Lord Mansfield, Chief Justice of the Royal Court of King’s Bench, found the silencing of the philosopher to be an error and granted a new trial on the ground that Smeaton’s theory “was very proper evidence.”

Lord Mansfield’s opinion in Folkes v. Chadd has served in the legal literature as the principal precedent that shaped the dominant option of using expert knowledge in the modern adversarial courtroom—calling experts to testify before the jury as partisan witnesses. It has been unanimously declared “the foundation of the rules governing expert evidence.”

James Thayer maintained that it inaugurated the new practice of calling experts as partisan witnesses before the jury. John Wigmore claimed that it certified experts to pronounce opinion without being personally familiar with the facts of the case. And recently, Stephan Landsman considered it “the court’s seal of approval on the whole adversarial apparatus

31 James Bradley Thayer, A Selection of Cases on Evidence at the Common Law 672-73 (1900).
32 4 Wigmore, supra note 7, § 1917, at 103.
including contending experts, hypothetical questions, and jury evaluation."

What follows below is a close study of *Folkes v. Chadd*. The study confirms that it was indeed a moment of great importance for the deployment of expert testimony, but for reasons other than those so far suggested by historians. In *Folkes v. Chadd*, Lord Mansfield was intent neither on inaugurating a new practice of calling experts as partisan witnesses before the jury nor on solving the difficulties that awaited such practice in the adversarial courtroom. Instead, Mansfield was trying to clarify the legal status of a nascent species of experts—proto-scientists, who presented in court knowledge claims whose legal status was yet to be settled.

A. *Folkes v. Chadd, Round I*

During the eighteenth century a new breed of capitalistic landlords evolved in the northern county of Norfolk, England, anxious to experiment with new methods of farming that would produce ever larger surpluses for sale. By the end of the eighteenth century, Norfolk's innovative husbandry came to be known worldwide as the Norfolk System and its harbors were shipping more grain than the rest of England combined.

One of these harbors was the tidal harbor of the town of Wells. Having no river or other inland fresh water source, Wells Harbor had relied for centuries on the strength of the ebbing tide to scour the rich silt that the violent tides and winds of the North Sea constantly deposited at its bottom. Overflowing much of the coast, the tide created a body of water covering thousands of acres. With the ebbing of the sea, much of this water ultimately collected in the main channel of Wells Harbor, providing sufficient scouring to maintain its depth and safety.

---


In the 1720s, Wells Harbor began to show signs of decay. The parts of the harbor furthest from the sea became increasingly clogged, and by mid-century the quay became inaccessible to shipping and the greater part of the cargo had to be carried to and from the town by lighters. Wells merchants and the ship owners blamed the deterioration of their harbor on some of the local landlords, who embanked and reclaimed from the North Sea significant tracts of land on both sides of the harbor’s main channel. These embankments, Wells inhabitants believed, greatly weakened the body of backwater available for scouring their harbor, thereby causing it to choke up.

Fearing for the loss of their livelihood, Wells’ inhabitants tried to save their harbor. They financed the constructions of two artificial sluices that would scour the harbor and keep it open, but to no avail. The conditions of the harbor continued to deteriorate. Finally, in 1780, confident that the embankments were the principal cause of its troubles, the Harbor’s board of commissioners decided to take legal action against one of the biggest landlords, whose embankment, it felt, was the most harmful to its harbor.

The trial took place in August 1781, at the summer Assizes in Norwich. The questions put before the jury were whether the said embankment harmed the harbor and whether the harm justified the cutting of the embankments. The trial lasted two days. During the first day, the commissioners’ lawyers marched a long line of traditional experts to the witness stand. Pilots, mariners, and other seamen, who had spent their entire life at the harbor, testified from their personal experience to the rapid deterioration of the harbor following the construction of the said embankment. On the

---


37 Nathaniel Kent, General View of the Agriculture of the County of Norfolk with Observations on the Means of Improvement 9-10 (London, C. Macrea 1794); Cuchlaine A.M. King, Beaches and Coasts 274-75 (1959).


39 See id. at 11-12 (citing Tidal Harbours Commission, supra note 35, at 444).

second day, it was the landlord's lawyers' turn. They presented only one expert, Robert Mylne, Fellow of the Royal Society and the owner of a thriving London-based architecture-engineering practice.41

Before the trial, upon the invitation of the landlord, Sir Martin Browne Folkes, Mylne traveled from London to Wells to study its harbor.42 He concluded that the troubles of the harbor could not be attributed to Folkes's embankment or to any other embankment. What caused the decay of the harbor were the vast quantities of materials discharged at the immense western estuaries by the six rivers of Ouse, Nene, Witham, Trent, Wharfe, and Swale, and deposited along the north coast of Norfolk by the strong tides and winds of the North Sea.43 The defense experts, Mylne explained to the jury, had been misled by their perceptions. The filling up of the harbor they had witnessed was but a mere link in the temporal chain of causes imperceptible to lay observation, a chain perceptible only to those intimated with the hidden operation of nature.44

The authoritative testimony of the famous engineer and Fellow of the Royal Society, who had made a special study of the cause of the trial, made a strong impression on the jurors, who, we are told, “relying on the weight of Mr. Mylne's abilities of knowledge, and not having the least doubt of the truth of his evidence, found a verdict for the Folkes.”45 Wells' inhabitants were outraged by the arrogance of the metropolitan expert. How could a foreigner to the county, lacking intimate knowledge of the facts of the case, claim, on the basis of the shortest inspection, to recognize forces at work unobserved by their own experienced experts, who had spent their entire life at the harbor? The commissioners' lawyers moved for a new

---

41 Robert Mylne owned a thriving private architecture/engineering/surveying practice. His specialty was bridges but his designs were various. Among other things he was appointed architect to St. Paul's Cathedral and chief engineer of the great New River Company that supplied water to London. ALBERT E. RICHARDSON, ROBERT MYLNE: ARCHITECT AND ENGINEER, 1733 TO 1811 (1955).

42 Id. at 114-16.

43 ROBERT MYLNE, MR. MYLNE'S REPORT, ON HIS SURVEY OF THE HARBOUR, &C. OF WELLS, IN NORFOLK 4 (Apr. 28, 1781).

44 Id.

45 See GOLAN, supra note 38, at 23; Notes of the Ordering for a 2nd Trial, Folkes, Bart & All Agst. Chadd, Esq. & Others, Norfolk Records Office, MF/RO 504/2, MS 486 [hereinafter Notes Ordering 2nd Trial]. This was a standard legal argument for a new trial made by a party who felt cheated by a case falsely made at the trial that it had no reason to expect and therefore could not come prepared to answer. See, e.g., Hartley v. Buggin, 3 Doug. 39, 40, 99 Eng. Rep. 527, 528 (1781) (Mansfield, J.).
trial “on the ground that the defendants were surprised by the doctrine and reasoning of Mr. Milne.”

The royal judges of the King’s Bench, who convened to discuss the commissioners’ request, agreed that the commissioners should have had the opportunity to counter Mylne’s performance with their own experts. “[I]n matters of science,” they dictated, “the reasonings of men of science can only be answered by men of science.” A new trial was therefore granted and was promptly set for the following summer term of the Norfolk assizes in July 1782. To avoid additional surprises in this important litigation, which “has influenced the whole county of Norfolk, and perhaps the whole country may be affected by it,” the judges directed the parties to exchange between them in writing, before the new trial, the opinions of the experts whom they intended to produce in court “so that both sides might be prepared to answer them.”

B. The Experts and Their Reports

The second round of the Wells Harbor litigation, it was clear to all, was going to be decided upon the opinions of men of science. Unwilling to be caught off guard again, the commissioners of Wells Harbor went out and recruited four senior experts—John Grundy, Joseph Nickalls, Thomas Hogard, and Joseph Hodkinson—to represent them in the coming trial. Grundy was an experienced engineer whose specialties included the improvement of river navigation and the drainage of adjacent low lands. Nickalls served as an appointed engineer to the Thames Commissioners, representing their cause in Parliament against loud opposition from promoters of competing canal schemes. Hogard specialized in fen drainage and served as a commissioner for several fen drainage schemes in Lincolnshire. Hodkinson was Vice-President of the Society of Civil Engineering and one of the most respected land-surveyors in England. No doubt, the commissioners of Wells Harbor prepared themselves well for the coming scientific battle.

48 See GOLAN, supra note 38, at 24 (citing Notes Ordering 2nd Trial, supra note 45, at 67).
49 Folkes, 3 Doug. at 157.
Sir Martin Browne Folkes added just one more expert to his team. But he chose exceptionally well. His expert was John Smeaton—Fellow of the Royal Society and a civil engineer who was considered the highest authority on harbors in the kingdom.51 By 1781, Smeaton had been consulted on more than thirty different harbors in England and Scotland.52 Furthermore, he was responsible for the recent successful rescue of Ramsgate Harbor, one of England’s largest harbors, from the vast amounts of sand that threatened to choke it completely.53 In addition to his flourishing engineering practice, Smeaton also developed a prominent scientific career. Smeaton had contributed fifteen papers to the Royal Society’s Philosophical Transactions. One of them, describing his experiments with waterwheels and windmills, was awarded the Copley Medal, the society’s highest award, thereby establishing Smeaton’s reputation as one of the Kingdom’s most celebrated natural philosophers.54

In March 1782, Smeaton traveled to Wells and spent three days there studying the harbor. Then, he returned to London to study further the history of the harbor, read the evidence produced in the first trial, and write his report.55 Like Mylne before him, Smeaton also concluded that the decay of Wells Harbor was caused not by the hand of man but by the hidden hand of nature. “To have a clear and comprehensive view of the cause of [the] decay,” Smeaton wrote, “it will be necessary to shew the natural causes by which the port of

---

52 M. Dixon, Some Account of the Life, Character, and Works, of Mr. John Smeaton, F.R.S., in 1 REPORTS OF THE LATE JOHN SMEATON, F.R.S. xv-xxx (London, Longman et al., 2d ed. 1837) (hereinafter SMEATON REPORTS); see Skempton, JOHN SMEATON, supra note 50, at 27-34.
53 Skempton, supra note 52, at 65-70.
54 J. Smeaton, An Experimental Enquiry Concerning the Natural Powers of Water and Wind to Turn Mills, and other Machines, Depending on a Circular Motion, in 51 PHILOSOPHICAL TRANSACTIONS, GIVING SOME ACCOUNT OF THE PRESENT UNDERTAKINGS, STUDIES AND LABOURS, OF THE INGENIOUS, IN MANY CONSIDERABLE PARTS OF THE WORLD 100-74 (London, L. Davis & C. Reymers 1760).
Wells has been formed.\textsuperscript{56} Thus, Smeaton commenced with a theoretical discussion concerning the general principles that govern the creation and decay of tidal harbors by the natural forces of the sea and the weather.

There was a time, Smeaton hypothesized, when nothing more than naked sand lay against the bare coast upon which the town of Wells stood. At that stage, the tide flowed and ebbed uniformly as a continuous sheet of water. But as the steady deposition of tidal silt increased the height of the sand, tidal waters were eventually left behind that started to cut gullies on their way back to the sea. As the land steadily increased in height, greater bodies of water were left on it and the gullies increased in number, depth, and size. Eventually, Smeaton wrote:

\begin{quote}
  \text{[I]f all were ultimately collected into one, as has been the case with the channel of Wells Harbor, the scour would be sufficient to maintain a channel through which vessels might be brought from the sea, and thus an useful harbour would be formed, which would increase in depth and utility by the continuance of the forming powers, but yet, only to a certain degree.}\textsuperscript{57}
\end{quote}

With the continual elevation of the coast above the reach of higher and higher tides, a height would eventually be reached from which the process would start to reverse itself. The volume of the water left upon the coast would start to diminish. Deprived of the scouring action of the backwater, first the gullies and finally the main channel would be choked. The true story of Wells Harbor, then, Smeaton recounted in his report, was that

\begin{quote}
  the progressional operation of nature, which originally formed the harbour of Wells and brought it to maturity, has also occasioned it to grow more and more into a state of decay; and will finally close it up, and convert into a firm ground, fit for arable purposes, and those of pasturage, the very spot where ships have rode at anchor . . . .\textsuperscript{58}
\end{quote}

Two weeks after Smeaton’s departure, the Commissioners’ scientific team also arrived to Wells and spent ten days there, inspecting the harbor and its surroundings.\textsuperscript{59} They did not bother themselves with the general laws governing the dynamics of tidal harbors. That would have been

\textsuperscript{56} 2 Smeaton Reports, \textit{supra} note 51, at 149.
\textsuperscript{57} \textit{Id.} at 151.
\textsuperscript{58} \textit{Id.} at 157.
\textsuperscript{59} See Golan, \textit{supra} note 38, at 26, 32-34.
to play the part of the natural philosopher. They were civil engineers, practical men whose business was to estimate the relative effectiveness of the backwater deprived by Folkes's embankment. To that end, they concerned themselves with the mappings, measurements, and calculations of the current state of the land; its existing areas, enclosed or not; the locations and orientations of the neighboring creeks; the length and the depth of the main channel; the directions of the winds and the tides; etc.60

Back in London, Hodskinson promptly delivered a detailed report to his employers. While Smeaton based his report on his dynamic theory of the tidal harbors, Hodskinson grounded his on interviews with the local harbor masters and on observations and measurements his team had made. The interviews convinced him that within local memory not only had the coast not grown in breadth or height, but, on the contrary, the sea had gained in many places along the coast. His copious calculations showed that Folkes’s embankment eliminated 214,122 tons of water, about a third of the volume of water previously available for scouring the harbor.61 Hodskinson concluded:

Upon the whole, I am of opinion that the present bad and ruinous state of the harbor is to be in a great measure, if not wholly, imputed to the [said] embankment . . . and that if the tide of the sea is permitted to flow and reflow in its ancient course and manner . . ., the navigation will be supported and maintained in a safe, useful and commodious state by the natural operation of the tidal waters thereon.62

C. Folkes v. Chadd, Round II

The second round of the Wells Harbor litigation began on July 25, 1782, before Henry Gould, Chief Justice of the Royal Court of Common Pleas, and a special jury. Two eminent counsels ran the trial. Leading the legal team for Folkes, unchanged from the previous trial, was Henry Partridge, a well-known barrister with strong Norfolk connections. This time the commissioners of Wells Harbor brought their own heavy legal artillery in the person of George Hardinge,

61 See GOLAN, supra note 38, at 32-33.
62 Hodskinson, supra note 60, at 449.
Barrister of the Middle Temple and solicitor-general to Queen Charlotte.63

The reports of the experts on both sides were made available to the jury a week before the trial. However, when the trial began, Hardinge chose not to call upon his experts to give oral testimony and be cross-examined. Instead, he summoned a long line of mariners and navigators to testify from their personal experience to the rapid deterioration of their harbor. Partridge, on the other hand, planned to repeat his successful strategy and summon his famous scientific expert to the witness stand to persuade the jury that it was nature, not his employer’s embankment, that was responsible for Wells Harbor’s troubles. However, when Partridge tried to call Smeaton to the stand, Hardinge objected. The illustrious Smeaton, Hardinge contended, should not be permitted to speak to the jury since his testimony concerning the hidden causes of nature “was matter of opinion, which could be no foundation for the verdict of the jury, which was to be built entirely on facts, and not on opinions.”64

Hardinge’s objection was characteristic of the evidentiary battles that flourished in the new adversarial courtroom.65 Central to these battles was the opinion doctrine, which sought to restrict witnesses from expressing their opinions where the subject matter was susceptible to factual statements. As John Wigmore summarized the status of the new opinion doctrine at the end of the eighteenth century: “[H]enceforth, the only question can be how far there are to be specific exceptions to it.”66 It was with this question in mind that Chief Justice Gould addressed Hardinge’s objection to Smeaton’s testimony.

Smeaton’s case seemed straightforward. He was a well-respected expert and the facts of the Wells Harbor case, which he had observed directly, fell well within his field of expertise, thus constituting a proper object for his expert opinion. Nevertheless, the defense counsel, Hardinge, had firm legal

65 Hand, supra note 2, at 37; Landsman, supra note 14, at 572.
66 Wigmore, supra note 19, at 448. Collecting his data from political state trials, Wigmore may have overstated the status of the opinion doctrine in non-political criminal and civil proceedings. In 1782, while no longer embryonic, the opinion doctrine was still subject to the wide discretion of the individual trial judge. See Hand, supra note 2, at 45; Landsman, supra note 14, at 572; Gallanis, supra note 17, at 512-13, 530.
ground to stand upon when he maintained that Smeaton’s opinion did not rest on the facts of the case but on speculation that had no place in court. Smeaton, Hardinge argued, transgressed the established legal range of expertise. Expert opinions were supposed to be based on calculations and empirical observations, readily traceable to the particular training and experience of the particular expert pronouncing them. Smeaton, on the other hand, propounded in court a high-minded hypothesis about some natural processes, imperceptible to anyone but himself, which allowed him to shift the blame for the undisputed decay of the harbor from the obvious human hand of his employers to the hidden hand of nature. But what kind of training or experience could have qualified a person like Smeaton in 1782 as an authority on such matters? And what kind of legal reasoning could acknowledge such latent causes as a proper foundation for the verdict of the jury, which was to be built entirely on facts presented before them in court?

Descending from a long legal lineage, the seventy-two-year-old Chief Justice Gould was known for the strictness of his law.67 His logic, the logic of the common law, suspected elegantly constructed theories and stressed the importance of empirical data in processes of proof. Yet, Smeaton’s evidence was based on a hypothetical natural process that could have taken centuries and could not be measured, tested, or otherwise verified. Thus, Chief Justice Gould accepted Hardinge’s argument that Smeaton’s evidence indeed “could be no foundation for the verdict of the jury” and did not permit Smeaton to address the jury from the witness stand.68 With Smeaton and his imponderable science out of the way, Hardinge won the day as the jury gave a verdict for the commissioners, allowing them to cut the embankment that choked their harbor.69 Folkes’s lawyers immediately asked for a new trial on the grounds that their expert was improperly silenced.70

69 Id.
70 See id.
D. Mansfield’s Decision

The request for a third trial was laid before Lord Mansfield, Chief Justice of the King’s Bench and probably the most influential judicial figure of the eighteenth century.71 His famous decision is worthy of a lengthy quotation:

The facts in this case are not disputed. In 1758 the bank was erected, and soon afterwards the harbor went to decay. The question is, to what has this decay been owing? The defendant says, to this bank. Why? Because it prevents the back-water. That is matter of opinion:—the whole case is a question of opinion from the facts agreed upon. Nobody can swear that it was the cause; nobody thought that it would produce this mischief when the bank was erected. The commissioners themselves look on for above twenty years . . . . It is a matter of judgment, what has hurt the harbour. The plaintiff says that the bank was not the occasion of it. . . . Mr. Smeaton is called. A confusion now arises from a misapplication of terms. It is objected that Mr. Smeaton is going to speak, not as to facts, but as to opinion. That opinion, however, is deduced from facts which are not disputed—the situation of banks, the course of tides and winds, and the shifting of sands . . . Mr. Smeaton understands the construction of harbours, the causes of their destruction, and how remedied. In matters of science no other witnesses can be called. An instance frequently occurs in actions of unskillfully navigating ships. The question then depends on the evidence of those who understand such matters; and when such questions come before me, I always send for some of the brethren of Trinity House. I cannot believe that where the question is, whether a defect arises from a natural or an artificial cause, the opinions of men of science are not to be received. . . . I have myself received the opinion of Mr. Smeaton respecting mills, as a matter of science. The cause of the decay of the harbor is also a matter of science, and still more so, whether the removal of the bank can be beneficial. Of this, such men as Mr. Smeaton alone can judge. Therefore we are of opinion that his judgment, formed on facts, was very proper evidence.72

Mansfield’s decision has long served as the origin story for the rise of partisan expert testimony in the modern Anglo-American legal system. The first report on Folkes v. Chadd was published in 1831, as edited by Henry Roscoe, an experienced barrister who based his reports on the records of various leading judges. According to Roscoe, early nineteenth-century courts regarded Mansfield’s decision as “the principal case on

---


the admissibility of matter of opinion.” Roscoe further elaborated on what was to be learned from Mansfield’s decision: “Professional men, when examined on the subject of their art or science, are of necessity allowed to state their opinions . . . .”

The next significant reference to Folkes v. Chadd came sixty years later from Professor James Thayer of Harvard Law School. Mansfield’s decision, Thayer explained, represented the onset of judicial recognition in the modern practice of party-called expertise. Introducing Folkes in his influential textbook, Thayer wrote that, for a long time, “experts were thought of in the old way, as being helpers of the court . . . . But at last the modern conception came in, which regards the expert as testifying, like other witnesses, directly to the jury.”

One difficulty with Thayer’s reading of Mansfield’s decision lies in the fact that the practice of party-called expertise was not novel in 1782. As early as 1678, some of the most eminent physicians in the Kingdom, including future president of the Royal Society Sir Hans Sloan, were called by both sides to testify in a murder case, as to the causes of certain symptoms observed in an autopsy and on the general proposition as to whether a man could die of wounds without a fever. In the eighteenth century, party-called expertise was also documented in civil proceedings, noticeably in the growing area of patent litigation. The practice was recurrent in the growing textile trade during the 1760s and 1770s. Tax litigation and nuisance litigation also saw the deployment of party-called expertise with intense adversarial spirit. Indeed, Lord Mansfield himself had presided over many patent trials and was well familiar with the practice of party-called expertise. In the 1760s, for example, he presided over an

73 See Folkes, 3 Doug. at 160 n.(b).
74 Id.
75 Thayer, supra note 31, at 673.
important patent case that involved a prominent London optician, Peter Dollond, and a group of other London opticians, in a struggle for patent rights for the design of the refracting telescope.\(^7\) The case brought to the witness stand a line of experts, who testified to optical principles, previous designs, and trade secrets involved. \(^8\) In the summer of 1781, concomitantly with the Wells Harbor litigation, Lord Mansfield presided over another important case involving expert testimony—one which revolved around Richard Arkwright’s attempt to enforce his monopoly of the carding machine on the textile business. The attempt failed, but a second one succeeded in 1785, largely thanks to the testimony of scientific figures such as William Herschel, Erasmus Darwin, James Watt, and Robert Mylne, who vouched under oath to the validity of the patent’s principle and specifications.\(^8\)

The practice of party-called expertise was therefore not new in 1782. Still, one could maintain that the adversarial context was novel, and that Mansfield’s decision was the first to recognize and legitimate the practice in this new context. However, little in Mansfield’s decision could support such claim. In fact, Mansfield’s decision displayed a complete disregard to Smeaton’s appearance as a partisan witness. If anything, Mansfield’s opinion treated Smeaton as if he was a court expert. “When such questions come before me,” Mansfield reasoned, “I always send for some brethren of Trinity House.”\(^8\)

The Trinity House was a famous club of retired sea captains, and its brethren functioned as arbitrators and official court experts in cases arising out of events on the high seas.\(^8\)

---


\(^8\) Id. at 41.


\(^8\) Like other common law judges, Mansfield used to sit with an admiralty judge. The Trinity House brethren served as court experts in such cases. On occasion Mansfield adopted this practice in the Court of the King’s Bench. The MANSFIELD MANUSCRIPTS, supra note 71, at 146 n.377, 395 n.5.
Clearly, the deployment of court-nominated experts is not the precedent one would choose if intent on inaugurating a new practice of calling experts as partisan witnesses selected and paid for by the parties.

John Henry Wigmore, Thayer’s student and the leading early twentieth-century authority on evidence, also saw an origin story at the bottom of Folkes v. Chadd. Wigmore was aware that the practice of party-called expertise was not new to the period.84 He also recognized that expert witnesses were yet to be differentiated from lay witnesses, who were also allowed to testify to their opinions, especially if they were based on an intimate knowledge of the facts of the case.85 Wigmore concluded therefore that the distinctiveness of the modern expert witness sprang not from the license to testify to opinion (which was still shared with lay witnesses), but from the exclusive privilege to pronounce an opinion whether or not the expert had observed the facts of the case directly.86 It was this distinction, according to Wigmore, which made its first successful appearance in Mansfield’s decision in Folkes v. Chadd:

Here was a man [Smeaton], who had never seen the place, had no “facts” to add, and was going to give . . . his opinion upon the general question in doubt, the cause of the decay. Why should he do this? Why waste time in listening to numbers of such persons when the twelve men in the box have been specially selected for the very purpose of having their opinions serve as decisive? There would be only one reason for listening to such outside opinions, namely, that the witness was such a person that the jury would be really aided by his opinion.87

Thus, Wigmore claimed, Mansfield’s decision epitomized “the general recognition by the end of the 1700s, that there was a class of persons, i.e., those skilled in matters of science, who, though they personally knew nothing about the circumstances of the particular case, might yet, perhaps by way of exception, give their opinion on the matter.”88

Wigmore, we know, had his facts wrong. Smeaton not only had seen the place, but had also written a detailed report on his findings that had been accepted by the court as primary

84 WIGMORE, supra note 7, § 1917, at 101-02.
85 Id. at 102.
86 Id. at 103.
87 Id. at 105-06.
88 Id. at 103.
evidence. But if Mansfield’s decision was neither about inaugurating a new practice of calling experts as partisan witnesses before the jury, as Thayer maintained, nor about allowing experts to pronounce opinion without being personally familiar with the facts of the case, as Wigmore maintained, then what was it about? What was the issue that Lord Mansfield tried to winnow from the chaff of the protracted litigation?

Having reconstructed the facts of the case, we are now able to propose an answer to this question. Lord Mansfield, just like Chief Justice Gould before him, was trying to decide the merits of Hardinge’s objection, which pitted men of science in the old sense (that is, men of large and tested experience) against men of science in the new sense (that is, Newtonian philosophers), who based their opinions on their privileged knowledge of the imponderable laws of nature. Mansfield’s decision delivered therefore the authoritative interpretation of the King’s Bench concerning the implications of the nascent opinion doctrine for experts like Smeaton and Mylne, proto-scientists who functioned like skilled professionals but cogitated like natural philosophers.

The Wells Harbor litigation therefore did constitute an important historical moment in the deployment of expert knowledge in the courtroom, but for reasons different from those so far suggested. It was a junction in which the expanding late eighteenth-century cultures of law and science finally crossed paths. The lawyers had been solidifying their control over the production and presentation of evidence in the legal courtroom. Meanwhile, natural philosophy had become a competent branch of applicable knowledge. A noted Newtonian, Smeaton presented a conspicuous example of the growing importance of natural philosophy as a useful pursuit. For example, his experimental studies of waterwheels, the most important source of energy during the early stages of the industrial revolution, revolutionized their design and improved their performance by over thirty percent.89 Thus, although late eighteenth-century Englishmen may have still considered the opinions of natural philosophers to be less than facts, they nevertheless were already more than mere opinions. That was also the case concerning the causes behind matters such as the

decay of harbors, which had been traditionally a matter for the experience of the craftsmen and artisans who had built them. By 1782, as Mansfield made certain to clarify in his decision, these causes were already a “matter of science [about which] such men as Mr. Smeaton alone can judge.”

Wigmore was therefore right when he considered Mansfield’s decision as illustrating the growing legal recognition by the end of the eighteenth century that there was a new class of witnesses, skilled in matters of science, who could give opinions that were not based directly on the traditional trustworthiness of the senses. However, this lack of positive first-hand evidence was not merely a contingent deficiency occasioned by the experts’ failure to personally inspect the facts of the case. Rather, it was an inevitable consequence of the knowledge these new experts brought to the courts, knowledge that often claimed to be based on the imponderables of nature, which “nothing but the most philosophic eye, by reasoning upon chain of facts, is able to discover.”

Hardinge’s objection to Smeaton’s testimony forced the chief justices of the two central royal courts to reflect on the epistemological status of this new style of scientific reasoning and on the status of its bearers in the courtroom. Gould, the conservative, chose to remain within the guarded line delineated by the evolving rules of evidence and excluded Smeaton’s theory for not being clearly reducible to hard and concrete evidence. This formalist approach, which denied the court the services of the most respected expert on the issue upon which the whole litigation turned, made no sense to Mansfield. “I cannot believe,” he remarked, “that when the question is, whether a defect arises from a natural or an artificial cause, the opinions of men of science are not to be received . . . .” Mansfield declared, therefore, that the opinions of men of science were an exception to the opinion doctrine. Unwilling to distinguish one science from the other, Mansfield measured professional reputation instead. If the proposed witness was known as an expert on the matter before the court, Mansfield prescribed, his opinion was proper evidence.

---

91 2 JAMES HUTTON, THEORY OF THE EARTH 90 (1795).
Smeaton’s appearance as a partisan witness for hire played no part in Mansfield’s decision. This disregard was also typical of other leading rulings from the late-eighteenth and early-nineteenth centuries. One is pressed to find in them judicial angst about the growing practice of calling experts as partisan witnesses. One may wonder then: Could it be that the experienced royal judges overlooked the difficulties that might await the deployment of partisan expertise in the new adversarial courtroom? Why did they not try to mold a procedure that would keep expert advice out of the adversarial fire?

The answer, I would like to suggest, is that late eighteenth-century judges counted upon men of science to give, by ties of honor, unbiased opinions on matters beyond the ken of the jurors. The scientific community had long adopted the gentlemanly code of honor as a necessary condition for the reliability of the scientific discourse. Gentlemen were bound to credit the word of their fellows. The status of the gentleman—his economic independence, the freedom of his actions, the moral discipline he imposed upon himself—guaranteed the credibility of his word. This social contract worked both ways. Nothing ruined gentlemanly status quicker than dishonesty. John Locke gave notice in his 1690 guide to the education of English gentlemen:

[T]wenty faults are sooner to be forgiven, than the straining of truth, to cover any one by an excuse. . . . [Lying was] a quality so wholly inconsistent with the name and character of a gentleman, that nobody of any credit can bear the imputation of a lye; a mark that is judged the utmost disgrace, which debases a man to the lowest degree of a shameful manners, and ranks him with the most contemptible part of mankind, and the abhorred rascality, and it is not to endure by anyone, who would converse with people of condition, or have any esteem or reputation in the world.

It was noted that among the seventeenth-century members of the Royal Society, “the far greater Number are Gentlemen, free and unconfine’d.” During the eighteenth century, the Royal

---


95 John Locke, Some Thoughts Concerning Education §§ 131, 139 (1692) (paragraph break omitted) (quoted in Steven Shapin, A Social History of Truth: Civility and Science in Seventeenth-Century England 74 (1995)).

96 Thomas Sprat, The History of the Royal Society of London for the Improving of Natural Knowledge 67 (J. Knapton et al. eds., 3d ed. 1722).
Society continued to strengthen its status as a body of disinterested gentlemen who impartially investigated nature and toil for no end but the improvement of public good. The royal judges therefore were not worried about the behavior of these new experts and trusted that their testimony would correspond to their true opinions.

In retrospect, one can only appreciate the irony in this late eighteenth-century judicial leniency towards the new partisan role men of science took as witnesses in the modern adversarial courtroom. This leniency seems to carry the mark of the aloofness of the eighteenth-century judiciary, who dominated the courtroom to such an extent that they could not imagine it otherwise—that a time may come when their judicial powers would no longer suffice to control the play of partisan expertise in the courtroom. They were soon proven wrong. The tremendous nineteenth-century expansion of science and technology into industry and other public sectors quickly established the scientific expert as a pivotal figure in the courtroom and turned partisan expert testimony into an acrimonious and persistent thorn in the side of the common law. We will attend to these developments in the next two sections.

III. THE DEVELOPMENT OF SCIENTIFIC EXPERT TESTIMONY IN ENGLAND DURING THE NINETEENTH CENTURY

The 1782 contest of expertise over the implications of the tide for Wells Harbor and Lord Mansfield's subsequent decision to allow opinion testimony given by men of science to lay anchor in the courtroom were signs of the time—the time of the Industrial Revolution and the rising tide it brought of legal cases involving technological and scientific argumentation. Thus, during the early nineteenth century, among the crowd of experts who were allowed into the witness stand, besides the traditional figures of the physician, the merchant, and the sea captain, we find the growing presence of men of science—chemists, microscopists, geologists, engineers, mechanists, etc.

97 See generally Miller, supra note 25.

98 The circumstances of Folkes v. Chadd provided a good example of this judicial trust. Mylne and Smeaton were elite members of the Royal Society. Lord Mansfield had previously benefited from their expert services and respected them enormously. Clearly, he was not worried that they would testify dishonorably. They were men of honor and their integrity guaranteed the truthfulness of their stories. See GOLAN, supra note 38, at 51; Notes Ordering 2nd Trial, supra note 45, at 67.
These experts untangled for the court and the jury the complexities of the growing number of cases involving science, in fields ranging from insurance, mining, and energy to toxicology, patents, and even regulation. They appraised the disputed claims with their experimental techniques and offered their knowledge of the principles of nature, which the jurors then could apply to the facts at issue before them.  

The growing judicial recognition of this special class of witnesses was not all good news for this rising species of experts. It may have underlined the experts' growing importance to the judicial process, but at the same time it perpetuated their marginalization within this process. Moving across professional and institutional boundaries, from the exclusivity of their lecture theaters, private laboratories, and societies to the public courtroom, men of science hoped to represent laws that were not controlled by human whim. Instead, they found themselves manipulated as mere tools in the hands of the lawyers. As members of the jury or as advisors to the court, they were independent and active participants in the legal decision-making process. As witnesses, they found themselves isolated in the witness box, away from the decision-making processes. Browbeaten and set against each other, they found their standard strategies for generating credibility and agreement unfitted for the adversarial heat of the courtroom. The result was a continuous parade of leading men of science zealously contradicting each other from the witness stand—a parade that started to cast serious doubts on their integrity and on their science in the eyes of the legal profession and the public.

A. Severn, King and Co. v. Imperial Insurance Co.

All this was clearly on display during a large insurance litigation in the early nineteenth century. The case was a fire damage claim made in 1820 by the owners of a large sugar factory in London against four insurance companies. Shortly before the fire, the factory introduced a new process that pumped hot whale oil through copper coils immersed in the

99 For detailed analyses of many of these cases, see GOLAN, supra note 38, at 52-106.
100 The ensuing description of this case is drawn largely from June Z. Fullmer, Technology, Chemistry, and the Law in Early Nineteenth-Century England, 21 TECH. & CULTURE 1 (1980), and from the London Times; see also GOLAN, supra note 38, at 54-70.
sugar pans to boil the sugar solution. The insurance companies argued that the new process, the use of which had not been reported to them, introduced an increased risk of fire that voided the terms of their policies. The ensuing litigation revolved therefore around the little known characteristics of whale oil and its behavior under frequent application of intense heat.

Both sides of the litigation built their strategies upon scientific expert evidence. Virtually every notable chemist and chemical technologist in the kingdom was recruited by one of the parties. The scientific team for the sugar factory included the likes of John Dalton, President of the Manchester Literary and Philosophical Society and the author of the first useful atomic theory of matter; William Brande, Secretary of the Royal Society and Sir Humphry Davy’s successor as professor of chemistry at the Royal Institution; Thomas Thomson, a professor of chemistry at the University of Glasgow, editor of *Annals of Philosophy*, and the author of the influential *System of Chemistry*; as well as many other top men of science. One by one, the members of this dream team stepped onto the witness stand and swore for the plaintiffs that the new process was infinitely less dangerous than the old process, in which the sugar solution was heated over open fire.¹⁰¹

The insurance companies put together a scientific team no less formidable. It included the likes of Michael Faraday, chemical operator at the Royal Institution; Arthur Aikin, Secretary to the Society for the Encouragement of the Arts and the author of the *Dictionary of Chemistry*; Richard Phillips, professor of chemistry at the Royal Military College and chairman of the London Chemical Society, and many more. These experts testified under oath that the new heating process was extremely dangerous because the repeated heating and cooling altered the nature of the whale oil, making it increasingly volatile and producing highly flammable gases apt to explode.¹⁰²

Both teams of experts made sure to back up their opinions with results from experiments, many of them performed especially for the trial. Alas, the experimental results presented by the two sides were completely contradictory. While the plaintiff’s experts described smooth

¹⁰¹ See *GOLAN*, *supra* note 38, at 56-59.
¹⁰² See *id.* at 59-62.
operation under all conditions, the defense experts described
gusts of fire, combustive vapors, and sudden explosions.
Charging the jury, the presiding judge, Chief Justice Lord
Dallas, could not hide his utter frustration:

[T]hey [the jurors] had heard the evidence, he [Dallas] would not say
of the most intelligent, but of as intelligent men in chymical and
scientific pursuits as were to be found in this country or in Europe.
He had himself read the works of some of them, had derived
pleasure from their labours, and entertained the greatest respect for
their talents and information. But they had, nevertheless, left the
Court in a state of utter uncertainty; and the two days during which
the results of their experiments had been brought into comparison,
were days, not of triumph, but of humiliation to science.103

Dallas advised the jury to throw “the contradictory results of
experiment” out the window, and stated his disgust from the
partisanship that had been displayed during the trial.104 “It
must be a matter of general regret,” he said, “to find the
respectable witnesses to whom he was alluding drawn up, not
on one side, and for the maintenance of the same truths, but,
as it were, in martial and hostile array against each other.”105

The jury found a full verdict for the factory owners and
the Imperial Insurance Company moved for a new trial on the
ground that the verdict went against the weight of the
evidence.106 Discussing the appeal, the judges of the Court of
Common Pleas approved the request but decided to delay the
new trial “till one of the other causes [that is, the suits against
the other three insurance companies] should also have been
tried, and the result of certain proposed experiments affecting
the point in dispute be made known.”107 The judges, it seems,
were hoping that further investigation would allow the
chemists to clarify their evidence and offer the jury a better
basis on which to draw an informed conclusion in this
important litigation, which involved not only large sums of
money but also the general practice and principles by which
fire insurance was regulated.

The stakes were therefore raised for the next trial,
against the Phoenix Insurance Company. All parties were well
aware that the results would affect not only the renewed trial

103 Court of Common Pleas, TIMES (London), Apr. 14, 1820, at 3.
104 Id.
105 Id.
106 Fullmer, supra note 100, at 18.
of the first case, but also two other suits still waiting in the wings. Both parties recruited therefore even larger crowds of scientific experts and doubled their efforts to produce convincing experimental data. Alas, the experts and their evidence remained as contradictory as they were in the first trial. Chief Justice Lord Dallas, again, did not hide his distress in his charge to the jury:

A vast body of evidence had been laid before the jury; medical men, chymical men, eminent men in every department of science, had been examined in the course of the trial; but what was the lamentable result? The jury had heard of opinion opposed to opinion, judgment to judgment, theory to theory, and what was still more extraordinary, they had seen the same experiments producing opposite results. Who should decide this mighty controversy? He [Dallas] professed himself unable to give an opinion. He was not unacquainted with scientific subjects, but the little he knew only convinced him how much was beyond the reach of his knowledge . . . . This he would say of science in its present state, that all that belonged to the theory was doubtful, and that all that rested on experiment was new."108

The jury, again, gave full verdict for the plaintiffs, and the insurance companies were directed to honor their policies and compensate the factory owners for their trial costs. However, the insurance company refused to pay the cost of the plaintiffs’ experts. Legal precedent entitled only professional men to be compensated for their time and efforts, and the insurance companies argued that the plaintiffs’ experts did not fall within this category, but belonged instead to the legal category of “men of skill,” which included artisan-like experts (mechanics, navigators, etc.), who were not entitled to be compensated for their troubles and time.109

A third legal round took place, therefore, which revolved around the professional status of the scientific experts involved in the trial. The standard was set by the two established professions—medicine and law—upper-class oriented vocations defined by highly formal and specialized trainings. Alas, the welter of experts produced in the trial professed none of these traits. Socially they were faceless; intellectually they had just been proved incoherent; and their expertise was not based on any regulated training but rather self-taught. Thus, it was not without a gleeful undertone of score settling that Chief Justice

Dallas refused to recognize them as professionals. It took chemistry another sixty-five years to reverse this perception.\textsuperscript{110}

B. Gillespie v. Russell

The courts may have been unwilling for most of the nineteenth century to recognize men of science as professionals and compensate them accordingly, but there were plenty of others who were ready to pay them for their services in court.\textsuperscript{111} Thus, the presence of scientific experts in the courtroom continued to grow, and with it the embarrassing display of disarrayed definitions, inconsistent experimental results, and contradictory opinions. Such cases disturbed the courts deeply.\textsuperscript{112} The royal judges were not naive. They were ready to tolerate, to a certain degree, the difference of opinion among the scientific experts. Whether a fever that raged in a certain neighborhood was or was not caused by the fumes of a factory in the vicinity was a matter which admitted no demonstration and was considered therefore to be a speculation admitted as evidence only out of a necessity because no one else was qualified to give a better opinion. Experimental evidence was a different story, however. It was considered to be among the surest species of evidence, and the judges found it exceedingly difficult to accept the fact that similar experimental procedures were constantly producing antithetical results when conducted by opposed experts. Such conflicting experimental results, they believed, represented the partisanship of the men of science who produced them, and since these men were highly paid for their services, their conduct was seen as the prostitution of their science, of selling its credibility to the highest bidder.

Thus, as the century advanced and the legal use of scientific expertise grew exponentially, the court began to develop a skeptical view not only toward the opinions of the scientific experts but also toward their data—not because nature could lie, but because its representatives could. Indeed, by mid-century the judiciary no longer expected further scientific investigations to clarify the issues before the court.

\textsuperscript{110} Fullmer, supra note 100 at 24-25; COLIN A. RUSSEL, EDWARD FRANKLAND: CHEMISTRY, CONTROVERSY AND CONSPIRACY IN VICTORIAN ENGLAND 445-57 (1996).

\textsuperscript{111} See GOLAN, supra note 38, at 81.

\textsuperscript{112} For examples from numerous cases, see GOLAN, supra note 38, at 52-106.
Gillespie v. Russell, heard in 1853 in Edinburgh’s Court of Session, provides an example of this change in legal mood.113

In 1850, William Gillespie, a Scottish landowner, sold James Russell and Co., a firm of iron masters, the rights to mine coal from his land for the fixed royalty of 13 shillings and 6 pence per ton.114 Later that year, James Young, an Industrial chemist from Manchester, patented a revolutionary process that allowed for the first time the profitable distillation of paraffin oil from coal by low-temperature pyrolysis.115 Finding that a certain mineral of the Scottish lowlands, known as Boghead coal, yielded particularly high quantities of paraffin oil, Young established a large oil production facility in Bathgate, Scotland, next to Gillespie’s estate, where the Boghead coal was mined. Realizing in retrospect the unexpected value of the Boghead coal, Gillespie attempted to exclude it from the lease on the grounds that it was not truly coal but shale.116 After a set of inconclusive negotiations, Gillespie sued Russell and Co. for mining a mineral not included in their lease.117

The case constituted a direct attack on Young’s profitable patent. If the famous Boghead coal was found not to be a coal, anyone could use it in Young’s new process without infringing upon Young’s patent. The financial stakes were huge and both parties assembled unprecedented crowds of mining experts, geologists, chemists, and microscopists to debate the true identity of said mineral. Alas, although the chemists agreed that the great element in coal was carbon, they disagreed on how to measure it. All experts concurred that since that coal was of organic origin, the classification of the mineral in question could be reduced to the microscopic examination of its organic content. But while one set of microscopists swore that they saw vegetable tissue in the substance, the other set was equally sure that there was none. Lord President of the Court of Session, Duncan McNeil, was exasperated: “I do not care what you call it. I do not care about theories of the formation of coal—I do not care about what

114 For more information surrounding this case, see Golan, supra note 38, at 89-92.
115 See James Young, Improvements in the Treatment of Certain Bituminous Mineral Substances, and in Obtaining Products Therefrom, 9 Chemical News 249-50 (1864).
116 Gillespie, 3 S.C. at 1-2.
117 Id.
chemists choose to call it.” 118 Whatever it was, he reasoned, Gillespie agreed to lease it. Thus, the verdict was given for the defendant. 119

Gillespie moved for a new trial on the grounds that the scientific evidence was so contradictory that a new trial was essential to the justice of the case, and that a new experimental investigation be ordered to clarify the science involved. 120 Recall the similar request made thirty years earlier, in the 1820 insurance case of Severn, King and Co. v. Imperial Insurance Co. 121 There, the Court of Common Pleas not only ordered a new trial, but also decided to postpone it until further experimentation would be able to reduce the discrepancies among the scientific witnesses. By the 1850s, however, the Court of Session no longer entertained such high hopes. “[T]hey [the scientists] all agreed upon the theory, but they all disagreed on what they look at with their own eyes. . . .,” mocked the Lord President, “My opinion is, that they would differ in the result to the end of time.” 122 Lord Rutherford, a second judge, iterated:

Are we sure that they will be ever agreed? Are we going to get better microscopes and better eyes? Shall this branch of science, not only new in its name, but in its scientific terms, become new in a much more remarkable feature—in the unanimity of its professors? I cannot expect that. I do not anticipate it. 123

The request for a new trial was flatly rejected. 124

C. The Disillusionment with Scientific Expert Testimony

Ironically, the process of disillusionment the legal system was going through in its relations with science was not the result of some sort of philosophical relativism or skepticism. On the contrary, it was the result of hard-dying positivism that saw in the scientific method the best passage to truth. One of the most authoritative legal texts of the nineteenth century, written by Chief Justice and legal historian James Fitzjames Stephen, serves as an example of this deep

118 Id. at 13.
119 Id. at 12.
120 Id. at 1.
121 See supra Part III.A.
122 Gillespie, 3 S.C. at 1.
123 Id. at 15. For a related litigation in North America in 1851-1852, see 1 PAUL LUCIER, SCIENTISTS AND SWINDLERS: COAL, OIL, AND SCIENTIFIC CONSULTING IN AMERICAN INDUSTRIAL REVOLUTION, 1830-1870, at 69-121 (1994).
124 Gillespie, 3 S.C. at 15.
irony. According to Sir James, the quest for both legal truths and truths about the physical world “rest upon the same foundations . . . . the same great assumptions—the general uniformity of Nature, and the general trustworthiness of the senses.” 125 Still, scientific knowledge was much closer to certainty:

In physical inquiries the relevant facts are usually established by testimony open to no doubt, because they relate to simple facts which do not affect the passions, which are observed by trained observers, who are exposed to detection if they make mistakes, and who could not tell the effect of misrepresentation if they were disposed to be fraudulent.126

No wonder, then, that Stephen too interpreted the constant spectacle of leading scientists contradicting each other from the witness stand, not as legitimate debates but as a sign of moral decadence. “As to want of will to speak the truth, the causes of it are infinitely various,” and in this matter, he concluded, “The case of experts is as strong a one as can be mentioned. No one expects an expert, except in the rarest possible cases, to be quite candid. Most of them—are all but avowedly advocates, and speak for the side which calls them.”127

The growing scientific controversies in court have resulted therefore in widespread judicial and public indignation. In 1862, the conservative Saturday Review commented:

It is a fact that in all matters which require to be investigated through the evidence of expert witnesses, the same remarkable discrepancies show themselves. Hardly a single patent case is ever tried in which men of the highest scientific eminence do not appear to contradict one another flatly on the newness of the invention, or of some of its parts or stages, and the commonest disputes concerning architects’ and engineering bills are constantly calling forth similar conflicts of skilled testimony. Even in criminal cases, where the point to be decided is whether a particular poison was administered, or whether a death was caused in a particular way, the evidence of the experts is generally more contradictory than would be supposed from the nature of the inquiry; and, in short, judges and lawyers are rapidly coming to the conclusion that skilled testimony, which ought

126 Id. at 190.
127 Id. at 198-99.
to be the most decisive and convincing of them all, is of all the most suspicious and unsatisfactory.  

Yet the *Review* admitted, “To suppose that courts can do without such evidence would indeed be a stolid and ignorant prejudice, for expert witnesses can supply materials for judgments not to be obtained from any other source.”  

Still, “[t]here is no doubt that a system has been growing up of late years under which men of special knowledge are consulted under such circumstances as to render their opinions almost worthless.”  

The growing public mistrust of science, and even more so of the integrity of the scientists, in such an important domain as the legal system, deeply troubled the scientific community. The evil done by the scandals in court, scientific leaders emphasized again and again, was great not only to the administration of justice, but also to the public image of the Victorian scientific community, which was toiling hard on its professional status and seeking to expand its influence into the public domains of education, industry, health, administration, and culture in general. The reform of expert testimony therefore became one of the hottest topics in the meetings of the various scientific and legal societies, and many proposals were put forward to remedy the situation.  

Almost all scientific commentators agreed that the disagreements among the scientific witnesses did not reflect uncertainties within the body of scientific knowledge itself. Most of them argued that the disagreements were largely created by the improper adversarial procedures by which the legal system processed scientific knowledge. Others were ready to concede that scientific opinions may legitimately differ. But even they did not believe that the judge, let alone the lay jury, could reliably assess these differences. Each pundit had his own reform proposal, but they all seemed to agree on at least one of two central elements—that the courts should be allowed to call their own independent scientific witnesses, and to  

---

129 *Id.*
130 *Id.*
131 For a comprehensive discussion of the Victorian debates on expert testimony, see GOLAN, * supra* note 38, at 107-43; see also Christopher Hamlin, *Scientific Method and Expert Witnessing: Victorian Perspective on a Modern Problem*, 16 SOC. STUD. SCI. 485 (1986).
appoint, at least in civil cases, scientific assessors who would sit next to the judge and advise him on technical matters.133

However, even those in the legal profession who empathized with the frustrated scientific community pointed out that the reforms proposed by the scientific community ran against the fundamental postulates of the adversarial legal system.134 Getting rid of the jury ran against the fundamental political right to a trial by a jury of one’s peers. And allowing the court to call in assessors or witnesses independent of the parties ran against two other equally fundamental postulates—the right of the parties to control the evidence in court and the neutrality of the court. The reform suggestions were therefore rejected by the legal profession, who cautioned that “the remedy should not be worse than the disease.”135

The legal profession also resisted the scientific attempts to monopolize expertise by drawing a line between scientific and non-scientific men. “A man who acquired a particular kind of knowledge by long training,” insisted Thomas Webster, a leading authority on Patent Law, who played a central role in many attempts to reform Patent Law, “was just as much a scientific man in his particular art as the man who contributed to those wonderful discoveries of science at which we all so much rejoice.”136 Thus, it was wrong to speak of scientific men as a class to the exclusion of such skilled witnesses.

With no resolution in sight, the legal and scientific fraternities grew belligerent. The legal profession remained disturbed by the scientific partisanship displayed in the courtroom, while the scientific community remained frustrated by the awkward position it occupied in the courtroom. Still, the deadlock did not stop the increasing deployment of expert testimony in the courts. The growing scope and accuracy of scientific knowledge, as well as the constantly enlarging application of its principles to the business of everyday life, had inevitably expanded the uses of experts and tended to make the

133 Id. at 120; see, e.g., Committee of Scientific Evidence in Courts of Law, Report of the Committee, in 36TH MEETING OF THE BRITISH ASS’N FOR THE ADVANCEMENT OF SCIENCE 456, 456-57 (London, John Murray 1867).
134 See, e.g., William Odling, Science in the Courts of Law, 7 J. SOC’Y ARTS 167 (1860).
135 CHEMICAL NEWS, Oct. 11, 1862, at 190.
136 Robert Angus Smith, Science in the Courts of Law, J. SOC’Y ARTS, Jan. 20, 1860, at 144 (1860). Thomas Webster wrote the standard textbook on patent law for his day, THOMAS WEBSTER, REPORTS AND NOTES OF CASES ON LETTERS PATENT FOR INVENTIONS (London, Thomas Blenkarn 1844).
courts more and more dependent upon their advice. Still, the increasing tendency of lawyers to fortify their cases by the testimony of experts did not reflect appreciation for its excellence, but rather, the requirements of the rising culture of Victorian professionalism. The result was an ironic schism that was clearly emerging during the second half of the nineteenth century—the same increasingly indispensable expert opinions that were treated in everyday life as safe and reliable under the mere good faith of social and business reputation were considered unsafe and unreliable when given in court under oath.

IV. THE DEVELOPMENT OF SCIENTIFIC EXPERT TESTIMONY DURING THE NINETEENTH CENTURY IN THE UNITED STATES

The sale of expert advice did not become widespread in America until the middle decades of the nineteenth century. Once it did, though, the deployment of expert testimony in American courts of law grew quickly, and with it all its familiar woes. The American legal system observed the same adversarial procedures of the common law, while the American scientific community advertised the same high expectations from the scientific method, as did its English counterpart. These two features ensured that in spite of the significant differences in the institutional and social dynamics of the legal and scientific communities between the two countries, the problem of expert testimony would develop in nineteenth-century America along the same basic pattern displayed in England. Thus, as in England, the growing deployment of men of science in divergent areas of litigation turned the American courts into a lucrative arena for scientific activity.137

England, this arena soon put on public display the curious spectacle of leading scientists disagreeing with each other from the witness stand, a view that served to cast doubts on the integrity of the experts and their science.138

A. The Declining Credibility of Scientific Experts

By 1870, a study on expert testimony was already able to report in detail on an “unmistakable tendency on the part of eminent judges and jurists to attach less and less importance to testimony of this nature,” explaining this result by “the surprising facility with which scientific gentlemen will swear to the most opposite opinions upon matters falling within their domain.”139 Many shared this bleak view. “[W]hoever has read the reports of trials or been present at them, in which experts are seen arrayed against each other, prostituting at times the science which they professed to represent,” wrote U.S. Supreme Court Chief Justice Morrison Remick Waite in 1874, “need not be told, that the subject of expert testimony as now understood, is one of no ordinary importance.”140

Like their English colleagues, American men of science were much concerned with the damage that the scandals in court were inflicting on the public image and credibility of their emerging community, and, like their English colleagues, they were bitter about the adversarial legal machinery that placed them in the awkward position of partisan witnesses.141 “[N]o class connected with the administration of justice is more frequently misunderstood, or abused,” complained Charles Himes, a professor of physics and chemistry at Dickinson University.142 The improper position of science in court seemed to have turned the scientific witness into a legal annoyance:


139 Expert Testimony, 5 Am. L. Rev. 227, 228 (1871).

140 Morrison R. Waite, Testimony of Experts, 8 W. Jurist 129, 134-35 (1874).


[A] sort of intractable, incompatible, inharmonious factor, disturbing the otherwise smooth current of legal procedure; too important or necessary to be ruled out, too intelligent and disciplined mentally to yield without reason to ordinary rules and regulations of the court . . . and at the same time possessing an undoubted influence with a jury that it is difficult to restrict by the established rules and maxims of legal procedures.  

Both the English and the American legal systems were well aware of the need to protect the credulous jury from charlatans. Still, neither system was able to lay down a precise rule for determining who was and who was not a competent expert. The only legal criterion was 800 years old: those persons are qualified to speak as experts who possess special training and experience in the subject in question. Everything beyond this point remained purely a matter of discretion with the presiding judge, who had to decide in each case afresh whether the particular person offered as an expert witness would be admitted or not. In most cases, it was very hard for the judge to satisfy himself as to the qualifications of the persons offered as experts. Scientific titles and diplomas, and professional reputation, carried little judicial meaning during the nineteenth century, and preliminary examinations were impossible to make. The judges therefore were continually forced to decide on the spur of the moment, and often on the most difficult subjects, upon the credibility of the persons offered as experts. Unable to discriminate with any reasonable degree of accuracy between experts and charlatans, the actual practice of the courts came to be to admit almost everybody presented as experts, leaving it for cross-examination to expose quackery, and for the jury to be the judge of the ensuing battles between expert witnesses and lawyers.  

No one, of course, trusted the jury to be able to do this job properly. As one distinguished psychologist of the day pointed out:

> [T]he average jury is unable to pass judgment on, or even to comprehend in any adequate way, many of the questions submitted to it—such as motives and capacity of the mind, the power of control, the analysis of conduct, and the conditions and influences which have been dominant in certain acts; the application of the law, and

143  *Id.* at 411-12.

the distinctions of responsibility and accountability; the distinctions of science as to the meaning of certain facts, or the recognition and discrimination of facts from the mass of statements.\textsuperscript{145}

Here, however, came into play two critical differences between the English and the American situations. While the English legal system recognized the jury as the final adjudicator on the facts of the case, it nevertheless granted its judges the freedom to take part in the questioning of the witnesses, advise the counsels in the framing of their questions, and comment fully on the weight of the evidence and the credibility of the witnesses in their charge to the jury. The authoritative royal judges did not hesitate to use these legal instruments to control the usage of expert testimony in their court and to guide the jury in its assessment of the scientific witnesses and their evidence. In addition, from the mid 1860s onward, the English courts began to divert cases that involved scientific expertise from jury trials to the juryless Chancery Court. At first it was mainly patent trials, but in 1875 a parliamentary bill was passed that officially granted trial judges unfettered discretion in all civil actions to order a trial without a jury in any matter requiring scientific evidence that, in their opinion, could not be handled by the jury.\textsuperscript{146}

The American court lacked access to these two instruments. The eighteenth-century English notion of the institution of the jury as a mainstay of liberty was adopted with added zeal in the American colonies, and the fact that many judges were laymen with no special claim to legal competence only added to the prominence of the jury. The Jacksonian faith in the ability of the common man and the enduring political philosophy that supported maximizing citizen participation in government kept this enthusiasm alive throughout most of the nineteenth century to an extent unfamiliar in England.\textsuperscript{147} Consequently, the early nineteenth-century American juries did pretty much as they pleased. The

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{145} T.D. Crothers, \textit{A Psychological Study of Jurors}, 60 ALB. L.J. 341, 342 (1899).
\item \textsuperscript{147} The Seventh Amendment to the American Constitution provides as follows: “In Suits at common law, where the value in controversy shall exceed twenty dollars, the right of trial by a jury shall be preserved, and no fact tried by jury, shall be otherwise re-examined in any Court of the United States, than according to the rules of the common law.” U.S. CONST. amend. VII.
\end{itemize}
\end{footnotesize}
second half of the century saw a growing pressure by the bar and the business community to instill more predictability and rationality in the operation of the jury, but the pressure also bred popular fears of undue influence. The result was a practical compromise that was attained by a sharpening of the law/fact dichotomy and the corresponding spheres of judge and jury. On the one hand, the power of the jury to determine the law, especially in civil cases, was eroded. On the other hand, fears of undue influence on the jury were eased by legislative and constitutional restrictions on the power of the American courts in charging juries. By 1889, in twenty-one out of the forty-nine U.S. states and territories, judges were expressly forbidden by statute or constitutional provision to charge the jury on questions of fact. And in about half of the remaining twenty-eight states and territories, the courts had voluntarily adopted the same restriction. Only in federal courts and a minority of state courts were judges allowed to comment on the weight of the evidence in their charge to the jury.

American men of science decried the absence of a judicial hand that would guide the jury in its difficult task of assessing the scientific evidence of the case. “If it be necessary to give juries authoritative instruction on points of law, how can it be less necessary that they should be similarly instructed in matters involving scientific knowledge?” wondered *Scientific American* in 1872:

Is it any wonder that the public is beginning to mistrust the value of this kind of [scientific] evidence? Such mistrust is based upon good grounds enough. As now presented to juries, the testimony of the both competent and incompetent witnesses, only serves to muddle their intellects, and to complicate rather than make plain the facts.

---


150 *Id.* at 308.

151 *Id.* at 309.


153 *Id.*
The nineteenth-century American scientific community lacked the organization, status, and political resources needed to challenge the legal system and its procedures. Most attempts to reform the legal procedures of expert testimony were initiated by members of the medical and the legal professions. The reform of expert testimony became one of the hottest topics in the meetings of the various bar associations that mushroomed in late nineteenth-century America, and many bills were drafted to remedy the evils of expert testimony. For the selection of experts it was suggested that they be chosen by the court, either reserving or denying the right of the parties to call additional witnesses; that the selection of the courts be unassisted or made from an official list chosen in some other manner; that the official list be either permanent or special for each case. In regard to the examination of witnesses, it was recommended that the examination be by the court, with or without the right of the parties to cross-examine, or that there be no examination at all and that the expert would submit a report. In regard to decisions when experts disagree, it was recommended that a jury of experts be selected or that an expert sit with the judge during the trial to advise him. Alas, the American legislature and judiciary seemed even more reluctant than their English counterparts to dissent from the legal axioms of the adversary system. Most reforms bills did not pass the legislative stage, and the few that did were promptly held unconstitutional.


155 One of the first reform bills was a joint effort of the American Academy of Arts and Sciences, the Suffolk District Medical Society, the Boston Society for Medical Observation, and the Boston Society for medical Sciences. The bill was written by Judge Emory Washburn who headed the committee of the Academy of Arts and Sciences. See Editorial, 90 Bos. Med. & Sur. J. 387-88 (Apr. 16, 1874). For a wider look on the efforts of the medical community, see James C. Mohr, Doctors and the Law: Medical Jurisprudence in Nineteenth-Century America 95, 100-05 (1993).

156 For a long list of related commentary, see “Evidence, Expert” in Index to State Bar Association Reports and Proceedings 176-77 (Dennis A. Dooley ed., 1942).


158 See id. at 224.


B. The Failure of the Law of Evidence

Unable to check either the selection of the experts or to guide the jury’s assessment of their evidence, nineteenth-century American courts concentrated their efforts on the law of evidence, in an attempt to check the growing problem by regulating the processes through which the experts communicated their information in court.

One major legal doctrine sought to protect the credulous jury from being uncritically influenced by the expert’s view by preventing the expert from giving his opinion upon the “ultimate issue,” that is, the precise factual issue before the jury.161 To permit that, it was held, would put the expert in place of the jury and invade their province.162 Rational as it may sound, the application of this doctrine created great confusion and led to absurd consequences. In theory, it made irrelevancy a ground for admission, and relevancy for exclusion. In practice, the “ultimate issue” was often exactly what the expert testimony was all about. The doctrine seemed, therefore, to exclude expert evidence exactly where it was most needed. Consequently, the courts developed various ways to bypass the rule and allow the witnesses to give their opinion on the ultimate issue.163

The most popular procedure was to allow an expert to state in general terms whether a certain cause may have produced the result under consideration, and leave it to the jury to decide whether it did produce it or not. To enable this, a second evidentiary doctrine came into play. Under the “hypothetical question” doctrine, the expert’s testimony was given in the form of answers to hypothetically framed questions.164 These questions specified a set of factual premises, already submitted in evidence, and the expert was asked to draw his conclusion from them, assuming that they were true. This cumbersome technique was justified on triple grounds: to enable experts to apply their general knowledge to facts that were not within their personal knowledge; to allow juries to

---

note 2, at 55, 56-58; Note, Appointment of Expert Witnesses by the Court, 24 Harv. L. Rev. 483, 483-84 (1911).


163 McCormick, Handbook supra note 161, at 120.

164 McCormick, supra note 161, at 820-25
hear the factual premises upon which expert opinions were based; and to allow experts to give their opinions on ultimate issues without “invading” the province of the jury. Juries were then instructed to credit the opinions only if they believed the underlying premises.165

Sound in theory, the technique broke down in practice. If counsel was required to recite all the relevant facts, the question became intolerably lengthy; if allowed to select the facts, as most courts did, it prompted one-sided hypotheses. Designed and controlled by the interested parties, the hypothetical question became a means to manipulate the facts of the case rather than to clarify them for the jury. “It was a strange irony,” Wigmore noted in 1904, that “the hypothetical question, which is one of the few truly scientific features of the rules of Evidence, should have become that feature which does most to disgust men of science with the law of Evidence.”166

Even the old and powerful hearsay doctrine turned out to be problematic in the context of expert testimony. The caution of the courts in admitting opinions not based on observation of the particular facts of the case, and the fear of misleading the jurors by reading to them scientific statements they were hardly competent to assess, had led many courts to exclude what many considered the most natural source of scientific information—standard textbooks, reports, etc. The exclusion of these written documents was justified by the hearsay doctrine on the premise that they were statements not made under oath or that their author was not available for cross-examination. As with other doctrines, the courts slowly devised ways to work around this one, too. Some courts permitted the use of scientific treatises, but only to discredit an expert. Others allowed experts to “refresh their memory” by reading from standard works. Others even allowed publications of exact science, assuming their statements to be of ascertained facts rather than of opinion, and excluded other treatises, especially medical works.167 Confusion and inconsistency, again, were rampant.

167 See John Henry Wigmore, Scientific Books in Evidence, 26 Am. L. Rev. 390, 390-403 (1892); Warren M. Dana, Admission of Learned Treatises in Evidence, Wis. L.
By the end of the nineteenth century, it was clear that the American law of evidence had failed to control the problem of expert testimony. Designed to be the crown of modern American jurisprudence, a corpus of legal procedures as rational as Euclidean geometry, the law of evidence had turned instead into a highly complicated and technical domain, sagging to the point of collapse under the burden of its own distinctions, exceptions, and exclusionary duties. In 1898, James Thayer called it a “piece of illogical, but by no means irrational, patchwork; not at all to be admired, nor easily to be found intelligible,” and by 1904, his famous pupil, John Henry Wigmore, needed four thick volumes to cover it. “There is a full realization now,” concluded the Chicago Legal News in 1909, after a long historical review of expert testimony, “that in the present practice we have carried a branch of procedure out to the utter defeat of its object, to an absurdity; and that the result has been a wide-spread disgust with methods of legal administration.”

Thus, although first raising its head in the English courts, it was in America that the problem of expert testimony reached its fullest expression. The diversion in England of technical litigation away from jury trials and the efforts of the authoritative royal judges seemed to have kept the thriving business of expert testimony under relative check. The last decades of the nineteenth century saw the bitter English debates concerning the problems of expert testimony subsiding. Across the Atlantic, however, things went from bad to worse. Unable to check either the selection of the experts or their evidence in court, or the jury’s assessment of this evidence, late nineteenth-century American courts saw the problem of scientific expert testimony reach a crisis.
The grim legal mood was captured by Judge Gustav Endlich in his 1896 address on expert testimony before the prestigious Law Academy of Philadelphia:

Indeed, it is difficult to conceive of language within the bounds of decent and temperate criticism, which ought to be regarded as excessively severe in commenting upon the expert testimony nuisance as it has, of late years, been infesting our courts. In the way of wasting the public time, in the way of burdening litigants with expense, and in the way of beclouding the real issues to be tried and effecting miscarriages of justice, it has grown to the proportions of an offensive scandal. Instead of being an aid in the administration of the law, it has become a positive hindrance to it. Instead of assisting in the approximation of the truth, it has become the means of obscuring it. . . . [E]xpert testimony is to-day discredited and rightly discredited by the courts, and ridiculed and rightly ridiculed by the hard common sense of the people.\(^{173}\)

This outlook was shared by many, in and out of the legal profession. Invited to speak on expert testimony before the New Hampshire Medical Society at its 1897 annual meeting, Judge William Foster chose to open his address with a joke popular within legal circles: “There are three kinds of liars,—the common liar, the [damned] liar, and the scientific expert.”\(^{174}\) Foster assures his scientific audience:

This characterization . . . is bestowed . . . not only by defeated lawyers and their enraged clients, but also by eminent members of the legal profession, both lawyers and judges, as well as by worthy and respectable members of the general public outside of the professions involved. It is the voice of the people and of the press, as well as that of the bench and the bar. It is the fashion.\(^{175}\)

V. DEVELOPMENTS DURING THE TWENTIETH CENTURY IN THE UNITED STATES

As Endlich’s and Foster’s choice of words made clear, the feeling by the end of the nineteenth century was that the problem of expert testimony had reached rock bottom.\(^{176}\) The

---

\(^{173}\) GUSTAV ENDLICH, EXPERT TESTIMONY: WHAT IS TO BE DONE WITH IT? 5, 12-13 (Phila. 1896).

\(^{174}\) See Foster, supra note 159, at 169.

\(^{175}\) Id.

\(^{176}\) Judicial complaints about scientific expert testimony could be compiled almost at will. For one of the more extensive compilations, see L.G. Kinne, Expert
debate over its causes and solution had been picking up steam for almost half a century, and the remedies suggested, as another judge put it, were “as numerous as prescriptions for the cure of rheumatism and generally about as useful.” 177 The disease was spreading fast, however, and the pressure for a remedy, any kind of a remedy, was mounting. Something eventually had to give way in the sacred triangulation of the adversarial system: either the political postulate of the lay jury, the traditional right of the parties to furnish all evidence, or the neutral position of the court.

In 1905, Michigan made the first attempt to accomplish such a reform by an act of the legislature. It passed a statute that embodied the most popular reform suggestion—that of allowing the court to nominate its own experts. The statute contained the mildest possible version of such a reform. It did not preclude the parties from using their own witnesses but provided in criminal cases for the additional appointment by the court of no more than three disinterested persons, whose identity should not be made public, to investigate issues involving expert knowledge and testify to their findings at the trial. Nevertheless, the Michigan Supreme Court held the statute unconstitutional. 178 The court considered it no part of the duties of the court to select witnesses. 179 Such activity, it pointed out, transferred the power of choosing witnesses from the prosecutor, an administrative officer, to a member of the judicial department, in violation of the provision of the state constitution for a separation of powers. 180 The court considered the statute a violation also of the fundamental right of the accused to a fair and impartial trial. The official sanction of judicial appointment, the court pointed out, would give the court experts an “extraordinary certificate of candor, ability, and truthfulness, while the other testimony in the case must be judged by the jury by ordinary standards.” 181 Declaring the legislation in question unconstitutional, the court expressed the opinion that the only available remedy for the acknowledged evils at which the statute aimed would have to

---

177 Kidd, supra note 159, at 217.
178 People v. Dickerson 129 N.W. 199, 199-201 (Mich. 1910).
179 Id. at 200-01.
180 Id. at 201.
181 Id.
be found in a “livelier sense of responsibility to the public for the proper and decent administration of justice.”

The decision dealt a serious blow to those who had been advocating reform of expert testimony by means of statutory enactment. Taking their cue from the Michigan Supreme Court, and accepting that the experts should remain party-chosen and the jury still be considered the final trier of facts, early twentieth-century American legal scholars concentrated their attention on improving the standards of admissibility for the party-chosen expert. Their renewed hopes of succeeding where their predecessors had so miserably failed hinged on a clear change in the market of scientific expertise, created by the rising professional culture in America. By the second decade of the twentieth century, the individual expert who developed and marketed his or her own expertise had already been replaced by a community of experts who shared, and were defined by, common standards of competence and ethics. A wide range of expertise, from the scientific and technological fields of the chemists, physicists, and engineers, to architects, surveyors, actuaries, realtors, insurers, and accountants, came to be dominated by professional associations of practitioners. These associations developed codes of ethics, standards of education, training and practice, and defined minimum qualifications of certification either through their own examinations or through those of the various state boards of examiners.

By the second decade of the twentieth century, one can find legal scholars pondering the ways in which the courts could take advantage of this standardized market of expertise to check the problem of expert testimony. “The remedy is not in the enactment of any new statute,” wrote one scholar in 1910. “No act of the legislature will make witnesses learned or honest. The reform must come from the professions themselves.” It was not clear exactly how such a reform should be carried out. “There is . . . the logical possibility that no remedy exists, or that any proposed remedy brings in its...

182 Id.; see also Note, supra note 160, at 483-84.
185 Friedman, supra note 183, at 252.
186 Id.
train new evils worse that the disease which it cures,” claimed another commentator in 1915.187 One thing was clear, though. “In the selection of experts no solution can be considered satisfactory that does not provide for the selection by the profession involved.”188 Eight years later, in 1923, the Court of Appeals for the District of Columbia came up with the first effective formulation of such a solution.189

A. The General Acceptance Standard

The decision was occasioned by the newly invented and highly publicized lie-detector technology. In 1922, a young African-American named James Frye was accused of murder in Washington, D.C.190 Frye pleaded not guilty, and his lawyer offered one of the inventors of the lie detector, William Marston, as an expert witness to testify to the results of a test performed on the defendant, which allegedly proved his truthfulness.191 The trial judge, Walter Irvin McCoy of the Supreme Court of the District of Columbia, refused to admit Marston and his sensational test in evidence.192 As far as he was concerned, McCoy proclaimed, such tests were inadmissible until “there is an infallible instrument for ascertaining whether a person is speaking the truth or not. . . ., [but] I shall be dead by that time, probably, and it will bother some other judge, not me.”193 Frye was found guilty, and his lawyer appealed on the ground that the scientific expert and his evidence were improperly excluded.194

The Court of Appeals for the District of Columbia likewise was not going to allow the sensational lie-detector test into the court. Still, being an appellate court, it needed to furnish a better rationale for its exclusion than McCoy’s peculiar infallibility standard. This was not an easy task, however. In 1923, there was no special rule for the admissibility of scientific evidence. Like every other type of evidence, scientific evidence was mainly evaluated according to

187 Kidd, supra note 159, at 218, 223.
188 Friedman, supra note 183, at 252; Kidd, supra note 159, at 218, 223.
189 Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).
190 See WILLIAM MOULTON MARSTON, LIE DETECTOR TEST 70-72 (1938).
191 See id. at 71-72.
193 Id. at 694.
194 Frye, 293 F. at 1014.
the traditional evidentiary criteria: the “logical relevancy” of the evidence and its helpfulness to the trier of fact, and the qualifications of the witness. \(^{195}\) Neither criterion offered the Court of Appeals for the District of Columbia much reason to exclude the lie detector. The logical relevance of the test and its potential helpfulness to the jury were obvious. So were the credentials of the test inventor, William Marston. He was a lawyer, a member of the Massachusetts Bar, and a well-published research psychologist, who possessed special training and extensive practical experience in the subject in question. \(^{196}\) Thus qualified, the weight of legal precedent of expert testimony was clearly for admissibility. \(^{197}\)

Unable to exclude Dr. Marston and his deception test on the basis of the existing admissibility rules, the Frye court put forward an innovative rationale that shifted the focus of the admissibility process from the expert’s credentials to the particular scientific knowledge he proposed to the court.

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs. We think that the systolic blood pressure deception test has not yet gained such standing scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made. The judgment is affirmed. \(^{198}\)

Proposing to look for a general acceptance in the particular field to which the expertise belonged, Frye offered a potent point of departure from the traditional deadlock of scientific expert testimony. The jury was still considered the final trier of facts, and the experts would still be party-chosen; but the

\(^{195}\) See McCormick, Handbook, supra note 162, at 489; James Bradley Thayer, A Preliminary Treatise on Evidence at the Common Law 266 (Boston, Little, Brown, and Co. 1898); Herman L. Trautman, Logical or Legal Relevance—A Conflict in Theory, 5 Vand. L. Rev. 385, 385, 392 (1952).


\(^{197}\) See Best, supra note 144, at §§ 513-515, at 867-88; Thayer, supra note 195, at 195-97; David L. Faigman et al., Check Your Crystal Ball at the Courthouse Door, Please: Exploring the Past, Understanding the Present and Worrying About the Future of Scientific Evidence, 15 Cardozo L. Rev. 1799, 1803-04 (1994).

\(^{198}\) Frye, 293 F. at 1014.
judicial ability to control the market of scientific expert testimony was significantly extended from the realm of the expert to the realm of expertise.\textsuperscript{199}

A century and a half earlier, in 1782, George Hardinge, barrister of the Middle Temple, suggested a similar rationale in \textit{Folkes v. Chadd}.\textsuperscript{200} John Smeaton was a worthy expert, Hardinge argued, but he propounded in court a theory whose scientific status was yet unsettled. Hence, his evidence should be excluded. Lord Mansfield rejected the proposed distinction between the expert and his expertise. Mansfield’s decision, maintaining that it was not for the court to qualify the expert’s opinion, shaped the nineteenth-century practice of expert testimony.\textsuperscript{201} If a person was qualified as an expert, his or her expert opinion came along; it was the job of the cross-examiner to expose the weaknesses of the testimony and for the jury to weigh it.\textsuperscript{202} A century and a half later, the \textit{Frye} court reintroduced Hardinge’s exclusionary logic into the rules of evidence; being an expert was no longer enough.\textsuperscript{203}

The \textit{Frye} opinion portrayed scientific knowledge as an evolutionary process that had to advance from the experimental to a demonstrable stage before it could be accepted in the court. This resonated admirably with the pragmatic vision of American professional culture that conceived of expert knowledge as a communal product that could be objectively evaluated, independently from the individual expert.\textsuperscript{204} In a similar fashion, the search for general acceptance within the relevant scientific community accorded well with the dominant progressive views of the age, which conceived of law as an organic part of greater society and emphasized the role of coordinated expertise in the joint attempt to run society

\textsuperscript{199} Wigmore offered a similar but less demanding formulation in a new section, entitled “Scientific Experimental Tests by Psychologists,” that he added to the 1923 edition of his famous treatise on evidence: “All that should be required as a condition is the preliminary testimony of a scientist that the proposed test is an accepted one in his profession and that it has a reasonable measure of precision in its indications.” \textsc{2 Wigmore, supra} note 7, § 990, at 419.

\textsuperscript{200} See supra text accompanying notes 63-67.

\textsuperscript{201} See \textsc{Thayer, supra} note 31, at 672-73; \textsc{4 Wigmore, supra} note 7, § 1917, at 102-04.

\textsuperscript{202} See \textsc{McCormick, Handbook, supra} note 162, at 363-64.

\textsuperscript{203} See \textsc{Frye, 293 F. at 1014}; \textsc{Black et al., Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge, 72 Tex. L. Rev. 715, 725-27 (1994)}.

\textsuperscript{204} \textsc{Arthur O. Lovejoy, The Thirteen Pragmatisms, 5 J. Phil. Psych. & Sci. Methods 1-12, 29-39 (1908)}; \textsc{H.S. Thayer, Meaning and Action: A Critical History of Pragmatism 50-66 (1968)}.
efficiently and uniformly.\textsuperscript{205} Still, originating in an extreme case and containing no precedential citations, the decision of the Court of Appeals for the District of Columbia remained at first an isolated solution to a particular problem. During the following three decades, the courts remained content with applying the requirement for general acceptance only to exorcize from criminal trials evidence derived from various lie-detection and truth serum schemes.\textsuperscript{206} It was only in the post-World War II years that the courts began to apply the general acceptance requirement as an exclusive test in a constantly broadening range of novel scientific evidence.\textsuperscript{207}

The growing adoption of \textit{Frye} signaled the onset of a new judicial trend towards ever-greater judicial scrutiny of scientific evidence. The expansion of this judicial role during the second part of the twentieth century can be described in terms of both the scope of the cases to which it has been applied and the depth of the judicial scrutiny of the scientific evidence. Thus, by the early 1950s, \textit{Frye} was already addressed in the legal literature as the leading criterion for the admissibility of novel types of scientific evidence in criminal trials.\textsuperscript{208} In the 1960s, the U.S. Supreme Court restricted the acquisition of evidence in criminal cases via traditional interrogation techniques.\textsuperscript{209} Perhaps in response, crime laboratories flooded the courts with innovative scientific technologies, and trial judges started to use \textit{Frye} as a ready-made tool to decide the reliability of evidence derived from new techniques such as voice prints, neutron activation analysis, gunshot residue tests, bite mark comparisons, scanning electron microscopic analysis,

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{207} After World War II, the case was only cited 6 times before 1950, 20 times during the 1950s, 21 times during the 1960s, 100 times during the 1970s, 470 times during the 1980s, and 350 times in the early 1990s. \textit{See} Black et al., \textit{supra} note 203, at 722 n.30.
\item \textsuperscript{208} \textit{See}, \textit{e.g.}, MCCORMICK, \textit{Handbook}, \textit{supra} note 162, at 363; Felver A. Rowell, Jr., \textit{Comment, Admissibility of Evidence Obtained by Scientific Devices and Analysis}, 6 ARK. L. REV. 181, 181 (1952).
\end{enumerate}
\end{footnotesize}
truth sera, and numerous others. By the 1970s, Frye’s general acceptance standard had become “not only the majority view . . . [but] the almost universal view” in the majority of criminal courts that considered the admissibility of new scientific evidence. By the end of the 1980s, the judiciary expanded its use from criminal to civil proceedings. Finally, by the 1990s, as the judiciary grew confident in their ability to measure the proffered expert evidence, the courts moved from deferring to the judgment of the scientific community to independently finding out things for themselves.

The expanding judicial dominion over scientific expert testimony met with increased criticism. The earliest critics considered judicial screening of the scientific evidence as an unnecessary procedure that deprived the jurors of their right to decide for themselves what facts are valuable. As the use of Frye multiplied, so did its critics—complaining that the general acceptance criterion was too vague, too narrow, and too slow. The courts responded by setting forth supporting rationales. Their main argument was that Frye finally provided the courts with a uniform “method for ensuring the reliability of the

---


212 Sterling v. Velsicol Chem. Corp., 855 F.2d 1188, 1208 (6th Cir. 1988). The Sixth Circuit did not cite Frye; nevertheless it excluded “clinical ecology” evidence for not being generally accepted in the relevant scientific community. The Fifth Circuit applied Frye soon after, in 1991, in Christophersen v. Allied-Signal Corp., 939 F.2d 1106, 1115-16 (5th Cir. 1991). Tellingly, both of these were toxic tort cases.


215 Professor McCormick criticized the general acceptance test and argued that it “is a proper condition upon the court’s taking judicial notice of scientific facts, but not a criterion for the admissibility of scientific evidence.” MCCORMICK, HANDBOOK, supra note 162, at 363.

scientific evidence.” 217 However, in practice, “the thing from which the deduction is made” has meant different things to different courts at different times. The ambiguities inherent in determining the particular field to which new scientific evidence belongs, and in deciding how to measure its “general acceptance,” left ample room for discretion. Consequently, Frye ended up having not one but many general acceptance criteria, which the courts seemed to apply in a selective manner, according to their own views about the reliability of the particular forensic technique before them. 218

Meanwhile, a new twist entered the plot. In 1975, the rules that federal judges must follow were finally codified. Completely disregarding Frye, the newly enacted Federal Rules of Evidence (“FRE”) prescribed no special test to ensure the reliability of scientific evidence, new or old. Instead, casting the widest net possible, the FRE provided:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of opinion or otherwise. 219

Having left open the question of how one defines “scientific, technical, or other specialized knowledge,” the FRE was generally interpreted as prescribing a more flexible judicial consideration of scientific evidence in order to create the opportunity for more types of scientific evidence to be used in court. 220 On the other hand, since the FRE did not state an explicit intent to abandon the Frye rule, some federal and almost all state courts remained committed to the “general

---

217 Giannelli, supra note 209, at 1207.
220 According to the U.S. Supreme Court, the general approach of the Federal Rules of Evidence was “of relaxing the traditional barriers to opinion testimony.” The Court referred to Frye’s as an “austere standard, absent from and incompatible with, the Federal Rules of Evidence.” Daubert, 509 U.S. at 589. According to Judge Weinstein, “The Rules were designed to depend primarily upon lawyer-adversaries and sensible triers of fact to evaluate conflicts.” Jack B. Weinstein, Rule 702 of the Federal Rules of Evidence Is Sound; It Should Not Be Amended, 138 F.R.D. 631, 631 (1991) (quoted in Daubert, 509 U.S. at 589).
acceptance” criterion as an absolute prerequisite to the admissibility of scientific evidence.221

The debate concerning the proper judicial attitude and standard for the admissibility of scientific evidence intensified during the late 1980s and early 1990s, heated by growing fears of a mass tort litigation explosion.222 Until the 1970s, judicial scrutiny of scientific evidence took place almost exclusively in criminal cases, where life and liberty are on the line.223 However, the 1970s saw the rise of tort law as a new major arena for scientific evidence. Dangerous drugs, industrial defects, environmental pollutants, and a host of other technological breakdowns have all become the subject of prolonged civil litigation with ever-escalating financial stakes.224 In the great majority of these cases, the central legal questions were of risk and causation, which invariably turned upon scientific evidence and put on display, again, the all-too-familiar sight of leading scientific experts producing from the witness stand conflicting data and contradictory conclusions.225

The customary complaints soon followed, and the warning was sounded that America’s courts were being swamped by “junk science,” produced by an unholy alliance between unscrupulous experts and opportunistic attorneys.226 The judges were urged to raise the bar and rely on the conservative interpretation of Frye in order to protect the credulous jury from pseudoscientific experts, and the deep-pocketed corporations from greedy lawyers.227 Others objected. The Frye test, they argued, sanctions a stifling and repressive scientific orthodoxy that could prevent the courts from learning

221 Giannelli, supra note 209, at 1228-31.
223 Bernstein, supra note 214, at 389-90.
225 For a plethora of examples, see Huber, supra note 222; Phantom Risk: Scientific Inference and the Law (Kenneth R. Foster et al., eds. 1993).
227 Huber, supra note 222, at 200-01.
of authentic scientific innovations and citizens from seeking justice.  

Hence, they urged the court to adopt the relaxed admissibility requirements of the FRE. In short, the two-centuries-old debate between the Goulds, who maintain that the law should exclude from the courtroom certain expert opinions for not being scientific enough, and the Mansfields, who maintain that the law has no means to give preference to one kind of science over another, was back with renewed vigor.

B. Daubert in Historical Perspective

The conflict came to a head in 1993, when a civil suit by two minors, Jason Daubert and Eric Schuller, and their parents, against the giant pharmaceutical corporation Merrell Dow was decided by the Supreme Court. Daubert and Schuller were born with serious birth defects and blamed them on Merrell Dow’s Bendectin, a popular anti-nausea drug their mothers took during their pregnancies. To prove a causal link between the drug and the birth defects, Daubert’s lawyers offered eight well-credentialed experts, who based their conclusion that Bendectin can cause birth defects on: animal studies that found links between Bendectin and malformation; chemical analysis which pointed to structural similarities between Bendectin and other substances known to cause birth defects; and re-analysis of previously published epidemiological data that found a link between the drug and birth defects. The federal judge was not impressed, finding that the animal studies and the chemical analysis insufficient to show causation. The meta-analysis of the epidemiological data, he further pointed out, was created especially for the trial and was neither published nor subjected to peer-review. Thus, it could not be considered to be generally accepted under Frye.

---


231 Id. at 573.

232 Id. at 575.

233 Id.
Consequently, the judge granted summary judgment to Merrell Dow and threw the case out before it could reach a jury.\textsuperscript{234} Daubert’s lawyers appealed all the way to the Supreme Court, arguing that the FRE supersedes \textit{Frye} and that according to the FRE, it is for a jury, not a judge, to determine the persuasiveness of scientific evidence. The Court agreed with the petitioners that \textit{Frye} was superseded by the FRE, but rejected their let-it-all-in interpretation of the FRE. Instead, the Supreme Court read the FRE as requiring that the trial judge ensure that any scientific evidence admitted into the courtroom be reliable. Addressing the main question left open by the FRE—how one recognizes valid scientific knowledge—the Supreme Court acknowledged that ready-made formulae, such as the “general acceptance” criterion, failed to provide the desired answer.\textsuperscript{235} Nevertheless, the Court emphasized that an answer must be provided, and if a general formula could not be furnished, then it was the responsibility of trial judges to make their own inquiries in each and every case in order to provide it.\textsuperscript{236} To that end, the Supreme Court took a courageous dip into the murky waters of modern philosophy of science and came up with a flexible recipe of four non-exclusive factors that could be used by the trial judge in determining the quality of the scientific evidence proposed: testability, peer review, standardization, and general acceptance.\textsuperscript{237}

Emphasizing the need for independent judicial inquiry into the reliability of all scientific evidence, \textit{Daubert} has been widely celebrated as a breakthrough in the attitude of the courts towards scientific evidence. The decision was labeled the “death of \textit{Frye}.”\textsuperscript{238} The seventy years since \textit{Frye} were declared to have been “seven decades of hiding from science,”\textsuperscript{239} and \textit{Daubert}’s new set of criteria was hailed as an attempt “to deal substantively with the problem of expert testimony in the

\textsuperscript{234} Id. at 576.
\textsuperscript{235} \textit{Daubert}, 509 U.S. at 593-94.
\textsuperscript{236} Id. at 594-97.
\textsuperscript{237} Id. at 592-94. The petitioners won the Supreme Court battle but lost the war. The Ninth Circuit reconsidered the evidence and affirmed its exclusion again, this time under the new \textit{Daubert} criteria. \textit{Daubert} v. Merrell Dow Pharm., Inc., 43 F.3d 1311, 1322 (9th Cir. 1995). For a detailed analysis of the \textit{Daubert} decision and its criteria, see \textsc{Kenneth R. Foster & Peter W. Huber}, \textsc{Judging Science: Scientific Knowledge and the Federal Courts} 4-16 (1997).
\textsuperscript{238} Note the title of the symposium published in \textit{Scientific Evidence After the Death of Frye: Daubert and Its Implications for Toxic Tort, Pharmaceutical, and Product Liability Cases}, 15 \textsc{Cardozo L. Rev.} 1938 (1994).
\textsuperscript{239} Black et al., supra note 203, at 722.
courts.” However, a perspective beyond seven decades allows us to see Daubert more as the offspring of Frye rather than its killer. True, Daubert displaced Frye, but it also inherited Frye’s exclusionary logic. No longer a passive umpire who watches over the rules of the game, the twentieth-century trial judge has become an active player, a gatekeeper charged with the responsibility for screening unreliable scientific evidence. First formulated by the Frye court, albeit in a milder form, this exclusionary spirit has since come to dominate American legal thought, and the Daubert decision only served to affirm it more vigorously. Its admissibility criteria may be different, but its mindset followed, and put the final stamp on, the twentieth-century trend toward ever-greater judicial scrutiny of scientific evidence.

VI. DISCUSSION

What can we learn from this concise history of scientific expert testimony? First and foremost, that scientific expert testimony in common law courts has a long and rich history. This simple lesson is particularly important in the context of the current tendency to present the malaise of science in the courts as a sign of our time, the result of the growing volume and complexity of late twentieth-century science. As this Article makes clear, this widely held assumption is flat wrong. Far from being new, the putative problems of scientific expert testimony in common law courts have existed since science was first introduced into the adversarial courtroom. The difficulties of fitting science within adversarial procedures; the reluctance of courts to mold a procedure that would shield science from the adversarial fire; the chronic inability of courts to bridge the gap between scientific experts and lay jurors; the failure of the legal and scientific professions to regulate the market of expertise; the resultant fear of a credulous jury bewitched in the name of science by charlatans and opportunists—none of these predicaments was new to the twentieth century.

Not only the problems but also the debate over their meaning and the ways to resolve them had by the late-nineteenth century all of the features that today are blithely

240 ANGELL, supra note 224, at 127.
242 See supra Parts I-IV.
assumed to be new. Almost every twentieth-century reform proposal can be traced back to the nineteenth century, including self-regulation, court-appointed experts, pre-approved experts lists, scientific tribunals and expert juries. All these reform proposals fell flat. Not surprisingly, then, the feeling by the late twentieth century was remarkably similar to that which prevailed a century earlier: the problem of expert testimony had reached rock bottom.

Legal scholarship has not been entirely oblivious of the problematic history of science in the courts. Still, the relevance and significance of this history have been undermined by at least two arguments. The first holds that experts who served as witnesses in the courts in the eighteenth and nineteenth centuries, and even in the early twentieth century, were not really scientists but experience-based experts. This argument is, again, wrong. Prominent scientific experts made regular appearances in courts of law throughout the nineteenth century, if not earlier. Furthermore, one could argue that in these earlier centuries law and science actually enjoyed closer, if not better, relations. Indeed, in the nineteenth century, which offered scientists very limited career opportunities, the legal system stood out as an exceptional and most lucrative market for scientific expertise—so much so that, while writing about the spectrum of Victorian science patronage, one historian of science could not help but wonder why it was that “on more than one occasion it would seem . . .

---

243 See supra Part III.C. For more on the nineteenth-century reform proposals, see GOLAN, supra note 38, at 107-43.
245 Id. at 1393-96; see also Joe S. Cecil & Thomas E. Willging, Court-Appointed Experts, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 525, 529-73 (1st ed. 1994).
246 Cheng, supra note 244, at 1395-96.
248 Cheng, supra note 244, at 1396-98.
249 See, e.g., HUBER, supra note 222, at 3-5.
250 See, e.g., Hand, supra note 2. For examples of historians’ accounts focusing on medical expert testimony, see Möhr, supra note 155; Landsman, supra note 33 (both articles); see also Hamlin, supra note 131.
251 See, e.g., Lee Loevinger, Science as Evidence, 35 Jurimetrics J. 153, 154 (1995) (“Many experts who served as witnesses in the courts in earlier centuries, and even in the twentieth century prior to Frye, were not scientists.”).
252 See supra Parts III, IV.
253 See supra Part II.
that all the chemists in London were being shared by aggrieved parties in a patent action."

In the 1850s, for example, August Hofmann, a major figure in the field of organic chemistry and a leading consultant to the artificial dye industry, was able to pull in a yearly income of eight to nine thousand pounds from his court appearances. Hofmann made this lofty income in patent litigation, which became a most profitable sideline activity for many leading men of science on both sides of the Atlantic.

It would be a mistake, however, to subsume the evolving relations between law and science under monetary interests alone. Early nineteenth-century scientists dreamt of becoming “true amicus curia . . . a terror to the guilty and a joy to the innocent.” But what had seemed in the early part of the nineteenth century as a central civil function of science had become by the end of the century a source of discontent. The courts were growing increasingly weary and wary of the conflicting scientific testimony, and the scientific community was growing increasingly bitter with the courts’ deployment of its services. Far from being irrelevant, then, the history of nineteenth-century scientific expert testimony tells a dramatic story: what had started as a great promise of cooperation between law and science ended with mutual disenchantment. Instead of bringing these two powerful fraternities closer, forensic science pulled them further apart.

A second assumption that served to discount the historical depth of the problems of science in the courts is that the growing volume and complexity of science made it increasingly harder for the judicial system to manage late


255 Id. at 5.

256 See GOLAN, supra note 38, at 81-96. For scientific experts appearing in American courts, see RALPH K. ANDRIST, STEAMBOATS ON THE MISSISSIPPI (1962) (steamboat explosions); DE VILLE, supra note 137 (malpractice cases); LAWSON, supra note 137 (agricultural cases); LUCIER, supra note 123 (patent cases); Paul Lucier, Court and Controversy: Patenting Science in the Nineteenth Century, 29 BRIT. J. HIST. SCI. 139 (1996) (same); Okun, supra note 137 (food and drug adulteration); Fred Quivik, Smoke and Tailings: An Environmental History of Copper Smelting Technologies in Montana, 1880-1930 (1998) (Ph.D. dissertation, Univ. of Pa.) (pollution cases); Spence, supra note 137 (mining cases).


258 See generally ENDLICH, supra note 173; Foster, supra note 159; Science in the Courts, supra note 152; HERSCHEL, supra note 141; Himes, supra note 142.
twentieth-century science-rich cases. However, this widely held assumption has never been corroborated by comparative historical studies. Moreover, it is not clear why we should even assume that present-day judges and jurors would find it harder to handle the scientific evidence presented before them in court. Is DNA profiling really more complicated than fingerprinting or nineteenth-century blood tests? Is the science involved in testing new pharmaceuticals drugs harder to comprehend than the nineteenth-century science involved in patenting artificial dyes? Many historians and philosophers of science would answer in the negative. Science, they teach, advances not by sheer accumulation of details and complexities, but by devising new theories with superior organizational and predictive powers, which are not necessarily more complex.

Even if we concede that modern science did grow more complex (certainly it grew richer and far more specialized), we can still point to other developments that actually made science more accessible to laypersons. Scientific language, literature, and training have all been systematized, and its credentials, equipment, and procedures standardized. The presentation of scientific evidence in court has also been improved by visual technologies such as medical imaging and computer animation and simulation. Finally, lay understanding of science has greatly improved during the twentieth century, following the vast increase in the availability of modern education. All this supports the notion that a present-day factfinder should find it easier, not harder, to handle the scientific evidence presented in court. Indeed, during the last two decades, courts have consistently pronounced their confidence in the ability of lay judges to do just that—to critically evaluate the scientific...

259 See, e.g., CARNEGIE COMMISSION ON SCIENCE, TECHNOLOGY, AND GOVERNMENT, supra note 226, at 1.

260 My own historical studies do not support this assumption. See GOLAN, supra note 38, at 211-64.

261 No philosopher has explicitly discussed this aspect of complexity in any detail. However, many philosophers have described scientific progress as responding to notions such as simplicity, organizational strength, and predictive power. This implies that complexity may actually decrease at any given developmental step, or at least that it would not necessarily increase. See generally THOMAS KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS 43-51 (1970); LARRY LAUDAN, PROGRESS AND ITS PROBLEMS: TOWARDS A THEORY OF SCIENTIFIC GROWTH (1977); see also PIERRE MAURICE MARIE DUHEM, THE AIM AND STRUCTURE OF PHYSICAL THEORY (Philip P. Wiener trans., 1954).
evidence presented to them. But if a lay judge can do so on her own, why not twelve lay jurors in their combined wisdom?

The question remains: if the problems of scientific testimony were neither new to twentieth-century courts nor necessarily harder to manage, then what did change during this period? What has driven the unmistakable trend toward ever-greater judicial scrutiny of scientific evidence? Further research is needed before a convincing answer can be provided. Still, any answer would have to take into account the history outlined here. The early twentieth-century judicial attitude, I suggest, was a direct response to the late nineteenth-century crisis of scientific expert testimony. Unable to check either the selection of the experts or their evidence in court, or the jury’s assessment of this evidence, the nineteenth-century courts were forced to admit almost everybody presented as experts, leaving juries to assess the ensuing battles between expert witnesses and lawyers. That changed in the early twentieth century with the professionalization of American scientists and the standardization of science education, literature, and practices. These developments provided the American legal system for the first time with the means to control the performance of scientific experts in the courtroom—by measuring the proffered individual expertise against the field’s own standards. First offered by the Frye court, this rationale flourished during the middle decades of the twentieth century, thereby inculcating the exclusionary judicial attitude toward scientific evidence, which reached its high-water mark at the end of the century with the Daubert opinion.

The rise of mass tort litigation during the last quarter of the twentieth century helped to extend the new judicial gatekeeping role from criminal to civil proceedings. The extraordinarily complicated nature of this new species of lawsuit forced trial judges to become even more active players in the litigation process, and to develop new rules and procedures with an eye toward both economic consequences as


263 See supra text accompanying notes 145-160. Whatever its failings, this arrangement was nevertheless justified as a necessary trade-off that provided the best “free-market” mechanism of proof testing and the best protection from the abuse of executive power. See, e.g., Choate on Trial by Jury; New York Lawyer Addresses the American Bar Association, N.Y. TIMES, Aug. 19, 1898, at 5.

264 See SCHUCK, supra note 224, at 3-15; Bernstein, supra note 214, at 385-86, 389-90.
The far-reaching effects of this litigation legitimated a new judicial role, less arbitral and more managerial in kind. Furthermore, the unprecedented financial risks for the defendants in mass tort cases induced powerful economic players to put their weight behind the long-stalling campaign for reform of legal procedures for handling expert testimony. The results were quick to show, and in the early 1990s the two-century-old debate was red-hot again. Having never addressed the problems of science in the courts, the U.S. Supreme Court found it necessary to do so three times during the 1990s, each time in the context of a product liability case. The important outcome of this series of Supreme Court opinions was not the oft-discussed four-prong Daubert test, but rather the further enhancement of the judicial gate-keeping role and its extension to civil proceedings.

Finally, one should note that the growing judicial scrutiny of scientific evidence has not been driven by the loss of judicial faith in science. On the contrary, despite the problematic history of science in the courtroom, the legal profession has never wavered in its trust in the scientific method. During the nineteenth century, this steadfast belief

265 SCHUCK, supra note 224, at 5-7
267 See, e.g., the twenty-two amicus briefs filed in Daubert (links to the briefs available at http://www.westlaw.com). The unusually large number of briefs included a large cohort of major commercial interest groups such as the American Insurance Association, the American Tort Reform Association, the Chamber of Commerce of the United States, the Pharmaceutical Manufacturers Association, the Product Liability Advisory Council, the Defense Research Institute, Inc., and the Washington Legal Foundation. The scientific community was represented by a host of academies, societies, colleges, and journals such as the National Academy of Sciences; the American Medical Association; the American Association for the Advancement of Science; the New England Journal of Medicine; the Annals of Internal Medicine; the College of Legal Medicine; a group of eighteen scientists, scholars, and teachers of science (including six Nobel laureates); and another group of prominent physicians, scientists, and historians of science (including the late Stephen Jay Gould).
268 See generally CARNEGIE COMMITTEE ON SCIENCE, TECHNOLOGY AND GOVERNMENT, supra note 226; HUBER, supra note 222; PRESIDENT’S COUNCIL ON COMPETITIVENESS, supra note 226; SCHUCK, supra note 224; Saks, supra note 224.
270 See Cheng, supra note 244, at 1401.
272 See supra text accompanying notes 125-127.
induced the judiciary to interpret conflicting expert testimony not as legitimate debate, but as a sign of moral decadence.273 During the twentieth century, it pushed the judiciary even further into scientific territory to exorcise charlatanism and differentiate good science from bad. Consequently, at the beginning of the twenty-first century, lay judges find themselves deeper than ever in the strange land of biostatistics, confidence levels, meta-analysis, and falsifiability, charged with the difficult task of weighing the merit of highly specialized scientific claims. How well the lay judges can meet these challenges and whether their gate-keeping role will lead to better adjudication are questions that will bear careful watching.

273 GOLAN, supra note 38, at 107-43; Hamlin, supra note 131, at 490-501.
Truth, Epidemiology, and General Causation

Douglas L. Weed†

INTRODUCTION

Does epidemiology, a basic science of disease causation, seek and find the truth? Certainly there are those who say it does: physicians,¹ epidemiologists,² even lawyers.³ Many philosophers of science these days, on the other hand, are extremely skeptical of claims to scientific truth.⁴ And were it not for some important practical matters facing epidemiologists in the courtroom and in other parts of our professional practice,

† Currently, Vice President, THE WEINBERG GROUP, INC. in Washington DC.
I want to acknowledge the helpful comments of Dr. Michelle Althuis in the preparation of this paper.

¹ For a discussion of the role of epidemiology “in the search for clinical truth,” see, e.g., Ann M. Kosloske, Epidemiology as the Search for Truth, 11 SEMINARS IN PEDIATRIC SURGERY 162, 162 (2002).

² As one epidemiologist stated in a discussion of post-viral fatigue syndrome:

For many, [post-viral fatigue syndrome] represents a “rag bag” diagnosis into which unsolved diagnostic problems are discarded. Others are in no doubt that there is a discrete syndrome, probably with a specific causation. The real answer almost certainly lies somewhere in between, but the truth can only be established through epidemiological studies designed to answer key questions.


³ “Both law and science are truth-seeking endeavors. In at least one respect, lawyers and scientists are like Agent Mulder on the ‘X-Files’: we believe that the truth is out there and our goal is to find it.” Arthur H. Bryant & Alexander Reinert, Epidemiology in the Legal Arena and the Search for Truth, 154 AM. J. EPIDEMIOLOGY (SUPP.) S27, S27 (2001).

⁴ For example, Peter Lipton writes:

The status of the truth hypothesis [according to which, science is generating increasingly accurate representations of a mind-independent and largely unobservable world], and so of realism, thus remains unsettled: it is neither undermined by the pessimistic induction nor confirmed by the miracle argument. Nor do I know of any other arguments that even come close to closing the question. This suggests that the rational attitude towards a scientific theory should never rise above the level of agnosticism.

we could leave it at that; epidemiologists seek the truth but have no theoretical justification for saying they have found it. Seeking the truth and actually obtaining the truth, in other words, are different activities evaluated in very different ways. The problem with this perspective is that those who rely on claims about disease causation—the patients, the public at large, and the litigants—may prefer that the professionals who have promised to help them—the physicians and epidemiologists, for example—act with “truth” rather than a “search for truth” on their side. At stake for all these individuals are actions that can change lives in profound ways: therapeutic decisions, public health recommendations, and legal opinions.

One popular escape from this conundrum is for epidemiologists (and any others who both do the science and use it) to lay claim to something close to the truth rather than truth itself. When asked about whether they have actually obtained the truth, epidemiologists typically do a little back-pedaling and say that they are only more-or-less certain about what is true. They say that their claims about what causes diseases are only approximations to the truth. In doing so, they allow themselves the luxury of never actually dealing with the nature of truth as such. They need only assert that they are trying to reach the truth and believe that they are somewhere close to achieving it. But they never actually find the truth. In the end, they are only more or less certain about being more or less close to the truth.

This approach seems to have worked pretty well in the practice of medicine and public health where no one ever promised to provide the truth in the first place. Promises are made, but not about the truth. Epidemiologists, for example, promise to prevent disease. Physicians promise to treat disease. Neither profession, however, promises to provide the truth about the scientific theories that undergird their respective practices. When pushed on this point, they will only

---

5 The idea that promises are made to society by epidemiologists and physicians emerges from the discussion of the nature of these professions in contemporary society. As I have described:

[M]odern professions are recognized by a common education, a common ethic, and professional standards. But at a deeper level, professions are characterized by whom they serve and what promise they make to assist those served. The original meaning of the word “profession” is to “declare publicly,” thus, professional declare they have special knowledge, they can help, and they will do so in the interest of others. Examples include physicians, teachers, and ministers. In each case, the professional fulfills a
say that they seek the truth and that they are more or less certain that they are somewhere near it.

The practice of law is another matter altogether. In the courtroom, when epidemiologists (and physicians and many others) provide expert testimony on disease causation in toxic tort litigation, they swear to tell the truth, the whole truth, and nothing but the truth. What does it mean, then, for epidemiologists (and for jurisprudence) if they make such a promise but only seek some approximation to the truth—something more or less close to the truth—rather than actually ever finding it?

My purpose in this essay is to examine this issue in terms that will require a critical assessment of the nature of truth in epidemiology. Of particular interest here is the role of truth in solving the problem of disease causation, the central scientific problem of the profession and a key concern in toxic tort litigation. I will limit my remarks to general causation, leaving aside causation in individuals, that is, specific causation.6

In Part I, I briefly discuss “telling the truth” well beyond the need for honesty and prohibitions against falsification and fabrication in the practice of science but also in the business of expert testimony. Telling the “truth” about science is to tell what has been published in the scientific literature, no more and no less. This is the truth of studies and their results and the methods used to obtain those results. In Part II, I discuss the implications of “telling the whole truth.” The whole truth, in this context, must include not only the so-called original published studies but also the commentaries on those studies. These typically appear as systematic reviews, editorials, and

promise inherent in the act of profession by making a claim and by following up on that claim by specific actions that identify that profession. Thus physicians claim to restore health through the central act of healing . . . .

[and epidemiologists] claim to prevent disease.


6 To consider the problem of general causation is to consider whether an agent (e.g., an exposure factor) is capable of causing disease, typically in a population. Michael D. Green et al., Reference Guide on Epidemiology, in FEDERAL JUDICIAL CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 333, 392 (2d ed. 2000). Specific causation, on the other hand, involves considering whether exposure to an agent was responsible for a single individual’s disease. Id. at 396.
other publications, including textbook chapters, where the studies’ strengths and weaknesses are described and where evidence is summarized, synthesized, and interpreted, for the purpose of making claims about causation. It is increasingly accepted in the practice of epidemiology that a systematic narrative review of the evidence is the appropriate venue for assessing cause and for making claims about disease causation. In Part III, I explore the most controversial component of the honesty oath: telling “nothing but the truth.” This will require a serious reconsideration of the nature of truth in epidemiologic science. I discuss the extent to which those components of the practice of epidemiology that transcend its written historical record can be considered true: the aims and values of its practitioners, their modes of reasoning, the theories and philosophies used to help solve the problem of general causation, as well as the causal claims themselves, whether published or not. Of particular importance is the fact that expert testimony goes beyond that which has been published. It can only be true in the same sense that published studies and published causal assessments are true if the expert applies the same methods and rigor to his testimony as that which is required for results and causal claims to appear and to be accepted in the peer-reviewed literature.

I. THE TRUTH

At first glance, no single concept seems better suited for what a scientist believes is her ultimate aim than the truth. The search for truth in science is legendary. It is a search for the really real, a search to separate fact from fantasy. The

---

7 Philip Kitcher writes about science and its search for truth as a point of view so popular and important that he dubs it “Legend.” PHILIP KITCHER, THE ADVANCEMENT OF SCIENCE: SCIENCE WITHOUT LEGEND, OBJECTIVITY WITHOUT ILLUSIONS 3 (1993). He notes that there have been differences among the versions of Legend. Some thought in ambitious terms: ultimately science aims at discovering the truth, the whole truth, and nothing but the truth about the world. Others preferred to be more modest, viewing science as directed at discovering truth about those aspects of nature that impinge most directly upon us, those that we can observe (and, perhaps, hope to control). On either construal, discovery of truth was valued both for its own sake and for the power that discovery would confer upon us.

Id.
search for truth is an independent and objective assessment of what is “out there” in the world in which we live.8

Or so we have been told, over and over, by those who have trained us. I realize now, however, some thirty years after the last lecture in graduate school, that no one ever explained to me during my scientific training what exactly this thing called the “truth” is. Indeed, no one has ever held something up in front of me, much less an entire audience at an international congress of epidemiology, and declared: This is the truth! We have found it! To put it more bleakly, no one has ever even proclaimed: This is the truth that we seek!

Nevertheless, scientists seem to deeply appreciate their special relationship to the truth. When they proclaim that they are in pursuit of the truth, they do not think that truth is a temporary stamp of approval or, at worst, merely a compliment that can be revoked.9 No. The truth that science seeks—when we hear that it has gone hunting—is a permanent special something, something to believe in and something to put our faith in. Small wonder there are those who align science with religion, as both seem to be in search of something that passes beyond understanding, yet still explains who we are, why we are here, how we live, why we get sick, and how we die.10

Perhaps it would help matters here to offer a definition of truth. The influential philosopher of science, Karl Popper, never a friend of definitions,11 nevertheless accepted what he

---

8 The scientist’s favorite philosopher, Karl Popper, created three worlds: “the physical world ‘world 1’, the world of our conscious experiences ‘world 2’, and the world of the logical contents of books, libraries, computer memories, and suchlike ‘world 3’.” KARL R. POPPER, OBJECTIVE KNOWLEDGE: AN EVOLUTIONARY APPROACH 74 (1972). The world “out there” is world 1, the physical world.

9 Richard Rorty, the somewhat radical neo-pragmatist, has expressed the view that the term “true” is merely a compliment or commendation for one’s point of view. He writes, “The pragmatist . . . . feels free to use the term ‘true’ as a general term of commendation in the same way as his realist opponent does—and in particular to use it to commend his own view.” 1 RICHARD RORTY, OBJECTIVITY, RELATIVISM, AND TRUTH: PHILOSOPHICAL PAPERS 23 (1991).

10 Rorty opines, for example, that we live in a secular world in which the scientist has replaced the priest. “The scientist is now seen as the person who keeps humanity in touch with something beyond itself.” Id. at 35.

11 Karl Popper writes:

One should never get involved in verbal questions or questions of meaning, and never get interested in words. If challenged by the question of whether a word one uses really means this or perhaps that, then one should say, “I don’t know and I am not interested in meanings; and if you wish, I will gladly accept your terminology.” This never does any harm. One should never quarrel about words and never get involved in questions of terminology.

POPPER, supra note 8, at 309-10.
called a common sense definition of truth: that which corresponds with the facts.\textsuperscript{12} With this definition, we can proceed as long as we have facts. For example, with this definition in hand, how can we not accept the truth of some well-accepted claims: the existence—the fact—of cancer, the existence—the fact—of viruses, and the claim (also a fact?) that a virus can cause cancer? To be specific, do we not accept as true that the human papillomavirus establishes itself in the deep recesses of the cells lining the cervix and then transforms some unlucky cell into something uncontrollable, malignant, and deadly? Do we not accept as fact and so true that the human papillomavirus causes cervical cancer?\textsuperscript{13}

This final question is a tough one to answer truthfully. For we have seen the cancer and the virus with our own eyes in photos of malignant cells from pathologists’ slides and electron microscopic images of the virus. These we take to be true. And we might also accept as true the explanation that this cancer is caused by this virus, if it were not for the fact that we have not seen causation. Nowhere in the peer-reviewed literature, in the tables, nor in the graphs and figures displayed in all the studies that have been published, can we find this thing called “causation.” Not one of these shows us that, in fact, human papillomavirus causes cervical cancer. Not one. To tell you the truth, we have seen causation only in words.

There are many examples of this phenomenon. There are many such causal claims, many more now than there were a few decades ago, that can be found in this same peer-reviewed literature, alongside the photos of the viruses and the mutated genes—the exposures—and the cancers and other diseases—the outcomes—and I believe these to be statements that those who wrote, reviewed, edited, and published them believe are true, these statements that some exposure caused some outcome. But this is a different sort of truth, is it not, than the truth we assign to the photos and the tables, the graphs, the figures, and the numbers themselves, and to the

\textsuperscript{12} “[W]e can define, purely verbally, yet in keeping with common sense: A statement is true if and only if it corresponds to the facts.” Id. at 46 (footnote omitted).

existence—the fact—of the causal claims themselves that sit there to be read in the peer-reviewed literature?

The photos and all the rest, including the causal claim, are true in the sense that they were recorded at a certain point and place in time and will remain there for perpetuity, barring some unnatural calamity along the lines of the great fire at the Alexandria library. Anyone can see them. There is also a clear trail for each back to its origin. The photo of the virus came from a laboratory where a researcher turned on an electron microscope to take a snapshot of the parasite in all its symmetrical and deadly beauty. Likewise, a pathologist photographed the cells of a tumor that a surgeon carefully cut from the cervix of a woman who had the bad luck to be so afflicted. Finally, there is the author of the claim itself—that this virus causes that cancer—perhaps the same surgeon who also trained in epidemiology (a rare individual). Each of these individuals has a name—the laboratory researcher, the pathologist, the surgeon, and the patient herself—and their contributions to the existence of the photos and the sentence (all of which could have appeared in the same review article about human papillomavirus and cervical cancer) can be documented, verified, observed directly, and accepted without any uncertainty at all. These really are truths, plain and not so simple.

But the causal claim itself—that this type of virus caused that sort of cancer—does not have this same sort of connection back to some unique event that can be documented, verified, and directly observed. The causal claim is a scientific hypothesis and we cannot ever know if it is true in the same sense as the existence of the virus, the cancer, and its author. The hypothesis can be well supported or not by the available evidence. It can be more or less certain, more or less proven, but it cannot ever be true. The reason is remarkably straightforward. Causation cannot be seen. Causation cannot be proven. And the evidence for causation always under-determines our capacity to choose between the causal hypothesis of interest and its various alternatives.14 Nor can

---

14 Causal hypotheses, however much we believe in them and regardless of how much scientific evidence supports them, can never be true. This remarkable claim can best be understood by explaining how the scientific practice of making claims about disease causation is affected by three fundamental (causal) problems. These problems, which can also be considered constraints on scientific practice, affect anyone who examines scientific evidence for the purpose of making causal claims. Like a plague, these constraints are not negotiable. They involve, but are not limited to, the extent to
which we can see (observe) the thing we call “causation” and the extent to which we can prove any causal hypothesis. In the end, these constraints prohibit us from claiming that scientific hypotheses about disease causation are true. For the purposes here, I refer to these three problems (or constraints) as (1) the fundamental problem of causal inference, (2) the fundamental problem of causal logic, and (3) the fundamental problem of causation.

First, the fundamental problem of causal inference is that we cannot observe on the same individual both the effect of a cause (e.g., a disease outcome) and what would have occurred had the cause not acted to produce its effect. See generally Paul W. Holland, *Statistics and Causal Inference*, 81 J. AM. STAT. ASS’N 945 (1986). This constraint is sometimes called the “counterfactual” condition. If, for example, an individual begins taking a new medication and is later diagnosed with an illness or condition, we cannot know if that person would have contracted that illness or condition without taking the medication. This fundamental problem of causal inference is the primary reason why randomized clinical trials are considered the gold standard of scientific research in therapeutic and preventive clinical research; randomized trials provide the best approximation to solving this problem by assuming that the individuals who do not receive the medication or other intervention (e.g., the placebo controls) are as similar as possible to the individuals who do receive the intervention. Certainly, the control group is not exactly the same as the treated group, but randomization assures that the differences between the two groups are distributed evenly between them.

This fundamental problem of causal inference also explains why control groups in any study of human health effects are so important. Epidemiological studies do not randomly identify controls, but controls are carefully selected, nevertheless, for precisely the same reason they are used in randomized trials: to provide an approximate solution to this fundamental problem. Claims about disease causation from studies that lack control groups are of questionable validity and reliability.

The second constraint, the fundamental problem of causal logic, is also known as the problem of “underdetermination.” Underdetermination means that the available scientific evidence cannot once-and-for-all determine which hypothesis is the true hypothesis among all those involved in a particular situation, such as the main causal hypothesis, alternative causes, and chance. It is a straightforward if frustrating fact that scientific tests of causal hypotheses cannot provide proof or disproof. By “proof” I mean the sort of absolute proof found in mathematics and symbolic logic, the kind of proof that cannot be questioned. See generally Ernan McMullin, *Underdetermination*, 20 J. MED. & PHIL. 233 (1995); Douglas L. Weed, *Underdetermination and Incommensurability in Contemporary Epidemiology*, 7 KENNEDY INST. ETHICS J. 107 (1997), available at http://muse.jhu.edu/journals/kennedy_institute_of_ethics_journal/v007/7.2weed.html. The available scientific evidence always underdetermines the choice of the best hypothesis that explains that evidence.

Underdetermination implies that there is always room for questioning the validity and reliability of any scientific test of any scientific hypothesis. Causal hypotheses in medicine—such as the hypothesis that a medication (such as pramipexole) causes a condition (such as pathological gambling)—are never proven nor disproven in a scientific sense. Put another way, the available scientific evidence underdetermines the choice between the various alternative hypotheses that can explain that evidence. A randomized controlled clinical trial, for example, is a very strong test of a causal hypothesis, but it does not prove—one and for all—that a factor (e.g., a treatment) causes an outcome (e.g., a cure). That does not mean, of course, that we do not use a medication once a randomized trial has shown it to be effective; acting on the basis of the results of a trial and using the results of a trial as proof of a scientific hypothesis are two very different phenomena, an important point to which I will return. Case series, on the other hand, are extraordinarily weak—essentially irrelevant and highly unreliable—tests of causal hypotheses; they cannot help us to choose between the alternative hypotheses. A “case series” is a report of a collection (or group) of individual patients who have experienced a disease or other adverse event and were also exposed to some purported causal exposure (e.g., a medication or a
causation be made certain. It is, at best, an expert’s judgment, at worst, an expert’s guess.\textsuperscript{15}

This is what I take to be the true that I, as an epidemiologist appearing in court to provide expert testimony, can swear to: that exposures exist, that outcomes exist, and that causal claims exist. One can also swear that there is evidence—scientific studies—linking the exposures to the outcomes (or not). And there are methods used in the design, analysis, and interpretation of those studies—methods both quantitative and qualitative, methods analytical and synthetic, methods of science, of epidemiology, of statistics, and of philosophy itself. That is the truth and nearly the whole truth.

II. THE WHOLE TRUTH

Telling the truth is telling that which corresponds with the facts. This telling of scientific facts is so straightforward and clear that it is unquestionable at best and, perhaps for some, unremarkable at worst. The facts of any case of disease causation include the published reports of studies designed to

\textsuperscript{15} "[Judgment] is regularly invoked when scientific evidence is used to make a claim about disease causation in the courts or in regulatory risk assessment." Douglas L. Weed, The Nature and Necessity of Scientific Judgment, 15 J.L. & POLY 135, 135 (2007). Indeed, judgment is probably always used in any such assessment. "[S]cientific judgment is not easy to define, although we are fairly clear about the kind of judgment we prefer: good, sound, and unbiased." Id. at 140. It is reasonable to assume that some scientists’ judgments are better than others.
describe the exposure and the disease in question as well as those studies designed to test the hypothesis that the exposure caused the disease. In order to approach the “whole truth,” however, to these studies we must add the reviews—ideally, systematic narrative reviews—of those studies where searching, selection, summarization, and interpretation in causal terms occur. Reviews are critically important publications. Within these reviews (and commentaries, editorials, textbook chapters, and even letters to the editor), experts discuss their positions on issues including internal and external validity, methodological concerns (both quantitative and qualitative), as well as sources of funding, ethics, and practical implications of the research findings under review. These all count as facts, in a very broad sense, scientifically.

Curiously enough, both science and the law share this feature; careful records of events are kept in both: studies, reviews, and commentaries in science, and cases, opinions, and commentaries in the law. On the science side, we could also add the textbooks, popular (media) commentary, and last, but certainly not least, what philosophers of science think about the whole complex and lengthy business just described. If by “facts” we mean that which has been published, then there is only one (major) problem for the expert who raises his hand and swears to tell the truth and nothing but the truth: just how much of this written record really matters?

It is no secret that, in practice, science sees this issue—what evidence really matters in assessing causation—differently than does the law. The rules for selecting the truth (as historical record) differ in these two practices. Scientific assessments tend to be much more inclusive than those in some courts, where judges have the power to accept and reject (as inadmissible) every potential “fact” that the parties bring to their attention, subject to the laws of evidence. There are many examples of courts selectively rejecting as inadmissible scientific evidence that would typically be included in a systematic review in the practice of epidemiology.16

While the admissibility of facts is governed by the rules of evidence in the law, science too has rules regarding the

16 For two different accounts of the admissibility issues post-Daubert in the context of general causation in matters relating to health, see Joe S. Cecil, Ten Years of Judicial Gatekeeping Under Daubert, 95 AM. J. PUB. HEALTH (SUPP. 1) S74 (2005); Ronald L. Melnick, A Daubert Motion: A Legal Strategy to Exclude Essential Scientific Evidence in Toxic Tort Litigation, 95 AM. J. PUB. HEALTH (SUPP. 1) S30 (2005).
inclusion and exclusion of facts for consideration in an assessment of disease causation, as well as rules covering which studies to include in a systematic narrative review, a process that has been evolving from around the mid-1980s to the present. Today, a systematic narrative review is a careful and transparent process involving the use of search terms and databases, and inclusion and exclusion criteria. The process includes searching (in some instances) for not-yet-published studies and studies required by regulatory agencies but never published, and lengthy discussions about the quality and validity of these same studies, published or not, but worthy of publication nevertheless. This process of systematically reviewing all (or nearly all) the evidence has become standard accepted practice in epidemiology and medicine, indeed, in all the biomedical sciences save perhaps for the so-called basic (laboratory) sciences—molecular biology, toxicology, etc.—which tend to produce interesting and important but not-so-systematic reviews of disease mechanisms.

A central purpose of a systematic review is to determine if the available evidence sufficiently supports and/or warrants a claim of causation. Causal claims, in theory, could be—have been—made based on the results of single studies and, indeed, based on reports of adverse events in single individuals, but these are rare and unusual circumstances and, to put it bluntly, more likely decisions about what needs to be done (to protect the public’s health) than well-tested claims about causation. Causal claims typically require a body of evidence comprised of many studies from different disciplines along with

17 For an example of methodological guidelines for systematic reviews, see Douglas L. Weed, Methodologic Guidelines for Review Papers, 89 J. NAT’L CANCER INST. 6 (1997).

18 A systematic review of the evidence is one of several “weight of evidence” methods. See Douglas L. Weed, Weight of Evidence: A Review of Concept and Methods, 25 RISK ANALYSIS 1545 (2005).

19 The use of case reports—that is, reports of an adverse event in a single individual or a series of individuals (also known as a case series)—as scientific evidence to withdraw a medication from the market is discussed and debated in the area of pharmacovigilance. See J.A. Arnaiz et al., The Use of Evidence in Pharmacovigilance: Case Reports as the Reference Source for Drug Withdrawals, 57 EUR. J. CLINICAL PHARMACOLOGY 88, 88-91 (2001); Bruno H.Ch. Stricker & Bruce M. Psaty, Detection, Verification, and Quantification of Adverse Drug Reactions, 329 BRIT. MED. J. 44, 44-47 (2004); Jan P. Vandenbroucke, Case Reports in an Evidence-Based World, 92 J. ROYAL SOC’Y MED. 159, 159-63 (1999); Geoffrey R. Venning, Identification of Adverse Reactions to New Drugs III: Alerting Processes and Early Warning Systems, 286 BRIT. MED. J. 458, 458-60 (1993).
the methods of causal inference used to summarize and interpret that evidence.

Given the thoroughness of systematic reviews and their centrality in the fields of epidemiology and medicine, it follows that they are not only important facts to be considered in an expert’s testimony, but the single most important facts, the single most important truths, to be recounted, examined, and critiqued. And that brings us to the final component of the honesty oath: “nothing but the truth.” For if the “truth” is that which has already appeared in print, what is left for the expert if he has already told that truth, indeed the whole truth with the studies and the reviews and commentary and all the rest? What remains if that expert has promised to tell “nothing but the truth?”

III. Nothing But the Truth

As long as we maintain the premise that the truth of science is that which has been published—as carefully recorded historical events—then the expert who promises to tell nothing but the truth has apparently no greater role in the proceedings than any other reporter of events. In this situation, the scientific expert provides an account of the studies that have measured exposures and diseases and have combined that information in such a way that the causal hypotheses of interest were more-or-less tested. He also provides an account of the commentary on those studies, and, most importantly, he provides an account of the systematic reviews of those same studies, with their causal claims, recommendations for actions, and/or recommendations for more research.20

The only way the expert’s opinions on these matters—his own assessment of the relevance, reliability, and validity of the evidence, the commentaries, and especially the systematic reviews and their claims—can be considered “true enough” is if

20 As I have written elsewhere,

[Systematic] literature reviews . . . are a longstanding form of synthetic method and are ubiquitous in science, appearing in peer-reviewed journals, in textbook chapters, and in background material for research funding applications. The purposes of systematic reviews can include: claims regarding general causation, efficacy of therapeutic or preventive interventions, recommendations for such interventions, and recommendations for future research.

he presents them in a manner that would be acceptable in the practice of science, using the same methods of assessment, the same systematic search procedures, and the same criteria (or guidelines) for making causal claims. To be compatible with the promise to tell nothing but the truth, an expert’s report for the court must be as well-prepared and carefully documented as a highly-regarded systematic review from any of the several institutions which regularly produce such reports: the Institute of Medicine (“IOM”) in the United States, the Cochrane Collaboration in the United Kingdom, and the International Agency for Research on Cancer (“IARC”) in France, to name only three.21 The expert, in short, must opine as if he intends to submit his views for peer-reviewed publication, following reasonably well-established guidelines for transparent systematic reviews. Only then will his opinion on causation be true enough to be consistent with the promise he made to the court.

By no means am I suggesting that the current methods for making causal claims from scientific evidence (and I include here all types of relevant evidence, from human epidemiologic studies to lab-based toxicological studies) are without errors and thus free from the need for serious improvements.22 Nor am I suggesting that the process of peer review in scientific publishing eradicates all errors—intentional, accidental, or fundamental. Far from it. Both processes can be (must be)

21 The Institute of Medicine, the Cochrane Collaboration, and the IARC regularly produce highly regarded systematic reviews of causation issues. For representative examples, see COMMITTEE ON ASBESTOS, INST. OF MEDICINE, ASBESTOS: SELECTED CANCERS (2006); Jimmy Volmink et al., Research Synthesis and Dissemination as a Bridge to Knowledge Management: The Cochrane Collaboration, 82 BULL. WORLD HEALTH ORG. 778 (2004); INT’L AGENCY FOR RESEARCH ON CANCER, WORLD HEALTH ORG., IARC SCIENTIFIC PUB’N NO. 100, CANCER: CAUSES, OCCURRENCE AND CONTROL (L. Tomatis ed., 1990).

22 As I have written elsewhere:

It is no exaggeration to say that any epidemiologist who claims he is an expert—that he can reliably make claims about causation—is either hopelessly naïve or a flagrant prevaricator. As noted earlier, I do not mean to suggest that prior claims about what factors or exposures cause illnesses are incorrect.

Douglas L. Weed, Causation: An Epidemiologic Perspective (In Five Parts), 43 J.L. & POL’Y 52, 43-53 (2003). My primary concern here is that a careful description of the methods of causal inference used in practice is an absolute necessity. It is also the case that these methods are in need of improvement. That does not mean that an individual epidemiologist cannot be an expert in the methodology of causation. My main message here is that the methods of causal inference need to be improved, and thus the user must recognize their limitations. Nevertheless, the best situation in practice is to comprehensively describe and systematically reference these methods when they are used. When no method is described, any causal claim is suspect.
improved. Both have a long way to go. The methods of causal inference, for example, which we use to synthesize and summarize the available scientific studies and which incorporate criteria (or guidelines) for interpreting the summarized evidence, are chock full of values—both scientific and extra-scientific—of the practitioners who invented and use them.23

To be specific, consider the list of causal criteria (sometimes called considerations, aspects, or guidelines) used by epidemiologists, physicians, federal regulatory agencies, and just about everyone else (including toxicologists) who examines scientific evidence for the purpose of making causal claims. There are nine such criteria in traditional accounts. Hill’s criteria, for example, were described by the medical statistician, Sir Austin Bradford Hill in 1965; they include strength, consistency, dose-response, biological plausibility, temporality, specificity, coherence, experimentation, and analogy.24 Fundamentally, these are qualitative values with links (in some but not all instances) to quantitative results. They are values because we believe them to be that which matters when assessing the scientific evidence for causation. Other important scientific values are the concepts of relevance, reliability, validity, and statistical significance. Interestingly, Hill’s criteria do not include any of the aforementioned concepts, nor do they include predictability or testability. That simple fact points out the complexity of the process. All these concepts/criteria/guidelines/values (or whatever else you want to call them) are important when making causal claims: relevance, reliability, validity, statistical significance, predictability, testability, consistency, strength, dose-response, plausibility,

23 As I have previously commented,

[In the practice of causal inference . . . reviewers of scientific evidence come to opposite conclusions using methods that are quite similar on the surface, but quite different at deeper levels, where numerous choices are available regarding criteria, rules of inference, and other components of the methodology. Differing conclusions emerge because reviewers hold dear different scientific and extrascientific values that affect these choices.]

Weed, supra note 14, at 122-23.

24 Hill’s criteria (or considerations) are one component of a general method of causal inference. For an account of Hill’s criteria in the practice of causal inference in epidemiology, see Sir Austin Bradford Hill, The Environment and Disease: Association or Causation?, 58 PROC. ROYAL SOC’Y MED. 295, 295-305 (1965); Weed, supra note 22, at 43-53.
temporality, experimentation, specificity, coherence, and finally the lowly and much-maligned concept of analogy.25

It is beyond the scope of this paper to sort out the relationships among these various values, including how best to prioritize them, their definitions, and rules of inference. One point to be made here is that all have been used—explicitly and, more likely, implicitly—in the practice of causal inference over the past fifty years. They are important components of an overarching method of causal inference. To these facts I will swear. That is the truth, a part of the whole truth, and nothing but the truth.

SUMMARY

Perhaps it is obvious by now that I do not see the “truth” as the aim of epidemiology if by “truth” we mean some final, absolute, and uncriticizable set of certain statements about the causes of diseases. The truth of my science is that which has been recorded in its literature. I can relate to that truth as the whole truth, if I give special attention to the systematic reviews that provide an overall assessment of general causation. In turn, I can provide my own assessment of general causation in terms of the methods used to make such inferences, applied to the available evidence, including prior reviews. In doing so, I provide nothing but the truth.

The scientific aim of epidemiology is explanation rather than truth. These explanations arise from and are tested in the use of our analytical and interpretative methods as well as judgment. They are the best explanations we have today. We may choose to act upon them, or not.

25 It is an unfortunate fact that the authors of one of the most influential intermediary textbooks of epidemiology believe that analogy is at best an exercise in creative thinking. They suggest that if you cannot find an analogy between one potential causal association and another, you simply lack imagination. KENNETH J. ROTHMAN & SANDER GREENLAND, MODERN EPIDEMIOLOGY 24-27 (1998).
Causation, Truth, and the Law

Richard Scheines∗

I. INTRODUCTION

Deciding matters of legal liability, in torts and other civil actions, requires deciding causation. The injury suffered by a plaintiff must be caused by an event or condition due to the defendant. The courts distinguish between cause-in-fact and proximate causation, where cause-in-fact is guided by the “but-for” test: the effect would not have happened, but for the cause.1 Proximate causation is a set of legal limitations on cause-in-fact.

As this conference is entitled A Cross-Disciplinary Look at Scientific Truth: What’s the Law to Do, I will ignore the distinction between cause-in-fact and proximate causation, and instead focus on both the sense in which cause-in-fact claims can be considered true or false and on the challenges to establishing them.

Before a court can decide on proximate causation, and thus on liability or damages, it must decide on the truth of the cause-in-fact question: Was the injury suffered by the plaintiff caused by the action(s) or inaction(s) of the defendant? For example, was John Smith’s liver cancer caused by his exposure to trichloroethylene (‘‘TCE’’) in a factory that employed him for ten years?

I will focus on what we must assume before cause-in-fact claims can even be said to have a truth value, that is, objectively true or false independent of whether we can know it. Philosophers distinguish individual level causal claims (cause-in-fact claims) from ‘‘general causal claims.’’2 I will try to

∗ Dr. Scheines is a Professor (and Head) of Philosophy at Carnegie Mellon University, with courtesy appointments in the Department of Machine Learning and the Human-Computer Interaction Institute.


make the difference clear and then discuss the difficulties that arise in claiming that either have a determinate truth value. I will argue that if there is a truth to the matter about individual causal claims, it is parasitic upon the truth of general causal claims, and that therefore the relevant issues for the law involve deciding on the truth of general causal claims.

Given the limited amount of relevant empirical scientific evidence that is typically available, however, deciding in a legal setting whether such claims are actually true or false can be extremely difficult. As courts have no choice but to decide such matters, they need a rational process by which to synthesize the evidence for or against causal claims—both with respect to our best scientific guess about the truth of the claim and with respect to the scientific uncertainty about such a guess.

I will sketch the various forms of evidence that are used to prove general causal claims and then describe the strategies and problems associated with synthesizing the totality of this evidence into a single judgment, both with regard to the truth of a causal claim and with regard to the uncertainty with respect to that judgment.

II. INDIVIDUAL VERSUS GENERAL CAUSATION, INDETERMINISM, AND TRUTH

A. Individual Versus General Causal Claims

Consider first the difference between individual and general level causal claims. In legal contexts, the goal is often to establish whether one particular event or condition was the cause of another particular event or condition. For example, if John Smith contracts liver cancer, a court might seek to establish whether or not his exposure to TCE in a factory that employed him for ten years was the “proximate cause” of his particular cancer. In such cases, by saying that exposure to TCE caused the disease, courts typically ask whether the cancer would not have occurred but for the exposure to TCE.\(^3\) This is an individual level causal claim, and one whose truth, if

---

\(^3\) See Wright, supra note 1, at 1775; see also Jane Stapleton, Legal Cause: Cause-in-Fact and the Scope of Liability for Consequences, 54 Vand. L. Rev. 941, 958-61 (2001).
it has one, depends on a counterfactual claim: what would have happened to John Smith had he not been exposed to TCE.4

General causal claims refer to a population of individuals, and concern the probability or average severity of a property (for example, a disease) in that population. For example, a qualitative general causal claim about TCE and liver cancer might be: in a population of factory workers who were exposed to TCE, the probability of getting liver cancer (the risk among the exposed) is higher than it would have been in the same population had they not been exposed to TCE. A quantitative version of the same general causal claim: the risk of liver cancer among those exposed to TCE was 2%, while the risk in the same group, had they not been exposed, would have been 1%. Another example is the following: among middle class American children between the ages of five and ten, if everyone had watched one less hour of TV per day, then the average Body Mass Index (“BMI”)5 of the group would have been .5 point lower than it was.

Clearly, claiming that TCE causes liver cancer in a population of workers does not entail that every worker who was exposed to TCE will develop liver cancer, and it does not entail that every case of liver cancer among the workers would not have happened but for TCE exposure. Similarly, not all children would have lost .5 point of BMI had they decreased their TV watching by one hour per day, etc. Again, the general causal claims each make a counterfactual claim. In the TCE and liver cancer case, the claim is: had the same population lived the same life, with the exception of not being exposed to TCE, then the probability of liver cancer would have been lower than it was in the actual world. In the TV and obesity case, the claim is: had the same population of American children lived the same life, with the exception that they had watched an hour less of TV per day, then the average BMI in the population would have been .5 point lower than it was in the actual world.

4 David Lewis developed the most influential account of counterfactuals in philosophy, see DAVID LEWIS, COUNTERFACTUALS (2001); and Donald Rubin developed the most influential account of causation based on counterfactuals in Statistics, see Donald B. Rubin, Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, 66 J. EDUC. PSYCHOL. 688 (1974).

5 BMI is calculated as 703 x (weight in pounds) / (height in inches)².
B. General Causal Claims and Truth

Since our target is the law, which for the most part deals with events that have already occurred, I will restrict my attention to causal claims about things that have already happened. For example, was John Smith’s liver cancer caused by his exposure to TCE, or was the rise in obesity in the late twentieth century caused by an increase in TV watching?

On this sort of counterfactual account of causation, consider what it might mean to assert that general causal claims have a truth value, that is, that they are either true or false in the actual world. Two objections arise immediately: the vagueness and objectivity of counterfactual possible worlds, and the meaning of probability.

That causal claims in the actual world might depend upon what would or would not happen in alternative possible worlds bothers almost everyone. The problem is that most descriptions of alternative possible worlds seem intolerably vague. For example, in the TV and obesity claim we are asked to consider a world in which the same children lived the same life but watched one hour less of TV per day. How exactly do we imagine the change in their world so that they each watch one hour less of TV per day? Do we make the TV inoperable one hour before they would have turned it off anyway? Convince their parents to intervene and select an hour of TV every day the child will no longer watch? Offer them just enough of a candy reward to get them to voluntarily shut down the TV one hour before they would have anyway? Make them replace the first hour of TV they would have watched with exercise? The counterfactual as we stated it is vague—it can’t answer any of these questions even though they all obviously matter for assessing the causal claim.

One can, however, fully specify a manipulation or intervention that would change the actual world to the possible world we are considering in a way that eliminates all this vagueness. Donald Rubin famously articulated a counterfactual theory of causation based on drug trials. If we consider an experiment in which some people received a pill containing a drug (the treatment) and the other half received a pill identical in appearance, taste, etc. (the control), then the causal inference problem with respect to the drug amounts to missing

---

6 See Rubin, supra note 4.
data from alternative possible worlds. For the people who took the treatment pill, we are missing the data on what would have happened to them if they had taken the control pill. For the people who took the control pill, we are missing the data on what would have happened to them if they had taken the treatment pill (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Took treatment pill</th>
<th>Took control pill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane Doe</td>
<td>Cured</td>
<td>??</td>
</tr>
<tr>
<td>Person 2</td>
<td>Ill</td>
<td>??</td>
</tr>
<tr>
<td>. . .</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person N - 1</td>
<td>??</td>
<td>Ill</td>
</tr>
<tr>
<td>Person N</td>
<td>??</td>
<td>Ill</td>
</tr>
</tbody>
</table>

Table 1  Missing data needed to assess causal efficacy of treatment.

Since the pills are identical in appearance, etc., surely it is not difficult to be perfectly precise about the alternative worlds under discussion. Jane Doe took the treatment pill and recovered—what would have happened if she had taken the control pill? It isn’t hard to imagine the antecedent: an alternative world in which we leave everything as we found it in the real world, except for removing the drug from the pill Jane took. Knowing whether or not, in this hypothetical world, Jane would have remained ill or recovered is not so simple.

The point is this: the problem of vagueness is not insurmountable. It requires being clear about the intervention performed to transform the actual world into the counterfactual world.\(^7\) In the TCE and liver cancer case, we might describe a world in which the factory workers behaved identically, but the de-greasing chemical used in the factory

\(^7\) The idea of making counterfactuals clear by formalizing the idea of an intervention has been developed extensively. See generally JUDEA PEARL, CAUSALITY: MODELS, REASONING, AND INFERENCE (2000); PETER SPIRITES ET AL., CAUSATION, PREDICTION, AND SEARCH (2d ed. 2000); James Robins, A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods—Application to Control of the Healthy Worker Survivor Effect, 7 MATHEMATICAL MODELLING 1393 (1986) (errata appears in 14 COMPUTERS & MATHEMATICS WITH APPLICATIONS 917 (1987)).
in which they worked was changed from TCE to a specific alternative known to be non-carcinogenic.8

Assuming that vagueness is not an issue, what about the probability part of general causal claims? Recall that to assess the TCE and liver cancer claim, we need to know the probability of liver cancer in the population exposed to TCE and in the same population not exposed to TCE.

What is it for a probability claim to be true? Given two claims about a coin in my pocket:

H1: the coin is fair (probability of heads = .5)
H2: the coin is loaded 75% toward heads (probability of heads = .75)

What is it for H1 to be true but H2 false? Unfortunately, neither hypothesis puts any binding constraints on any experiment we might conduct in the actual world. We might say that H1 implies that the proportion of heads in a very long sequence of flips should converge to .5 as the sequence gets longer and longer, but that this is not the case for H2. But the “should” in this sentence is itself probabilistic. Any finite sequence of coin flips is consistent with both of these hypotheses. Perhaps this is just philosophical obstructionism. Even though we don’t yet possess an entirely satisfactory account of what it means for H1 to be true and H2 false in the actual world, several accounts are out there.9 We don’t want to put the legal system on hold until the philosophers can agree on a semantics for probability.

To summarize, provided we can be sufficiently precise about the manipulations (interventions) that will transport us from the actual to a counterfactual possible world, and provided probability claims have a coherent semantics, then general causal claims have a truth value as well. At minimum, their truth depends on the probability of the effect in two populations.

C. Individual Causal Claims and Truth

Now consider whether individual causal claims have a determinate truth value. Was John Smith’s liver cancer caused by his exposure to TCE? Was Jane Doe’s illness cured by the experimental drug she took? Again, these claims depend upon

---

8 Presuming there is such a thing!
evaluating counterfactuals. What happens to John Smith in the world in which he wasn't exposed to TCE? What happens to Jane Doe in the world in which she takes the placebo pill? Having argued that these counterfactual worlds are not necessarily insurmountably vague, let us now consider whether these questions have definite answers.

Again, the answer depends on probability, but in a completely different sense than we have already discussed. It depends upon whether the world is deterministic, as we are psychologically built to expect, or indeterministic, as the physicists tell us it is.

Roughly, in a deterministic world, the past fully determines the future. What happens to Jane Doe after she takes the pill is not a lottery, but a sure thing. We might not know enough about the details to be able to predict what will happen, but that is an epistemic limitation and not a feature of the world we inhabit. In an indeterministic world (still guided by physical laws), the past does not determine the future, but at most determines the probabilities of possible futures.

Although Einstein detested the thought of an indeterministic world and said so (“[God] does not play dice”\textsuperscript{10}), in modern physics God is indeed a gambler. When a single electron is shot at a screen that records the point where it “hits,” all that can be predicted about its landing site is its probability. The \textit{probabilities} can be predicted perfectly, and experiments using thousands of electrons have confirmed the accuracy of these probabilistic predictions, but the \textit{exact position} on a single trial cannot be predicted perfectly, \textit{no matter what we know} about the electron. Even if we learned everything there was to know about the electron midway through its flight, and the universe within which it traveled, it would not be enough to determine where it will hit the screen. We are simply not evolved to accept this idea fully, but this is the way the world seems to be.

Translated back to Jane Doe, if the world is truly indeterministic, then although she took the treatment pill and recovered in the actual world, there is \textit{no fact to the matter} about what would have happened to her if she had taken the control pill. And it has nothing to do with the vagueness of the counterfactual.

If the world is truly indeterministic, then even if we could go back in time and replay the world millions of times from the exact spot we like, for example, leaving everything the same but changing the contents of Jane’s pill, the outcome on each play is still truly a lottery. Analogous to the electron’s landing site on the screen, the probabilities for Jane’s outcome might be determined and perhaps knowable, but the outcome on each play is not. Thus, if one uses the basic legal test for causation, that is, the but-for test, then individual level causal claims simply do not have a truth value in a world in which God actually does play dice.

If the world is deterministic, then probability statements capture only our epistemic uncertainty, not something more fundamental about the world. For example, suppose that a population exposed to TCE has a probability of getting liver cancer of .02. It might be the case that some people in the population have an unusual genetic makeup such that, if they are exposed to TCE they will definitely get liver cancer, but if they are not exposed they definitely will not get cancer, and that 2% of this population has the unlucky gene. In 1960, before we could sequence an individual’s genes, we would not be able to tell which of the individuals were lucky or unlucky. Thus, for any individual J, chosen at random, J’s fate is determined, but our epistemic access to it is limited to the true claim: “J was exposed to TCE and thus the probability that he will get liver cancer is 2%.” The underlying situation is deterministic, but due to our limited access appears indeterministic. Philosophers refer to such a world as pseudo-indeterministic.11

In a pseudo-indeterministic world, individual level causal claims do have a truth value. Although we might not have access to his genome, individual J either has the unlucky gene or he doesn’t, and whichever it is determines his cancer outcome. If individual J got liver cancer, and he had the unlucky gene, then the claim that J would not have gotten liver cancer but for the TCE exposure is true, even though we cannot know it until we can sequence his genome or find some other marker that correlates perfectly with the unlucky gene.

So which world are we in? Electrons may be truly indeterministic, but is cancer? Even if a gene exists which makes an individual vulnerable to TCE exposure, cancer

11 See SPIRTES ET AL., supra note 7, at 19-29.
requires a complicated series of genetic mutations and other developments, all of which can happen in a number of different ways, including insult from TCE, cosmic rays, a failure of the DNA repair mechanisms, etc. Perhaps quantum mechanical indeterminism does play a role in TCE exposure. Perhaps TCE interacts with some molecule in codon 61 of the H-ras protooncogene,\(^{12}\) the result being to move the probability of mutation in this gene slightly higher but leaving us with nothing, even in principle, that we could measure or observe about an individual of whom we could say that but for TCE, they would not have gotten liver cancer. I don't know, and right now I think it is safe to say that no one else does either.

If the question of whether individual level causal claims have a truth value depends on whether the world is pseudo-indeterministic or truly indeterministic, this is not so for general causal claims.

General causal claims involve the probability of the effect in a population that was exposed to the cause and the probability of the effect in the same population not exposed to the cause. These probability claims have a truth value regardless if the world is pseudo-indeterministic or truly indeterministic. Consider again the probability of liver cancer and TCE exposure, and suppose our general claim is that the probability of liver cancer is .02 if you are exposed to TCE, and .01 if not. In the pseudo-indeterministic world depicted in Figure 1, 10% of the population has an unlucky gene (gene L) that produces liver cancer always, and another, separate 10%+

---

of the population has a vulnerability gene (gene V) that gives you liver cancer whenever you are exposed to TCE (and fails to prevent the first gene from giving you cancer when you are not exposed to TCE).

Likewise, we can imagine a truly indeterministic world in which everyone’s propensity for getting liver cancer is moved from 10% to 20% with TCE exposure, but that there is nothing hidden to be discovered that will determine the outcome.

D. The Primacy of General Causal Claims

So individual causal claims have a truth value in a pseudo-indeterministic world, but upon what does their truth depend? Essentially, it depends on answering a counterfactual question of the form: would the ‘effect’ have failed to occur ‘but for’ the cause? Ignoring the many difficulties with this simple account of actual causation (such as overdetermination13 or pre-emption14), consider again what is required to evaluate such counterfactual claims. Consider two claims: (1) John Smith would not have gotten liver cancer but for exposure to TCE and (2) John Smith would not have gotten liver cancer but for wearing brown socks to work on Mondays. We must be able to assess whether John Smith’s life would have produced liver cancer in each of the two counterfactual worlds: (1) the world in which he lives his life exactly as before but is not exposed to TCE and (2) the world in which he lives his life exactly as before but does not wear brown socks to work on Mondays. Assessing whether he gets liver cancer in these worlds requires the general causal knowledge about how the world would have responded to such changes.

Once we decide, in our counterfactual world, exactly how to change John Smith’s circumstances, then the question of whether or not he gets liver cancer in this alternative world depends entirely on the causal laws we take to hold in all the possible worlds we consider. Given that the shift from wearing brown socks to black socks on Mondays is sufficiently minimal,

13 For example, when several soldiers in a firing squad shoot real bullets accurately, the prisoner’s death is overdetermined. For any individual soldier, it is false to say that the prisoner would not have died but for the soldier. We want each soldier to come out as a cause.

14 For example, when spy 1 pokes a hole in the canteen of an enemy about to cross a desert, he pre-empts the effect of spy 2, who had previously filled the canteen with poison. It is false to say that the enemy would not have died, but for either spy. We want spy 1 to be the cause and not spy 2.
then Smith still gets cancer because of the general causal claim: sock color has no causal influence on liver cancer.

Halpern and Pearl,15 Woodward,16 and others who have articulated clear accounts of individual level causal claims all require as input: (1) what happened in the real world, (2) precision about how the counterfactual world is to differ from the actual world as a result of removing or adding the “cause,” and (3) the general causal laws (usually called the structural equations) relevant to the events discussed. The moral is clear: we cannot assess the truth of individual level causal claims until we have the general causal laws relevant to the events at issue.

E. The Probability of Causation

If we know the general causal claims, that is, the risks of those exposed and of those not exposed, then we can turn to a weaker notion than truth for assessing our individual but-for causal claims. We can compute what is called the probability of causation (“PC”), a number that roughly corresponds to the probability that someone exposed would have avoided the disease had they not been exposed.17 The PC is based on what is called the attributable fraction of risk in a population (AF):

\[
AF = \frac{\text{Risk(exposed)} - \text{Risk(unexposed)}}{\text{Risk(exposed)}}
\]

\[
= \frac{P(\text{disease|exposed}) - P(\text{disease|unexposed})}{P(\text{disease|exposed})}
\]

For example, if the risk of liver cancer among TCE exposed workers is 2%, and would have been 1% had they not been exposed, then the AF = .5, so half of the liver cancers

---

observed in TCE exposed workers can be assumed to be the result of TCE exposure.

In many cases we have data on widely defined populations, for example, factory workers in Ohio, but no data on subpopulations, for example, workers 40 to 45 years in age, or workers who have a close ancestor with liver cancer, etc. In such cases it is typical to use the AF from the most narrowly defined and most informative population possible. For example, the risks for the Ohio factory workers might be 2% for exposed and 1% for unexposed, but among 40- to 45-year-old workers with a history of liver cancer in their family, the risks might be 6% for exposed and 5% for unexposed, making the AF = .06 – .05/.06 = 16.67%. So if Robert Jones is a non-descript Ohio factory worker who was exposed to TCE and got liver cancer, then the probability that his liver cancer was caused by TCE was 50%, but if Tim Lewis is a 43-year-old factory worker exposed to TCE with a father who had liver cancer, then the probability that his liver cancer was caused by TCE was much lower: 16.7%.

As to truth, the probability of causation is no help at all. Just as it is not true to say that a given coin flip came out heads because the coin was loaded 75% heads, it is not true to say that a given individual’s cancer was caused by TCE exposure because the probability of causation was 75%. A high probability of causation, or something like it, might be what the law must resort to in deciding torts and similar issues, but it should not be confused with assenting to the truth of a but-for claim.

III. EVIDENCE FOR GENERAL CAUSAL CLAIMS

So the truth or probability of an individual causal claim depends upon the general causal laws. It is to the problem of deciding on the general causal laws that I now turn. As I have already stressed, assessing general causal claims requires comparing a real population (e.g., Actual Population 1, Figure 2)

---

18 PC is objectionable for several reasons, not just that it falls short of truth. Sander Greenland and Jamie Robins have argued that a better measure of the plaintiff’s injury is an estimate of the years of life lost (“YLL”). Since YLL is also either something that might be determined by exposure, or only have its probability determined by exposure, I have no stake in this debate: all the issues regarding the truth of individual causal claims apply to YLL and PC equally. See Greenland & Robins, supra note 17, at 346.
Figure 2  Association versus causation.\textsuperscript{19}

with a counterfactual one (e.g., Counterfactual Population 1, Figure 2). Clearly we cannot go back in time and remove TCE or change the commercials children were exposed to. Can we compare workers exposed to TCE to other workers not exposed to TCE? Yes, but comparing an exposed population (Actual Population 1, Figure 2) to another population that we observe not to be exposed (Actual Population 2, Figure 2) reveals association, but not necessarily causation.

This is because an actual population of workers not exposed to TCE (Actual Population 2, Figure 2) might differ from the workers exposed to TCE (Actual Population 1, Figure 2) in other ways that affect cancer, such as diet, income, etc.

For example, it is nearly certain that children five to eight years of age who were in fact exposed to fewer than ten junk food commercials per day had a lower frequency of obesity in 2005 than children who were exposed to more than ten. But because it is also nearly certain that this group of children differs in other important ways from the population described in the claim, for example, their parents are more educated, wealthier, etc., this is \textit{not} the appropriate contrast class for causation. The appropriate contrast class is the \textit{same} group of

\textsuperscript{19} The expression $P(\text{Cancer} \mid \text{TCE})$ denotes the probability of cancer among those exposed to TCE. It also might be referred to as the conditional probability of cancer, given exposure to TCE. The expression $P(\text{Cancer} \mid \text{Set to TCE})$ would denote the probability of cancer among those upon whom we intervened to force exposure to TCE.
children, living the same life, but with the junk food commercials removed, or replaced with other sorts of commercials. In the first case, we are comparing children we observed being exposed to a lot of junk food commercials to children we observed being exposed to few commercials, and this comparison would undoubtedly reveal a statistical association between junk food and obesity. In the second case, we are comparing children we observed being exposed to a lot of junk food commercials to the same children after a hypothetical intervention in which we go back in time and change their exposure to commercials from a lot to a little. It is this comparison that reveals the causal relationship, but which is, in the deepest sense, unobservable.

As the population under a counterfactual, hypothetical intervention cannot be observed, how are scientists to gather evidence about this counterfactual population? This is the problem of causal inference.

A. Randomized Trials

Sir Ronald Fisher, the brilliant and prolific British statistician, provided in the 1930s what is still the gold standard today for causal inference: the randomized trial (“RT”). In its simplest form, an RT randomly splits a population into two subgroups (which we can expect on average to be identical), thus creating two versions of the same population, and then exposes one subpopulation to the cause (the “treated” group) and one to the absence of the cause (the “control” group). The frequency of the effect in the two groups provides evidence of the probability of the effect in the two populations we seek: one in which the cause is present, and an identical copy in which the cause is not present. Subtleties abound, but the basic strategy is sound and taught in every introductory research methods course.

The problem, of course, is that in a number of situations performing an RT is either ethically or practically impossible. We simply cannot intentionally expose half of a population to TCE and look for liver cancer.

There are essentially two recourses to an RT: (1) we can statistically adjust for naturally occurring differences in two populations, or (2) we can perform very small versions of RTs.

on animals we don’t seem to mind harming, for example, rodents.

B. Epidemiological Studies

Epidemiological studies involve observing human populations in which we do not control exposure to a cause and thus must resort to recourse (b): that is, epidemiological studies must statically adjust for naturally occurring differences in the exposed and nonexposed populations before they can claim evidence of causation. Statistical adjustment requires that we know all the relevant features upon which individuals differ besides being exposed to the cause or not. For example, a subpopulation that is exposed to TCE, for example, automotive factory workers who handle paint strippers that contain TCE, and a subpopulation that is not exposed to TCE, for example, workers on a chicken farm, may differ in more ways than just TCE exposure. The chicken farm workers may be different in age, have different diets, etc.

If we measure all the relevant differences, that is, those that also might cause liver cancer, then we can often adjust for these differences statistically and test for differences in liver cancer rates among the groups after this adjustment. If we do not know all the relevant differences, however, then this strategy fails. For example, if, unbeknownst to us, the autoworkers’ drinking water contains some other set of chemicals that cause liver cancer, while the farm workers’ water does not, and we don’t adjust for this, then our inference will be unsound.

A raft of other methodological issues confront epidemiologists, but the scientific evidence from such studies can in some instances be compelling, for example, cigarette smoking and lung cancer.

C. Toxicological and Animal Studies

In many cases, animals like rats or mice or rabbits or chimps share enough of human physiology to make it plausible to extrapolate from experiments with animals to what would happen in a similar experiment with humans. Biologists frequently perform controlled experiments on rodents to garner evidence to show whether some chemical causes cancer. They expose some rodents to a control, and others that are genetically identical and raised in the same environment to the
chemical of interest, and then compare the frequency of cancerous tumors. In some cases they can examine extremely detailed mechanisms by which the chemical might lead to a tumor by doing cell physiology on both the animals under study and human cells.

There are a number of informative and accessible discussions of the wide variety of evidence that can be used for causal inference. Although in my view the topic deserves dozens of books, it is out of my scope to say more here. In summary, scientists have long recognized that there are at least three distinct kinds of evidence that bear on the truth of general causal claims:

- Interventional studies on humans (e.g., RTs)
- Non-interventional (observational) studies on humans (e.g., Epidemiological studies)
- Mechanistic/toxicological evidence (e.g., animal and cell studies)

IV. COMBING THE EVIDENCE FOR CAUSATION

To come to a reasoned position on the status of a general causal claim, especially in a legal setting, we must (1) combine all the available evidence into a single judgment on whether the claim is true and (2) express the degree of our uncertainty about the claim.

In many cases, the evidence for a general causal claim is mixed. On some questions, there are RTs that show that a drug or treatment has a positive effect, others which show no effect, and still others which show a negative effect. As they are more complicated methodologically, epidemiological studies often present mixed evidence for a general causal claim. In many situations animal studies also show mixed results. Rationally combining multiple pieces of similar evidence, for

---


22 See, e.g., sources cited in supra note 20.

example, multiple RTs, is difficult but often feasible. The real challenge is to rationally combine different sources of scientific evidence into a single judgment about the truth of a general causal claim.

A. Meta-Analysis

For many general causal claims, for example, hormone therapy and breast cancer, there are often several distinct RTs in the published literature. Optimally combining the evidence from such studies is a topic of its own, called meta-analysis. A meta-analysis involves mathematically combining the results from multiple but comparable RTs to derive a summary estimate of the effect of some cause on some effect, often involving health, that appropriately combines the results of all the individual studies.\textsuperscript{24}

The technique is not limited to combining RTs, but can also be used to combine the evidence from several epidemiological (observational) studies, provided the populations studied are comparable. If the populations vary, then a related technique called meta-regression\textsuperscript{25} sometimes allows pooling of data in a principled way. Recent work by Eloise Kaizar\textsuperscript{26} improves on meta-regression when data from both RTs and observational studies are available on a similar general causal claim.

Meta-analytic methods are in general quite useful when there are multiple studies on the same causal claim. For example, the 2006 Institute of Medicine Committee on Asbestos did a quantitative meta-analysis on studies which individually estimated the effect of asbestos exposure on any of a number of different cancers, reporting a quantitative estimate that is a combination of the estimates from the

individual studies, as well as measures expressing the uncertainty of such estimates.27

No matter how sophisticated the meta-analytic technique, however, it is still limited to combining statistical evidence from different human studies into a single statistical estimate of the effect size, for example, the probability of the effect in the population exposed to the cause and the probability of the effect in the same population not exposed to the cause.

Meta-analysis of any form, however, cannot incorporate toxicological/mechanistic knowledge, nor can it easily factor in the quality of the studies being combined into a single estimate.

There is a technique, called the Bayesian approach,28 for combining all available evidence, including a scientist’s background knowledge, judgment, etc., into a single judgment about the nature and uncertainty of a general causal claim.

B. The Bayesian Approach

Several forests have been sacrificed explicating and debating the pros and cons of the Bayesian approach,29 so I will try to avoid piling on and provide only the briefest of sketches. There are many forms of the Bayesian approach, but the most appealing, in my view, is the most extreme. In this view, almost any statement—for example, “TCE causes liver cancer” or “your next child will be born with blond hair” or “the moon was formed by a collision between a proto-planet and the Earth”—can be assigned a credence, or degree of belief between 0 and 1, and the degrees of belief can be interpreted as probabilities. In some cases the probability can be assigned objectively, for example, the objective probability of your next child being born with blond hair, or any of a number of heritable diseases, can be worked out by a genetics counselor

28 Named after the Reverend Thomas Bayes, who lived during the first half of the eighteenth century.
with appropriate access to your family history and perhaps some of your blood. In other cases the probability corresponds to nothing more than a subjective degree of belief. For example, having little or no evidence to go on, in 2007 I may assign a probability of .2 to the statement, “Blu-Ray will win the format wars over HD-DVD for the next generation of DVDs.”

The approach provides a principled way to compute the probability one should assign to a hypothesis \( H \), after you have seen a new piece of evidence \( E \), notated as \( P(H \mid E) \). The fundamental theorem which drives the approach is extremely simple to state and prove:

\[
P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}
\]

The numerator on the right involves \( P(E \mid H) \), called the likelihood since it represents the probability of the evidence \( E \) given \( H \) is true, and \( P(H) \), called the prior, the probability assigned to \( H \) prior to seeing the evidence. The denominator, \( P(E) \), is the probability of the evidence without any consideration of the hypothesis \( H \). The target, \( P(H \mid E) \), is called the posterior as it reflects the probability of \( H \) after seeing the evidence \( E \).

A classic use of the formula is in computing the probability of having a disease, given a diagnostic test result. For example, suppose a 20-year-old upper middle class heterosexual male Jim gets a blood test for HIV, and it comes out positive. Jim is scared, but what is the probability of \( H \): that he is actually infected with HIV, given the evidence from the blood test \( E \)? First suppose that the test is 98% reliable. That is, suppose that the probability of the test coming out positive given HIV infection, \( P(E \mid H) \), is .98. Now Jim is truly terrified. Next suppose that \( P(H) \), the prior probability of a 20-year-old upper middle class heterosexual male having HIV, is 1 in 1000 (.001). Finally, suppose that \( P(E) \), the probability of a blood test coming out positive is 1 in 125 (.008). Then, to Jim’s relief, the posterior probability of HIV = .1225:

\[
P(H \mid E) = \frac{P(E \mid H) = .98 \ast P(H) = .001}{P(E) = .008} = .1225
\]

---

30 Also referred to as the probability of \( H \) conditional on \( E \), or \( H \) given \( E \).
The apparent discrepancy between the 98% reliability of the HIV test and posterior probability of HIV of 12.25% is due to the low prior probability of HIV in Jim’s cohort, and the frequency of positive blood tests. If one views the situation as a Bayesian, however, the evidence did make a big difference. Before taking the test, the probability of Jim having HIV was 1 in 1000. After the positive test, the probability moved to about 1 in 8, a huge jump.

The Bayesian approach has been used (and misused) in assessing forensic evidence like DNA testing in courtrooms for well over a decade.\(^3\)

In our context, the potential utility of the approach should be apparent. Beginning with some prior belief over a general causal claim \(H_c\), for example, that TCE causes liver cancer, then for each new piece of evidence, in the form of an RT, or epidemiological study, toxicological study, or what have you, one can use the Bayesian approach to compute the probability that \(H_c\) is true given this evidence. When all the evidence is in, we emerge with a posterior probability of \(H_c\), that is, the probability of \(H_c\) given all the evidence available.

In practice, the Bayesian approach is far from a panacea. In typical scientific contexts involving a community of scientists, it is very difficult to move from inchoate and diverse sorts of background knowledge to a prior, that is, a degree of belief in the causal claim prior to reviewing the evidence. Further, while “updating” to a posterior from certain kinds of evidence is reasonably straightforward, updating from other kinds of evidence is not. For example, consider computing the probability of liver cancer as a function of TCE exposure. After a few studies estimating the dependence of liver cancer risk on TCE concentration in work environment air, we can use the Bayesian approach to incorporate a new sample of 400 factory workers who were exposed to air with varying amounts of TCE concentrations for ten years, for example. As the probability of liver cancer as a function of TCE exposure is the hypothesis under study, the likelihood in Bayes’ formula is objective. That is, the probability of seeing a particular frequency of liver cancers given that TCE does cause liver cancer, is objectively, mathematically derivable, and from that we can apply Bayes’ theorem.

\(^3\) See, e.g., Joseph B. Kadane, Misuse of Bayesian Statistics in Court, CHANCE, Spring 2006, at 38.
What we cannot do, at least in any objective way, is to use the approach to update on evidence that shows rats exposed to 50 ppb TCE get tumors at three times the rate of those exposed to 1 ppb. What is the likelihood of this evidence, assuming TCE does cause liver cancer in humans? Here we are often beyond mathematics and statistics and into opinion about the comparability of rats and humans. Other features of studies, for example, the quality of measures employed in an epidemiological survey, are also extremely hard to incorporate in any objective way into a Bayesian analysis. In general, this is referred to as the “objectivity of the likelihood” problem. Besides the sheer difficulty in performing the appropriate computations, the approach is a regulative ideal, but it is still far from the practical device we want to get us to a rational judgment on the truth of a general causal claim.

Finally, the probability of a hypothesis does not correspond in any simple way to the chances of it being true in the world. The justification for the technique is decision-theoretic, and is based more on a theory of rationality than it is on correspondence to the truth.

This is not to say that the technique is hopeless. It isn’t. An actual example of using it for assessing a general causal claim relevant to the law comes from the National Academy of Science’s (“NAS”) BEIR IV report, which sought to estimate the carcinogenicity of plutonium in humans. By assuming that the ratio of carcinogenic potencies of plutonium to various other radionuclides like radium would be roughly constant across species, the Committee managed to combine very limited human data involving plutonium, with extensive animal data on plutonium and radium, and more extensive human data on radium, to emerge with a posterior over the hypothesis concerning the carcinogenicity of plutonium in humans.

32 This is not always the case. See William H. DuMouchel & Jeffrey E. Harris, Bayes Methods for Combining the Results of Cancer Studies in Humans and Other Species, 78 J. AM. STATISTICAL ASS’N 293 (1983).
33 For a recent philosophical discussion of the epistemological view of the Bayesian approach, see Luc BOVENS & Stephan HARTMANN, BAYESIAN EPISTEMOLOGY (2003).
Although I have been focusing on the problems of combining different sorts of evidence and ignoring the issue of expressing the uncertainty over causal claims, this comes for free with the Bayesian approach. For simple propositions, like R: “it will rain on Labor Day 2010 in Brooklyn,” the posterior will be a probability between 0 and 1. If, in scientist A’s posterior R has a probability of .93, and in scientist B’s posterior R has a probability of .55, then A is in some sense more certain about R than B. Both scientists are more confident that R is true than that R is false—so they both in some sense believe R is true—but their degree of uncertainty is not the same. They should both take an even bet on R, but A would give much longer odds than B.

For more complicated hypotheses (for example, “it will rain $x$ inches on Labor Day 2010”), where we are asked to put a probability over each possible value of $x$, then the posterior is not a number but a probability distribution. For example, scientist A might have a posterior like the left side of Figure 3, while scientist B has a posterior like the right side of Figure 3. Roughly, the height of the graph corresponds to how much probability the scientist distributes over that number of inches of rain. The rectangle over 0 represents the probability of no rain (0 inches), which is .07 for Scientist A and .45 for Scientist B, while the rest of the posterior is distributed over rain from 0-1.2 inches on that day. If it does rain, then Scientist A seems to put the most probability over around .3 inches, while Scientist B, although he is less confident that it will rain, if it does, he deems it most probable to rain a little over .4 inches.

We can do the same with causal claims regarding the relative risk of liver cancer after 10 years of exposure to 10 ppb TCE exposure compared to no TCE exposure. A relative risk of 1.0 means that TCE has no effect on liver cancer. A relative

![Figure 3](image-url)
risk of 2.0 means that a person exposed to 10 ppb for 10 years has twice the probability of getting liver cancer as someone not exposed. In Figure 4, Scientist A believes that overall there is an effect of TCE, but it is relatively small, that is, a relative risk of around 1.3. Although her assessment of the size of the effect is small, she is quite confident that the relative risk is close to 1.3 as her posterior is narrowly distributed around 1.3. Scientist B seems to put the most probability around a larger effect, a relative risk of 2.0, but as her posterior is much wider and more diffuse, she is more uncertain about the size of the causal effect than is Scientist A.

So the Bayesian approach, although imperfect in many ways and practically always a challenge to apply, provides one way to synthesize the evidence and to express uncertainty about general causal claims that might be appealing to the law.

V. WHAT’S THE LAW TO DO?

The final question to tackle is the hardest: what’s the law to do? Currently, the courts deal with complicated matters of causation in something like the following way. The judge must act as gatekeeper and decide which experts will be allowed to testify as to the scientific case for or against the general causal claim. The plaintiffs then mount a case by summoning the experts (whom the judge allowed) to argue to the jury that the scientific evidence for the general and the individual causal claim is compelling. The defense then summons their own experts, who argue that the scientific evidence...

---

35 This is true since *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 592-93 (1993).
Evidence is not compelling, sometimes by impeaching the credibility of the plaintiff's experts, sometimes by emphasizing alternative evidence, etc.

Neither the judge nor the jury is trained to synthesize a diverse body of complicated scientific evidence, especially evidence presented by highly partial, highly sophistical experts. A better system would have the community of scientists—who are presumably both less invested in the outcome and more qualified to rationally assess a wide body of complicated evidence—come to consensus as to the truth of the general causal claims at issue as well as the scientific uncertainty around these claims. This output from the scientific community could then be used as input to the legal system. This would not preclude plaintiffs or defendants from mounting their own experts and cases, but it would give the judge and jury a perspective on the science to fall back to when they are overwhelmed by the briefs submitted or the pyrotechnics in the courtroom.

Such a system is in fact already in place and used widely. For example, in decisions as to whether to compensate Vietnam veterans who were exposed to Agent Orange and now have some illness (like liver cancer), the Veteran’s Administration does not restrict itself to hearings involving experts from both sides; rather, it consults a bi-annual report on the general causal claims true of Agent Orange exposure produced by a distinguished panel of independent scientists retained by the Institute of Medicine, a branch of the National Academies of Science.36

The International Agency for Research on Cancer (“IARC”),37 the U.S. Environmental Protection Agency,38 the Centers for Disease Control and Prevention,39 the National Institutes of Health,40 and the National Toxicology Program41

---

39 Centers for Disease Control and Prevention, supra note 21.
have all developed systems for retaining panels of scientists to assess the status of general causal claims and to classify the level of evidence in support of such claims, particularly where such claims regard chemicals and cancer.

These panels rarely do a full Bayesian analysis to synthesize all the evidence, although as I illustrated with the NAS’s BEIR IV report, it is not unknown. Almost universally, however, they assign the causal claim to one of four or five discrete categories which differ on the truth of the claim and on the scientific uncertainty surrounding the claim.42

For example, IARC forms committees consisting of biologists, epidemiologists, and toxicologists. They instruct these scientists to first categorize the level of evidence within three subcategories—human, animal, and mechanistic—and then to synthesize the subcategories of evidence into an overall evaluation on a five-category scale ranging from carcinogenic to probably not carcinogenic.43 Figure 5 depicts the IARC scheme.

Each of the categories has natural counterparts to a Bayesian posterior, as explicitly described in the Institute of Medicine’s 2007 report, “Improving the Presumptive Disability

---

41 NATIONAL TOXICOLOGY PROGRAM (NTP), 11TH REPORT ON CARCINOGENS (U.S. Dep’t of Health and Human Servs. 2005).
42 They of course accompany this categorization with a long and inaccessible report.
43 See IARC MONOGRAPH, supra note 37, at 22-23.
Decision-Making Process.” 44 This sort of categorical output is accessible and comprehensible to a judge and to a jury, and the steps to reach a consensus on the output are much better handled by relatively impartial scientists than by jurors trying to weigh evidence described to them by highly partial and well-paid experts.

The output of these panels regards the general causal claims that must be invoked in a legal case in which an individual causal claim is at stake, but the court must still decide the individual causal claim. In calculating the probability of causation or the years of life lost, for example, the plaintiff must appeal to the general causal laws connecting the purported exposure and the injury it allegedly caused, but the court must still decide whether the estimate of PC of YLL is high enough, even allowing for uncertainty in this estimate, to warrant liability. This seems like more than enough complexity for juries to be asked to handle.

44 Disability Decision-Making, supra note 17.
Of Truth, in Science and in Law

Susan Haack

When Questions of Science Come to the Courtroom, Truth Has Many Faces

—headline in the New York Times

I. INTRODUCTION: FACTUAL TRUTH, SUBSTANTIAL JUSTICE, AND THE LAW OF EVIDENCE

In 1966, the United States Supreme Court averred that “[t]he basic purpose of a trial is the determination of truth.”

In 1993, in the landmark ruling in Daubert v. Merrell Dow Pharmaceuticals, Inc., that set new standards for the admissibility of expert scientific testimony, Justice Blackmun was a bit more cautious, writing that “there are important differences between the quest for truth in the courtroom and the quest for truth in the laboratory.”

In my view, however, a trial—even a trial at which the main issue is a matter of fact, such as whether it was the defendant or someone else who pulled the trigger, or whether it was the drug in question or something else that caused the injury—"isn’t exactly a "search for truth." Rather, a trial is better described as a late stage of a whole process of determining a defendant’s guilt or liability: the stage at which, under the legal guidance of the court, advocates for each side present evidence in the light most favorable to their case, and

---

† Distinguished Professor in the Humanities, Cooper Senior Scholar in Arts and Sciences, Professor of Philosophy, and Professor of Law, University of Miami. My thanks to Mark Migotti for helpful comments on a draft, and to Barbara Brandon and Mark Plotkin of the University of Miami Law Library for help in tracking down relevant materials. I also learned a good deal from discussions at a workshop held at Albany Law School in December 2006 on a forthcoming book by Wendy Wagner and Thomas McGarity (BENDING SCIENCE (forthcoming 2008)).


the finder of fact sifts through it and assesses whether it establishes guilt or liability to the required degree of proof.4

Qua litigators, the attorneys for parties at trial are primarily engaged, not in inquiry, but in advocacy. And the whole process is constrained by legal desiderata which are not directly, or even not at all, truth-related: that defendants’ constitutional rights not be violated,5 that certain policy considerations (such as not discouraging landlords from making repairs that might prevent accidents)6 be served, and so on. Relevant evidence is thus sometimes excluded for reasons that have nothing to do with truth. Moreover, legal proceedings operate under a kind of time constraint alien to the search for truth in physics, history, and the rest. To quote Justice Blackmun again, the legal system seeks “quick, final and binding . . . judgments”;7 the desideratum of promptness imposes time constraints at one end of the process, the desideratum of finality-and-bindingness at the other.

Nevertheless, truth is surely relevant to legal proceedings, for we want, not simply resolutions, but just resolutions; and substantial justice requires factual truth.8 In its efforts to arrive at factually correct verdicts, the legal system has come to rely a good deal on scientific experts, who by now testify on just about every scientific, and quasi-scientific, subject imaginable: experts on blood, bullets, bite-

---

5 As the Supreme Court recognized in Tehan: “By contrast, the Fifth Amendment privilege against self-incrimination is not an adjunct to the ascertainment of truth. . . . [but] stands as a protection of quite different constitutional values.” Tehan, 382 U.S. at 416.
6 FED. R. EVID. 407 (providing that “evidence of the subsequent [remedial] measures” is (normally) inadmissible; e.g., that the landlord subsequently fixed the steps on which the plaintiff fell and broke her leg cannot be introduced in arguing that the landlord’s negligence makes him liable for the injury).
7 Daubert, 509 U.S. at 597.
8 This principle has sometimes been deliberately compromised; the classic instance is Summers v. Tice, 199 P.2d 1, 3 (Cal. 1948) (holding that “where a group of persons are . . . engaged in the use of firearms, and two of them are negligent in firing in the direction of a third person who is injured thereby, both . . . are liable for the injury . . . although the negligence of only one of them could have caused [it]”). See also the DES (diethylstilbestrol) cases: Sindell v. Abbott Laboratories, 607 P.2d 924, 937 (Cal. 1980) (holding that “[e]ach defendant will be held liable for the proportion of the judgment represented by its share of that market unless it demonstrates that it could not have made the product which caused plaintiff’s injuries”), and, most striking of all, Hymowitz v. Eli Lilly & Co., 539 N.E.2d 1069, 1078 (1989) (holding that “there should be no exculpation of a defendant who, although a member of the market producing DES for pregnancy use, appears not to have caused a particular plaintiff’s injury”).
marks, battered wives; experts on PCBs,\textsuperscript{9} paternity, poisons, post-traumatic stress; experts on radon, recovered memories, rape trauma syndrome, random-match probabilities; experts on psychosis, asbestosis, silicosis (and for all I know, on psittacosis)!\textsuperscript{10} There’s no question that the testimony of scientific experts is often essential, or that it can be enormously useful; but there’s also little question that these interactions between science and the legal system can be quite problematic.

The announcement of this conference invites participants to consider “if and how [the inquiry into the reliability of proffered scientific testimony mandated by \textit{Daubert}] relates to ‘truth’ and whose view of the truth should prevail.” I’m not entirely sure what to make of this, especially of those scare quotes around “truth.” Could they be an expression of skepticism, in the fashion of the post-modernists, about the legitimacy of the very concept of truth?\textsuperscript{11} Or are they perhaps an implicit endorsement of the strong forms of social constructivism fashionable among some practitioners of Sociology of Scientific Knowledge (“SSK”), Science and Technology Studies (“STS”), “ethnomethodology of science,” and other related fields?\textsuperscript{12} Or might they be just a way of conveying the modest-enough idea that certain scientific claims come to be thought of as true, not as a result of scientific investigation, but by becoming entrenched in legal decisions?

In any case, I hope the account of truth in the sciences offered here will be on point. It will combine a full acknowledgment of the fallibility and incompleteness of the scientific enterprise with a robustly objective conception of


\textsuperscript{12} As this suggests, there are stronger and weaker forms of social constructivism, some (in my opinion) true but not very exciting, others exciting but false. They are distinguished and disentangled in \textit{Susan Haack, Defending Science—Within Reason: Between Scientism and Cynicism} 190-91 (2003).
truth. It will, in consequence, be distinctly inhospitable both to fashionable forms of skepticism about the very idea of truth, and to strong forms of social constructivism about the sciences. But its combination of fallibilism and objectivity will help us understand why the legal system often gets less than the best out of science, shed some light on the vagaries of the legal use of the word “reliable,” and clarify the process by which scientific claims, true or false, can become legally entrenched as (supposedly) reliable science.

The first step (Section II) is to sketch the legal history leading up to Daubert, and to explore some of the difficulties Daubert brought in its wake; the next (Section III) to develop an account of truth in the sciences; then (Section IV) to articulate why the legal system so often fails to get the best scientific information available; and finally (Section V) to show how false scientific claims sometimes get legally entrenched as reliable science.

II. A LEGAL TANGLE: FROM FRYE TO DAUBERT TO KUMHO TIRE TO REVISED FEDERAL RULE OF EVIDENCE 702

Ever since the legal system began to call on scientific witnesses, there have been complaints both from legal commentators and from scientists themselves about the defects and drawbacks of the process. Judges, attorneys, and legal scholars complain about the venality of scientists willing to testify to just about anything for money, about the naiveté of scientists who apparently don’t think concerns about conflict of interest apply to them,¹³ and so forth; scientists complain about judges’, jurors’, and attorneys’ ignorance and credulity on scientific matters, about the professional insult of being “Dauberted” or “dauberted out,”¹⁴ and so on. And while for

¹³ See, e.g., Winans v. N.Y. & Erie R.R. Co., 62 U.S. 88, 101 (1858) (“Experience has shown that opposite opinions of persons professing to be experts may be obtained to any amount.”); Lee M. Friedman, Expert Testimony, Its Abuse and Reformation, 19 YALE L.J. 247, 247 (1910) (“[T]here is a constant complaining and mistrust on the part of judges, juries and lawyers of the expert witness.”).

many decades now, from Frye (1923)\textsuperscript{15} to the Federal Rules of Evidence (“FRE”) (1975) to Daubert (1993)\textsuperscript{16} and beyond, there have been legal efforts to domesticate scientific testimony by means of rules of evidence, these efforts have hardly been an unqualified success.

The Supreme Court’s ruling in Daubert, with its many ambiguities and confusions, is emblematic of the difficulties. The core of the Daubert ruling is that FRE 702, according to which the testimony of a qualified expert is admissible provided it is relevant and not otherwise legally excluded, supersedes the old Frye Rule,\textsuperscript{17} according to which novel scientific testimony is admissible only if it is “sufficiently established to be generally accepted in the field to which it belongs.”\textsuperscript{18} But the Daubert ruling goes on to explain that courts should screen proffered expert testimony for reliability as well as relevance.\textsuperscript{19} Moreover, the intended meaning of “reliable,” in the specialized, “evidentiary” sense Justice Blackmun gives it, and its relation to our ordinary notions of reliability, remains far from transparent.

Courts are to assess reliability, Justice Blackmun continues, by determining whether proffered expert evidence is genuinely “scientific . . . knowledge,”\textsuperscript{20} and this determination is to be focused on experts’ methodology, not their conclusions. He seems to run “reliable” together with “scientific,” and “genuinely scientific” with “conducted in accordance with the scientific method” (elisions perhaps aided and abetted by his conflation of Karl Popper’s and Carl Hempel’s incompatible philosophies of science).\textsuperscript{21} The reference to “peer review and publication” in his list of indicia of reliability veers unsteadily between suggesting that courts determine whether the science

\textsuperscript{15} Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).
\textsuperscript{17} Id. at 589.
\textsuperscript{18} Frye, 293 F. at 1014.
\textsuperscript{19} Daubert, 509 U.S. at 589.
\textsuperscript{20} Daubert, 509 U.S. at 590, 590 n.8. This is the Court’s interpretation of the 1975 Federal Rule of Evidence 702; but as the ellipses indicate, Justice Blackmun omits several words from the text of the Rule, which spoke of “scientific, technical, or other specialized knowledge.” FED. R. EVID. 702 (1975).
\textsuperscript{21} For more detailed discussion, see Susan Haack, \textit{Trial and Error: The Supreme Court’s Philosophy of Science}, 95 AM. J. PUB. HEALTH S66 (2005), reprinted in 41 INT’L SOC’Y BARRISTERS Q. 376 (2006), and HAACK, PUTTING PHILOSOPHY TO WORK, supra note 11.
on which proffered testimony is based has survived the pre-publication peer-review process of the scientific journals, and suggesting that courts determine whether it has survived (or will survive?) the long-run scrutiny of the relevant scientific community. And so on.

Moreover, while the ostensible intent of the Daubert ruling was to relax the “austere standard” of the older Frye rule in accordance with FRE 702—a point that states that decided to stick with Frye sometimes emphasize—it is far from clear that this has really been its effect. Though a 2001 study from the RAND Institute for Civil Justice indicated that the full story may be more complicated than this, commentators often suggest that on the whole, at least in civil cases, Daubert has made it harder, not easier, to get scientific testimony admitted.

Since Daubert, the Supreme Court has twice returned to the question of scientific testimony, first in Joiner (1997) and

---

22 For more detailed discussion, see Susan Haack, Peer Review and Publication: Lessons for Lawyers, 36 STETSON L. REV. 789 (2007).
23 E.g., People v. Leahy, 882 P.2d 321, 331 (Cal. 1994) (declining to adopt Daubert on the grounds that “Kelly sets forth the various reasons why the more ‘conservative’ Frye approach . . . [is] an appropriate one”) (referring to People v. Kelly, 549 P.2d 1240 (Cal. 1976)); Brim v. State, 695 So. 2d 268, 271-72 (Fla. 1997) (“Despite the federal adoption of a more lenient standard in Daubert v. Merrell Dow Pharmaceuticals, Inc. we have maintained the higher standard of reliability as dictated by Frye.” (citation and footnote omitted)); Blum v. Merrell Dow Pharm., Inc., 764 A.2d 1, 3 (Pa. 2000) (while the Blums’ expert testimony was arguably admissible under “the somewhat less exacting standard of Daubert,” it was inadmissible under the Frye Rule); Olivier A. v. Christina A., No. 30779/2002, 2005 WL 2171176, at *24 (N.Y. Sup. Ct. Aug. 25, 2005) (“The current controversy seeks to use Daubert to restrict and invalidate prior types of admissible evidence rather than expand what is admissible. This is inapposite . . . .”).
24 The study, based on data from 399 federal district court opinions in civil cases between 1980 and 1999, found that in the years immediately following Daubert judges scrutinized reliability more carefully, applied stricter standards, excluded more evidence, and issued more summary judgments; but that after about 1996 the rate of successful challenges began to fall, in part, they conjecture, because the parties responded to the changes in standards. LLOYD DIXON & BRIAN GILL, CHANGES IN THE STANDARDS FOR ADMITTING EXPERT EVIDENCE IN FEDERAL CIVIL CASES SINCE THE DAUBERT DECISION 7-8, 22-27 (RAND Inst. for Civil Justice, 2001).
then in *Kumho Tire* (1999).\(^{27}\) In these later rulings, the Court seems to have backed quietly away from some of the *Daubert* Court’s more ambitious philosophical observations: Justice Rehnquist’s ruling in *Joiner* suggests that the distinction between methodology and conclusions stressed in *Daubert* may be less than robust;\(^ {28}\) Justice Breyer’s ruling in *Kumho Tire* insists that when FRE 702 refers to “scientific or other technical knowledge” the key word is “knowledge,” not “scientific.”\(^ {29}\) But these later rulings hardly solved all the old problems; and they introduced new problems of their own. *Joiner*, confirming that evidentiary rulings under *Daubert* are reviewable only for abuse of discretion,\(^ {30}\) brings the possibility to the fore that different courts in the same jurisdiction may rule inconsistently on the admissibility of the very same evidence. *Kumho Tire*, confirming that *Daubert* applies to non-scientific as well as to scientific expert testimony, and that courts may use any, all, or none of the *Daubert* factors, or such other factors of their own devising as they deem appropriate,\(^ {31}\) leaves courts with startlingly little substantive guidance.

In December 2000, FRE 702 was modified to make courts’ obligation to screen proffered expert testimony for reliability explicit: to be admissible, expert testimony must be based on “sufficient” data, the result of “reliable” methods, “reliably [applied] to the facts of the case.”\(^ {32}\) With its stress on the need for courts to determine the reliability both of the underlying data and of its application to the case at issue, this revision went somewhat beyond simply articulating what, according to the *Daubert* Court, was already implicit in the original Rule 702; and some courts have understood the revised Rule as having tightened the *Daubert* standard.\(^ {33}\) Still, however often the word “reliable” and its cognates are repeated, it is

---


\(^{28}\) *Joiner*, 522 U.S. at 146.

\(^{29}\) *Kumho*, 526 U.S. at 147. Surprisingly, Justice Breyer suggests that this was already clear in *Daubert*. *Id.* at 147-48.

\(^{30}\) *Joiner*, 522 U.S. at 143 (“The question of admissibility of expert testimony . . . is reviewable under the abuse of discretion standard.”).

\(^{31}\) *Kumho*, 526 U.S. at 141 (“*Daubert’s* general holding . . . applies not only to testimony based on ‘scientific’ knowledge, but also to testimony based on ‘technical’ and ‘other specialized’ knowledge . . . But . . . *Daubert’s* list of specific factors neither necessarily nor exclusively applies to all experts or in every case.”).


\(^{33}\) See, e.g., *Rudd v. General Motors Corp.*, 127 F. Supp. 2d 1330, 1336 (M.D. Ala. 2001) (“The new Rule 702 appears to require a trial judge to make an evaluation that delves more into the facts than was recommended in *Daubert* . . . .”).
hard to see how any such verbal formula could, by itself, enable courts to discriminate genuinely reliable testimony from the unreliable stuff.

Even after all these legal efforts, I suspect, no one believes that the legal system always gets the most or the best scientific information that, ideally, it might. This, however, presupposes that there are more and less reliable scientific claims, and that scientific claims are (normally) either objectively true or else objectively false. But this presupposition stands in need of more detailed articulation and more careful argument.

III. TRUTH IN SCIENCE: THE CRITICAL COMMON-SENSIST PERSPECTIVE

The word “truth” is sometimes used as an abstract noun, referring to the concept of truth (or as some would prefer to say, to the property of being true, or to the meaning of the word “true” and its synonyms in other languages). The word is often also used, however, to refer to the particular propositions, etc., which are true; in this use it takes the plural form, as in: “we hold these truths to be self-evident . . . .”

People also write (or, “drawing” the quotation marks in the air with their fingers, speak) of “‘truth,’” and “‘truths.’” In this use, quotation marks signal doubts, reservations, or outright skepticism about the propriety of the word or phrase they enclose, and are known in the jargon of philosophers as scare quotes. Their effect is, roughly, to turn a term meaning “X” into a term meaning “so-called ‘X’.” So “truths” means “so-called ‘truths’” or “purported truths,” that is, propositions

---

34 “Critical common-sensism” (a phrase borrowed from C.S. Peirce) is my name for the philosophy of science offered in HAACK, DEFENDING SCIENCE—WITHIN REASON, supra note 12, on which I shall draw throughout this section.

35 Though I write here of “the property of being true,” I do not mean to foreclose questions in the philosophy of logic about whether “true” really is a predicate, really does represent a property, or whether it is, as Frank Ramsey suggested, something more like a sentential quantifier. See generally FRANK PLUMPTON RAMSEY, ON TRUTH: ORIGINAL MANUSCRIPT MATERIALS (1927-1929) FROM THE RAMSEY COLLECTION AT THE UNIVERSITY OF PITTSBURGH (Nicholas Rescher & Ulrich Majer eds., 1991); María-José Frápolli, The Logical Enquiry into Truth, 17 HIST. & PHIL. LOGIC 179 (1996); Susan Haack, The Unity of Truth and the Plurality of Truths, supra note 11, at 90-91.

36 Though I write here of the truth (and falsity) of propositions (and later of the truth (and falsity) of claims, theories, etc.), I do not mean to foreclose issues about the precise nature of the truth-bearers, a topic long debated in philosophy of logic. See SUSAN HAACK, PHILOSOPHY OF LOGICS 74-85 (1978).
which are *taken to be or presented as* truths; but conveys the suggestion that they may not really be true. And “truth” means “so-called ‘truth,’” that is, the *supposed* concept of truth or property of being true; but conveys the suggestion that the concept may not really be a legitimate one.

This much, though seldom spelled out explicitly,\textsuperscript{37} is not very controversial; now we get to the controversial part, for as the philosopher Nicholas Rescher once observed, “If two people agree, one of them isn’t a philosopher.”\textsuperscript{38} What I offer here is not “the philosophical perspective” on truth in the sciences; it is *my* philosophical perspective on truth in the sciences. But it is not (as I see it!) just one perspective on an equal footing with any other; for that it is my perspective means precisely that this is the conception I believe to be correct. I have no reservations about the legitimacy of the concept of truth, and so, where appropriate, will speak without apology of truth, not “truth.” However, I also have no doubt that propositions are often taken for true which are in fact false, that is, are only “truths,” not truths.

I should add that what I offer here is not an account of *scientific* truth, if that is taken to imply that “true” has a special, distinctive meaning when applied to scientific propositions. My observations about the concept of truth are intended to apply to the truth not only of claims and theories in the sciences, but also to the truth of propositions of other kinds.

As I see it, the terms “science” and “the sciences” are best construed as referring to a loose federation of kinds of inquiry, roughly characterizable by the questions within their scope. A large range of factual questions, including questions about the consequences of putting this or that policy into effect or about the risks and benefits of a drug, fall within the scope of one (or sometimes more than one) of the various, ramifying branches of science; but questions about whether this or that policy is preferable, or whether the risks outweigh the benefits, are not themselves scientific questions. This is not to suggest that the distinction is always perfectly clear; nor, more generally, that there is a clean, sharp line dividing scientific

\textsuperscript{37} The honorable exception is the now-classic critique of the deceptive power of scare quotes in D.C. STOVE, POPPER AND AFTER: FOUR MODERN IRRATIONALISTS (1982); reprinted under the title ANYTHING GOES: ORIGINS OF THE CULT OF SCIENTIFIC IRRATIONALISM (1998).

questions from others—cosmological from metaphysical questions, for example, or questions in theoretical psychology from questions in philosophy of mind. 39 Nor is it to deny that, when scientific inquiry tackles old questions, it almost always raises new ones, and sometimes leads to the conclusion that an older question has no true or false answer, but is flawed by false presuppositions.

The goal of scientific inquiry, as of any kind of inquiry, is to discover the answer to some question or questions; the true answer, that is. 40 This is not to suggest that scientists seek THE TRUTH, in some quasi-religious sense (the sense, whatever it is, of Jesus’ claim that he is “the Way, the Truth, and the Light”); it is not to suggest that scientists collect true propositions, as some people collect rare stamps or antique furniture; and it is not to intimate that scientific truths are the only truths there are. But it is to say that the goal of an investigation into the structure of DNA is to reach the answer that DNA is a double-helical, backbone-out macromolecule with like-with-unlike base pairs if DNA is a double-helical, backbone-out macromolecule with like-with-unlike base pairs, or the answer that DNA is a triple-helical, backbone-in macromolecule with like-with-like base pairs if DNA is a triple-helical, backbone-in macromolecule with like-with-like base pairs, . . ., or that it’s more complicated than that if it is more complicated than that; . . . and so on, mutatis mutandis, for other questions.

This is fully in accordance with Aristotle’s dictum that “[t]o say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true”; 41 and with Frank Ramsey’s laconic observation that “[a] belief that p . . . is true if and only if p; for instance, a

39 See also Susan Haack, Not Cynicism, But Synechism: Lessons from Classical Pragmatism, 41 TRANSACTIONS CHARLES S. PEIRCE SOC’Y 239 (2005), reprinted in COMPANION TO PRAGMATISM 141 (John Shook & Joseph Margolis eds., 2006) and in HAACK, PUTTING PHILOSOPHY TO WORK, supra note 11.

40 As this suggests, I accept neither an Instrumentalist view (according to which theoretical “statements” in science are not really genuine statements, and hence are neither true nor false), nor a Constructive Empiricist view (according to which, though theoretical statements are statements, and do have truth-values, the goal of science is empirical (observational) adequacy, not truth). I cannot argue either point here, but they are discussed in HAACK, DEFENDING SCIENCE—WITHIN REASON, supra note 12, at 137–41.

belief that Smith is either a liar or a fool is true if Smith is either a liar or a fool and not otherwise.\textsuperscript{42} (This is “merely a truism,” Ramsey continues, needing to be stressed only because “there is no platitude so obvious that eminent philosophers have not denied it.”\textsuperscript{43}) It is also enough to tell us that whether a scientific claim or theory is true or is false is normally an objective matter; that is, it is neither necessary nor sufficient for a proposition’s being true that you, or I, or anyone believe it.\textsuperscript{44}

To be sure, scientific claims and theories come into being as the result of scientists’ intellectual work; so in one sense scientists might be said to make scientific truths. But it is not scientists’ intellectual work, but the nature of the phenomena and events in the world that those claims and theories describe, that makes those scientific truths true. Natural-scientific claims and theories are about natural phenomena and events, and so (except in special cases involving phenomena and events that occur only in the laboratory) are about things not of our making; and while social-scientific claims and theories are about social phenomena, institutions, roles, rules, etc., which are of our making, whether those claims and theories are true or are false is still independent of whether you, or I, or anyone, thinks they are.

So scientific claims and theories are (normally) either true or else false, and their truth or falsity is (normally) an objective matter. The objectivity of scientists, however, is another question, for the meaning of “objective” shifts somewhat when it is applied to persons. To say that not all scientists are objective means, first, that not all scientists are unemotional, stolid types; some are deeply engaged with and excited by their work. And second, most to the present purpose, it means that not all scientists are unbiased. A few scientists have been outright dishonest or fraudulent; many more, probably, are self-deceived on some matters and in some degree; most, almost certainly, have some preconceptions on

\textsuperscript{42} RAMSEY, supra note 35, at 12.

\textsuperscript{43} Id.

\textsuperscript{44} Of course, this falls well short of a complete theory of truth. I do not claim to have such a theory, but I have nibbled away at the problem in HAACK, PHILOSOPHY OF LOGICS, supra note 36, at 86-134; SUSAN HAACK, Confessions of an Old-Fashioned Prig, in MANIFESTO OF A PASSIONATE MODERATE: UNFASHIONABLE ESSAYS 7, 7-30 (1998); Susan Haack, The Unity of Truth and the Plurality of Truths, supra note 11.
some questions that make them less responsive to certain evidence than they should ideally be.

The evidence with respect to scientific claims and theories is usually very complex, ramifying in every direction. It often depends on the reliable working of instruments of various kinds, or on the soundness of elaborate statistical techniques or computer programs; and it is almost always the work of many people—whether collaborators or rivals, and whether working together or many miles or decades apart—who rely, explicitly or implicitly, on the competence and honesty of others involved. Evidence can be misleading, ambiguous, hard to interpret—and it is virtually always incomplete. Getting more evidence may be difficult, prohibitively expensive, or outright impossible in the current state of knowledge or technology; worse, it isn’t always clear even what additional evidence is needed (for whether this evidence is relevant to that claim depends on facts about the world, facts about which we may be mistaken). Whether or not they articulate it explicitly, most serious scientists have a firm-enough grasp of the complexities of evidence; this is why, wary of claiming to have found the truth, they prefer to say, “this seems like a promising idea,” “this model seems to fit what we know so far,” “probably the value of $c$ is approximately $n$,” “perhaps the explanation might be this,” “possibly, it’s this way,” and such.

At any time there is a whole continuum of scientific ideas, claims, and theories: some so well-warranted by such strong evidence that it is most unlikely they will have to be revised; some not quite so well-warranted but still pretty solidly established; some promising but as yet far from certain; some new and exciting but highly speculative and as yet untested; and some so wild that few mainstream scientists are willing even to listen to them. (The proportion of the well-warranted to the highly-speculative varies, obviously, across fields and sub-fields.) A few of the exciting but as yet untested ideas, and a very, very few of the wildest ideas, will eventually turn out to be warrantable, but most will not. A few of the now pretty-well-warranted ideas, and perhaps even a very, very few

---

45 Was I the only person in the country who didn’t laugh derisively at Donald Rumsfeld’s observations about those “unknown unknowns” in military intelligence? From a strictly epistemological point of view, this was an unusually sophisticated remark.
of the best-warranted ideas, will eventually be overturned by overwhelming contrary evidence, but most will not.

There is no algorithmic scientific method guaranteeing success, or even progress. Rather, the many and various techniques and technologies the sciences have gradually devised—the instruments of observation, the cunningly devised experiments, the advanced mathematical and statistical techniques, the ever-fancier computer programs, etc.—have gradually made it possible to get more evidence, more exact evidence, and more focused evidence, and to work out the consequences of evidence, and assess its worth, more accurately. Nor are there “rules of acceptance and rejection” governing when a claim is well-enough warranted to be accepted into the corpus of scientific knowledge, and when badly-enough undermined to be rejected. Rather, virtually any scientific community will include some who are temperamentally inclined to work patiently to modify and adapt an old idea in the light of awkward new evidence, and some who are temperamentally more disposed to move on, to start looking for a fresh approach. And as new evidence comes in, there will be ongoing shifts of opinion out of which—though it may take many years or even decades—consensus will eventually emerge.

“Warrant” is an epistemological term of art: how warranted a claim is depends on how good the evidence with respect to that claim is. “Acceptance” and “consensus,” by contrast, are psycho-sociological terms, referring to scientists’ agreement on the correctness of this or that idea. Though doubtless individual scientists often hope to persuade others that their approach or their conjecture is the right one, scientific investigation is not an effort to reach consensus; it is, rather, an effort to answer the questions at issue. Consensus is ordinarily a byproduct that arises when enough members of the relevant scientific community come to see the evidence as warranting a claim or a theory.

As the previous paragraph suggests, warrant and consensus would, ideally, be tightly correlated: as the evidence favorable to a claim gets stronger, more scientists in the field will be inclined to accept it, or will be inclined to accept it more firmly; and as the evidence against a claim gets stronger,

46 “Epistemology” (from the Greek, episteme, “knowledge”) refers to the field of philosophical investigation focused on such concepts as knowledge, evidence, warrant, justification, inquiry, method, and intellectual character.
more scientists in the field will be inclined to reject it, or to reject it more confidently. At any time, there will be many scientific questions with respect to which the available evidence is still incomplete or ambiguous enough to leave room for legitimate disagreement among competent specialists about whether this account or that is more likely right, or even about whether any conclusion at all can be drawn on the basis of the evidence thus far. As more and stronger evidence comes in, however, more will be persuaded, until the former skeptics are convinced (or perhaps marginalized, or retired), and consensus solidifies.

This seems to have been pretty much what happened, for example, with James Watson and Francis Crick’s model for the structure of DNA—where, by Crick’s reckoning, it took about 30 years, and a lot more evidence than they had in 1953, before theirs became the standard, well-accepted view. But there is no guarantee that warrant and acceptance will always run in parallel; for the process by which bad ideas and false theories get filtered out and rejected, and good ideas get established and accepted, is fallible, untidy, and ragged. Nor, of course, is there any guarantee that all the scientific claims accepted as true at any time are true; almost certainly, some will eventually turn out to have been, not truths, but only “truths.”

Given the investigative character of the scientific enterprise and the pervasive reliance of individual scientists on evidence discovered by others, the core values are honesty (both with yourself and with other people) about what the evidence is and where it leads, and willingness to share evidence unreservedly with other investigators. Along with its constantly evolving ways of acquiring and appraising evidence, the scientific community has found ways to sustain commitment to these core epistemological values. This has been a matter less of regulation than of ethos: ideally, the values of honesty and evidence-sharing will be instilled in young

48 My terminology is new, and I hope less potentially confusing; but these terms correspond approximately to what Robert Merton called the norms of “disinterestedness” and “communism.” Robert Merton, Social Theory and Social Structure 307-16 (1949); see also Susan Haack, The Integrity of Science: What It Means, Why It Matters, Ética e Investigação nas Ciências da Vida, Actas do 10º Seminário de Conselho Nacional de Ética para as Ciências da Vida, 9 (2007), reprinted in Putting Philosophy to Work, supra note 11.
scientists in the course of their long apprenticeship, and will be sustained by the incentive of renown for successful work and the disincentive of loss of reputation for cheating, by conscientious peer review at journals and grant-giving bodies, and by the commitment of the universities to a culture of investigation. So long as these social mechanisms work tolerably well, dishonest or sloppy work will be discouraged or, at worst, will be detected and discarded—not always, but at least when it is significant enough scientifically that, sooner or later, someone thinks it worth his while to try to build on it.

But the social mechanisms that sustain commitment to the core epistemological values of the scientific enterprise are far more fallible and fragile than the technical helps to scientific inquiry—that is, the instruments, computer programs, and such. These social mechanisms are especially susceptible to pressures from the competing values of the society in which scientific work is undertaken: the commercial values of industrial sponsors of science, for example, or the ideological or political values of government sponsors. Presently, moreover, they are under such strain from changes in science funding, changes in the universities, changes at the scientific and medical journals, etc., that the integrity of science is in some danger of erosion.49

The pressure of competing values is not felt evenly, but is more severe in some areas of science than others; and it is most severe, probably, in certain areas of the sciences, the biomedical sciences in particular, on which the legal system very often finds itself calling. This, as we shall see, is one of many reasons why the science that enters the legal system so often disappoints us.

IV. SCIENCE IN THE BRAMBLE PATCH

“One of many reasons”—almost too many, and too untidily intermingled, to list; but I will do my best.

The first thing to stress, though, is that the enterprise to which we refer by the commodious word “science” is enormously complex, and uneven both in character and in quality. “Science” encompasses some of the most remarkable achievements of the human mind, some near-miracles of

49 See HAACK, The Integrity of Science, supra note 48.
“patience and postponement, [of] choking down of preference” as those heroes of the history of science have figured out this or that aspect of the world, and a good deal more routine but solid and significant investigation. But it also includes plenty of (sometimes far-fetched) speculation; and, as with everything human, much that is lazy, sloppy, corner-cutting, self-promoting, or self-deceptive, and some that is flaky or even fraudulent. Moreover, the scientific enterprise—complex, ragged, fallible, fumbling, but thus far remarkably successful, as human enterprises go—operates within the larger society on which it relies for resources, and is vulnerable to pressures to conform to commercial, ideological, etc., values at odds with its own core epistemological norms.

Against this background, it isn’t hard to see why the legal system has had difficulties in handling scientific testimony. It often calls on the weaker areas of science and/or on weak or marginal scientists in an area; moreover, its adversarial character may mean that even solid scientific information gets distorted; it may suppress or sequester relevant data; it may demand scientific answers when none are yet well-warranted; it may fumble in applying general scientific findings to specific cases; and it may fail to adapt appropriately as a relevant scientific field progresses. Let me take these points in turn.

Much of the scientific work on which the law calls comes from weaker or less mature sciences or fields of science. Some comes from fields the market for which is largely, if not exclusively, created by the legal system itself: for example, psychiatric theorizing about (purportedly) recovered memories, surely at the weaker end of the social, or as we sometimes say, the “soft” sciences; and forms of forensic identification, such as latent fingerprint identification, about the reliability of which much is claimed, but remarkably little seems to be known.

---

50 William James, *The Will to Believe*, in *The Will to Believe and Other Essays in Popular Philosophy* 1, 7 (Dover 1956) (1897).

51 And, as Justice Breyer observes, general acceptance in a discipline is little assurance of reliability if the discipline in question is weak or dubious. *Kumho*, 526 U.S. at 151.

Even when the science crucial to litigation comes from relatively stronger scientific fields where there is a substantial body of well-established knowledge, these are quite often the very areas of science where scientists’ commitment to the norms of honesty and evidence-sharing is most sorely tried—making it more likely that awkward evidence will be withheld, or “spun” to reach a desired conclusion. The scientific evidence in toxic tort litigation, for example, often hinges on epidemiological studies undertaken by defendant manufacturers, whether for the purpose of obtaining Food and Drug Administration (“FDA”) approval or in anticipation of the possibility of such litigation. Sometimes such studies are designed to make it more likely that they will detect favorable data than that they will detect unwelcome trends.  

Moreover, even when the relevant areas of science are relatively solid and respectable, the scientific issues in litigation tend to turn not on firmly-established, well-warranted core scientific principles, but on still-controversial scientific issues where there remains room for reasonable disagreement even among competent, honest scientists in the field. The better established it is that this substance is harmless and inert in the human body, for example, the less likely it is that it will be the subject of toxic tort litigation, and the better established it is that it is dangerous to humans, the likelier that cases will be settled; the better established it is that this technique of forensic identification is bullet-proof, the likelier that cases will be pled out.

And because of its adversarial character, the legal system tends to pull in scientists from the farther ends of the spectrum of scientific opinion: those ready to give a confident answer before others think any answer is warranted, those more scientifically radical, or more scientifically conservative, than most of their colleagues, those whose views have become

---


53 For example, the first large clinical trial of Vioxx, the VIGOR trial, conducted by Merck scientists, continued to track gastrointestinal effects (anticipated to be favorable) after it stopped tracking cardiovascular effects (anticipated to be unfavorable). See HAACK, The Integrity of Science, supra note 48; David Armstrong, Bitter Pill: How the New England Journal Missed Warning Signs on Vioxx, WALL ST. J., May 15, 2006, at A1.
dogmatically entrenched in the course of their involvement in the litigation process—and, no doubt, a few outright cranks and a few outright whores. But even when a scientific question is regarded by most people in the relevant scientific community as pretty well-settled, attorneys for the parties to a suit are motivated to seek out as expert witnesses those scientists who still have reservations on the matter. And even when a question is still within the realm of reasonable scientific disagreement, the attorneys are motivated to seek out those scientists who are already most firmly convinced one way or the other. As a result, the legal process can sometimes create spurious, artificial scientific certainty, and spurious, artificial scientific doubt.

Moreover, because advocates will try to get evidence unfavorable to their case excluded, and because settlements may require confidentiality, the adversarial system can contribute to scientific secrecy (withholding evidence) as well as spin (distorting evidence). Not surprisingly, also, when scientific developments attract the attention of the legal community, researchers may find their work interrupted by subpoenas and depositions, and sometimes attacked, as attorneys seek to discredit it.

Because they are specific to a particular case or individual, the questions to which the legal system needs answers are rarely exactly the questions on which the relevant scientific work would ordinarily focus. There may, for

---

54 It has also been suggested that the unappealing possibility of being Dauberted out makes some scientists, including some of those one would like to have involved, reluctant to participate in legal disputes. See, e.g., Heinzerling, supra note 25, at 75-77.


56 See, e.g., Catharine Arnst, Keeping Lawyers Out of the Lab: Researchers Gripe that Suits Arising from Their Findings Waste Time and Hurt Their Reputations, 4020 BUS. WK., Feb. 5, 2007, at 54 (reporting that “[s]cientists who would prefer to plug away quietly in their labs say they are dragged into lawsuits the moment their research turns up a hint of the possible cause of a disease”).

57 This may in part explain, for example, why despite the fact that the dangers of PCBs were already so well known that the stuff had been banned for decades, General Electric was able successfully to defend itself against Mr. Joiner’s claim. Precisely because PCBs had long been banned, there were no large, directly relevant epidemiological studies on which Mr. Joiner’s attorneys could rely; and GE’s attorneys could suggest other, arguably more plausible, explanations of what caused him to develop an aggressive lung cancer at an early age. See Michael Gottesman, From Barefoot to Daubert to Joiner: Triple Play or Double Error?, 40 ARIZ. L. REV. 753, 766-69 (1998). (Mr. Gottesman represented both the Dauberts and the Joiners at the Supreme Court.)
example, be solid scientific work on the distribution of a disease or disorder in the population as a whole, and reasonable theories about its etiology, but there is unlikely to be scientific work directly relevant to whether this risk factor was a significant cause of this person’s contracting the disease.\textsuperscript{58} Or, as in the case of forensic identification by DNA, where the relevant science is very well-warranted indeed, its application to the samples from this crime scene and this defendant introduces a whole raft of opportunities for corner-cutting, sloppiness, self-deception, and plain dishonesty.

The very structure of evidence law can create further problems. The evidence that warrants a scientific claim is likely to include a complex mesh of interlocking reasons, none of which by itself would be sufficient to warrant the claim. (This is not to suggest, as some proponents of “weight of evidence methodology” seem to think, that the combined weight of evidence can be assessed by some mechanical formula; it is only to say that, where many lines of evidence interlock in the right way, they may jointly warrant a claim even though none would do so by itself.)\textsuperscript{59} But courts screening expert scientific testimony will determine, for each proffered witness, whether his testimony is admissible, and often on what specific matters he may offer an opinion. This may mean that even when there

\textsuperscript{58} When \textit{Daubert} was remanded to the Ninth Circuit, Judge Kozinski ruled the plaintiffs’ expert epidemiological testimony inadmissible, finding (1) that “California tort law requires plaintiffs to show not merely that Bendectin increased the likelihood of injury, but that it more likely than not caused their injuries”; and (2) that

\textit{[i]n terms of statistical proof, this means that plaintiffs must establish not just that their mothers’ ingestion of Bendectin increased somewhat the likelihood of birth defects, but that it more than doubled it—only then can it be said that Bendectin is more likely than not the cause of their injury.}

\textit{Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311, 1320 (1995). Judge Kozinski cited Jones v. Ortho Pharm. Corp., 209 Cal. Rptr. 456, 460 (1985) (requiring that it be “more likely than not” that the injury was caused by defendant’s action or product) and DeLuca v. Merrell Dow Pharm., Inc., 911 F.2d 941, 958 (3d Cir. 1990) (“[T]he relative risk of limb reduction defects arising from the epidemiological data [Dr.] Done relies upon will . . . have to exceed ‘2’”). (I note, however, that there is no reference in Jones to a doubling of the risk; and that in DeLuca, where there is, the court of appeals reversed and remanded the lower court’s grant of summary judgment for Merrell Dow, which had been based on its exclusion of the testimony of Dr. Alan Done that Bendectin did double the risk of limb reduction. \textit{DeLuca}, 911 F.2d at 959.)}

\textsuperscript{59} See Susan Haack, \textit{An Epistemologist Among the Epidemiologists}, 15 \textit{Epidemiology} 521 (2004), reprinted in \textit{PUTTING PHILOSOPHY TO WORK}, supra note 11. Cf. Oxendine v. Merrell Dow Pharm., Inc., 506 A.2d 1100, 1110 (D.C. 1986) (“Like the pieces of a mosaic, the individual studies showed little or nothing when viewed separately . . . but they combined to produce a whole that was greater than the sum of its parts: a foundation for Dr. Done’s opinion that Bendectin caused appellant’s birth defects.”).
is a congeries of evidence which, taken together, strongly suggests that exposure to this substance is causally related to that disease or disorder, the rules of evidence may preclude its admissibility because no individual component, by itself, is deemed to meet the *Daubert* standard of reliability.60

Not only does the legal system quite often want scientific answers when no warranted answers are available, it also quite often fails to adapt, or adapts painfully slowly, as new scientific answers become available. The law looks to precedent, and courts sometimes continue to follow earlier rulings based on now-superseded science. In Texas death-penalty sentencing hearings, for example, courts continue to rely on psychiatric or “soft” social-scientific testimony as to the likelihood that a defendant convicted of first-degree murder will be dangerous in future—even though (somewhat) more accurate actuarial methods of prediction are now available.61

It is no mystery, then, why the law often gets less than the best out of science. But one more task remains: to explain how scientific claims that are not true are sometimes transmuted into legal truths.

V. HOW SCIENTIFIC “TRUTHS” GET ENTRANCED AS LEGALLY RELIABLE

“True” has the same meaning when it is applied to legal propositions as it does when applied to propositions of other kinds. It is true that domestic cats are related to tigers if and only if domestic cats are related to tigers; similarly, it is true that Florida law in 2007 requires that novel scientific testimony be “sufficiently established to be generally accepted in the field to which it belongs,” if and only if Florida law in 2007 does require that novel scientific testimony be sufficiently established to be generally accepted in the field to which it belongs.


61 See, e.g., Thomas Regnier, *Barefoot in Quicksand: The Future of Future Dangerousness Predictions in Death Penalty Sentencing in the World of Daubert and Kumho*, 37 AKRON L. REV. 469, 485-89 (2004); Erica Beecher-Monas & Edgar Garcia-Rill, *Genetic Predictions of Future Dangerousness: Is There a Blueprint for Violence?*, L. & CONTEMP. PROBS., Winter/Spring 2006, at 301, 301-02. (I don’t mean to imply that the fact that a person convicted of murder will be dangerous in future is a reason to sentence him or her to death; but this is not the place to engage with that issue.)
belongs. However, such legal truths are not exactly like, say, the truths of physics, but are more like social-scientific truths.

Truths to the effect that the law is thus and so have to be understood to be specific to a jurisdiction and to a time (as many truths about social institutions have to be understood to be specific to a society and a time). For example, in 2003 it was true that Michigan law required admissible scientific testimony to satisfy the Frye test, but since January 1, 2004, it has been true that Michigan law requires expert, including scientific, testimony to satisfy the Daubert requirements. Sometimes it is neither true nor false that the law in such-and-such a jurisdiction is thus and so; for example, until the Supreme Court settled the question in Daubert, it was neither true nor false that FRE 702 superseded the Frye Rule in federal courts. And what makes it true that the law in such-and-such a jurisdiction at such-and-such a time is thus and so is what legislators and courts do.

It is standard to distinguish statements of fact, such as “there was a stop sign at the intersection on the day of the accident,” from statements of law, such as “felony murder is an unlawful homicide occurring in the commission or attempted commission of a felony.” (It might, perhaps, be better to speak of statements of law versus statements of non-legal fact; but the more standard terminology will serve.) Usually, the intended distinction is straightforward enough, but there are significant borderline and mixed cases, one of which is directly to the present purpose. Questions about admissibility would be classified as questions of law, falling within the province of the court; questions about the weight of evidence would be classified as questions of fact, falling within the province of the jury. But rules about the admissibility of expert testimony have blurred this apparently simple dichotomy.

Specifically, by requiring courts to screen proffered expert scientific testimony for reliability as well as relevance, and thus extending their preserve, Daubert has shifted some questions formerly conceived as concerning the weight of

---


evidence into the category of questions bearing on admissibility. This is why attorneys sometimes complain that, since Daubert, they are in effect obliged to try their cases twice—first at an evidentiary hearing before the court, and then a second time before the jury; and why, if one party’s expert testimony is excluded, the case may well end in summary judgment against them. In short, the concept of evidentiary reliability seems to straddle the line between law and fact.

Like the legal concepts of insanity, causation, or intoxication, the legal concept of reliability is at once closely related to, and yet importantly distinct from, ordinary, non-technical concepts. The legal concept of insanity isn’t exactly the same as everyday conceptions of madness, but isn’t entirely disjoint from them; the legal concept of causation isn’t exactly the same as everyday conceptions of cause, but isn’t entirely disjoint from them; the legal concept of intoxication isn’t exactly the same as everyday conceptions of drunkenness, but isn’t entirely disjoint from them; and similarly, the legal concept of reliability isn’t just the same as everyday understandings of reliability, but isn’t entirely disjoint from them, either.

In fact, evidentiary reliability is a puzzling, Janus-faced concept. The doubleness is already close to the surface in Daubert, in that curious footnote where Justice Blackmun tells us that reliability, in the specifically legal sense he is articulating, is not to be identified with “scientific reliability” (which he takes to mean that “application of the principle gives consistent results”), but corresponds to “scientific validity” (which he takes to mean that “the principle supports what it purports to show”). But then he goes on to assure us that evidentiary reliability is simply a matter of the “trustworthiness” of the testimony—which, however, is surely significantly stronger than “the principle supports what it purports to show.”

You might wonder what degree of reliability is required for expert testimony to be admissible. For in ordinary speech,
though “reliable” has quite a tangle of uses, in all them—whether we are speaking of inanimate objects such as watches or motor vehicles, or of persons, or of information, data-bases, etc.—it connotes something, fitness to be relied upon, that comes in degrees. But admissibility is not a matter of degree, which means that the Daubert ruling requires evidentiary reliability to be categorical.67 Most to the present point, though, the rationale for introducing the concept of evidentiary reliability depends crucially on its connection with the ordinary concept of reliability-as-trustworthiness; but qua legal concept evidentiary reliability simply requires that the testimony at issue satisfy certain legally specified conditions.68 There is, however, absolutely no guarantee that all, or only, testimony deemed by courts to meet the standards of evidentiary reliability is trustworthy, that is, is reliable in the ordinary sense of the word.69

In Joiner, rejecting respondents’ (Joiner’s attorneys’) argument that the trial court’s evidentiary ruling excluding their proffered expert testimony should have been subject to especially stringent review, given that this exclusion was outcome-determinative, Justice Rehnquist writes: “On a motion for summary judgment disputed issues of fact are resolved against the moving party . . . . But the question of admissibility of expert testimony is not such an issue of fact . . . .”70 Here—even though, had the evidence been admitted, its degree of reliability would have been a factual matter for the jury to determine—the question of evidentiary reliability is treated as a purely legal matter.

Claims to the effect that this or that scientific evidence is legally reliable can be made true by legal decisions.

———

67 See, e.g., In re Paoli RR. Yard PCB Litig., 35 F.3d 717, 744 (3d Cir. 1994) (observing that “[t]he evidentiary requirement of reliability is lower than the merits standard of correctness”); see also Nenno v. State, 970 S.W.2d 549, 561 (Tex. Crim. App. 1998) (holding that social-scientific evidence should not be required to meet standards of reliability as high as those demanded for the admissibility of natural-scientific testimony).

68 Conditions which, however, are significantly less precisely specified since Kumho.

69 Compare this, on the preponderance of evidence standard of Federal Rule of Evidence 104(a), from Bourjaily v. United States, 483 U.S. 171, 175 (1987): “The inquiry made by a court . . . is not whether the proponent of the evidence wins or loses his case on the merits, but whether the evidentiary Rules have been satisfied.”

70 Joiner, 522 U.S. at 143 (emphasis added); see also Heinzerling, Doubting Daubert, supra note 25, at 80 (arguing that Joiner makes evidentiary reliability a kind of legal chimera, neither an issue of fact to be decided in favor of the nonmoving party nor an issue of law subject to de novo review).
However—verbal appearances to the contrary—it doesn’t follow that scientific truths can be legally constructed. Legal truths are made true by legal decisions; and so, sometimes, are scientific “truths.” But it is the character of phenomena and events in the world that scientific propositions describe—not legal decisions about evidentiary reliability, and not arguments and cross-examination in court—that make true scientific propositions true, and false scientific propositions false.
Expert Evidence, Partisanship, and Epistemic Competence

Jennifer L. Mnookin†

In various ways, skilled witnesses have been used in courtroom processes since just about the dawn of the jury trial. The expert witness in its modern form—a witness whose presence in court results not from being a percipient witness to material facts, but instead because of education, training, experience, or other specialized knowledge relevant to the case, and who is called by one party to testify, and is typically compensated by that party as well—can be traced back to at least the middle of the nineteenth century.

But the use of adversarial expert witnesses in court has been problematic from just about the moment of its invention. In this brief essay, I will explore two fundamental causes of the awkward fit between expert knowledge and our adjudicatory processes: the twin problems of partisanship and epistemic competence.

The use of expert evidence in court has been criticized for a remarkably long time. Consider the following three quotations:

But the practice [of using expert witnesses] under the present method has for years exhibited shortcomings which are lamentable. . . . The principal feature of the breakdown seems to be the distrust of the expert witness as one whose testimony is shaped by his bias for the party calling him. That bias itself is due, partly to the special fee which has been paid or promised him, and partly to his prior consultation with the party and his self-committal to a particular view. His candid scientific opinion thus has had no fair opportunity of expression, or even of formation, swerved as he is by this partisan committal.1

† Vice Dean for Faculty & Research and Professor of Law, UCLA School of Law. I thank the participants for an engaging and provocative symposium and particularly Margaret Berger and Larry Solan for organizing it. Thanks also to Jennie Katz and Janelle Ureta for research assistance.

Experts in other fields see lawyers as unprincipled manipulators of their disciplines, and lawyers and experts alike see expert witnesses—those members of other learned professions who will consort with lawyers—as whores.²

Now in the present instance I have, as usual, the evidence of experts on the one side and on the other, and, as usual, the experts do not agree in their opinion. There is no reason why they should. . . . A man may go . . . to half-a-dozen experts. . . . He takes their honest opinions, he finds three in his favor and three against him; he says to the three in his favor, “will you be kind enough to give evidence?” And he pays the three against him their fees and leaves them alone; the other side does the same. It may not be three out of six, it may be three out of fifty. I was told in one case, where a person wanted a certain thing done, that they went to sixty-eight people before they found one. . . . Therefore I have always the greatest possible distrust of scientific evidence of this kind.³

These views span more than a century, but they sound a remarkably consistent note. Though the rhetorical styles are quite different, the underlying message is strikingly similar: Expert witnesses in court are often not deserving of our confidence. Their conclusions cannot be relied upon, and their words cannot be trusted. Indeed, a century’s worth of writing about expert evidence circles around the same themes and consistently reaches the same conclusion: that the use of party-selected expert witnesses in an adversarial legal system is fraught with difficulties.⁴

Why is this so? At root, the use of expertise in our adversarial system raises two equally significant fundamental dilemmas: the problem of partisanship and the problem of epistemic competence. First, given that experts are called by one party and paid by that party, there is an inevitable danger of bias in favor of that party. The less extreme version of this concern is that as the expert prepares for and becomes enmeshed in the case, he will increasingly, if unconsciously,

side with the party that hired him, lose some degree of objectivity, and slant his testimony in that party's favor. The more dramatic version of the same fear is that some unscrupulous experts will literally offer themselves for hire, selling their opinions and their credentials to anyone who meets their price.5

Despite these dangers, refusing to permit payment to experts is obviously not a viable option. It is wholly unrealistic to imagine that those highly qualified experts whom we want to have participating in our adjudicatory process would (or should) devote their time and energy to the courts pro bono on a regular basis. Preparing for and testifying for trial can be extremely time consuming, and experts can otherwise be spending that time engaging in other professionally and/or economically remunerative activities, or enjoying their favorite leisure activities. Hardly anyone would view giving expert testimony as one of her favorite leisure-time activities. The reality is that experts must be paid.

The acute difficulty comes not simply from the fact of payment, but rather from the fact that it is the parties who choose and pay their experts. What a particular party views as the greatest value for its dollar—effective expert testimony that persuades the factfinder—will often not be commensurate with what a more systemic perspective would see as most valuable, which would presumably be careful, accurate expert testimony rather than testimony most persuasive to a non-expert. What this means is that those witnesses who succeed in the marketplace for experts within our adversarial process will often not be those with the most knowledge or actual expertise in a particular area, but rather those whom parties believe will succeed in persuading the factfinder. The confluence of adversarialism with the need for expert information also has permitted the creation of a class of “professional” expert witnesses, those for whom expert witnessing is no longer a sideline, a once-in-a-while add-on to their primary work as a physician, economist, epidemiologist, statistician, or whatnot, but rather is now a significant, or even primary, source of their

5 For the classic (pre-Daubert) article on the widespread frustration with the use of expert evidence and the structural problems with the use of expert knowledge in an adversarial system with lay fact-finders, see Gross, supra note 2. For a polemical but influential account of the problems with the use of expert evidence in civil cases, see Peter W. Huber, Galileo’s Revenge: Junk Science in the Courtroom (1991). For a description of the significant concerns about partisanship in the late nineteenth century, see Mnookin, supra note 4.
earnings. This group, obviously, has an especially strong interest in maintaining its marketability by being a “team player,” and telling potential employers (that is, parties) what they want to hear. The marketplace for experts cannot, therefore, be trusted to produce reliable information. To whatever extent price can be correlated with quality in other domains, the relationship cannot be counted on with respect to expert witnesses.

The second fundamental problem with the adversarial expert is epistemic. Experts are necessary precisely because of what the jury does not know. They are supposed to provide information useful to the jury’s decision-making that goes beyond what a jury would know without their assistance. But if the jury lacks the knowledge that the expert provides, how, then, can it rationally evaluate the expertise on offer? To be

---

6 As part of tort reform efforts, several states have attempted to curb “professional experts” in medical malpractice cases by statutory provisions limiting who can testify as an expert. Kansas, for example, requires that testifying experts have spent at least fifty percent of the two years preceding the incident giving rise to the claim in “actual clinical practice.” KAN. STAT. ANN. § 60-3412 (2007). Connecticut requires a testifying expert to have been active in the practice or teaching of medicine within the five years preceding the incident giving rise to the claim. CONN. GEN. STAT. § 52-184c (2007). Michigan requires a testifying medical expert to have spent a majority of the year preceding the incident giving rise to the claim engaged in active clinical practice or teaching. MICH. COMP. LAWS § 600.2169 (2008).


8 Federal Rule of Evidence 702, which controls the use of expert testimony, states:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

FED. R. EVID. 702. In order to “assist the trier of fact,” the expert testimony must go beyond what the trier of fact would have known and understood even without the expert.

9 As Learned Hand wrote in his well-known 1901 article about expert evidence:

The trouble with all this is that it is setting the jury to decide, where doctors disagree. The whole object of the expert is to tell the jury, not facts, as we have seen, but general truths derived from his specialized experience. But how can the jury judge between two statements each founded upon an experience confessedly foreign in kind to their own? It is just because they
sure, one might not need to be an expert in order to assess expertise, but the main mechanisms for assessing expertise outside of one’s domain of knowledge are, by necessity, secondary indicia, proxies: demeanor, perhaps, or credentials, or superficial explanatory plausibility. But because each party has the power to select its experts from the whole universe of experts willing to testify, parties will presumably attempt to select those experts who best satisfy the parties’ best guesses about what the jury will use as its proxy criteria. That might be a Ph.D. from a prestigious institution or a lengthy publication record. Perhaps it is certain forms of speech or dress, or an honest face and a winning testimonial manner. Most likely it is a mixture of all of the above and more. Whatever the specific criteria, the point is that parties (sometimes with the help of jury consultants) will deliberately select experts who satisfy their beliefs about the jury’s expectations for experts. Parties will, quite rationally, seek out precisely those experts most capable of “performing” the role of expert in just the way that the parties expect that a jury will find credible. Parties do not have infinite latitude, as they will have to choose from whatever array of experts is willing to testify in a way that substantively helps their case. But compared to fact witnesses, they have a great deal of leeway.10 With fact witnesses, a party is typically severely limited by the happenstance of who was there and who saw what; not so with expert witnesses, who can be selected from a national or even global pool, resources permitting. As a result, the power of proxy criteria, like demeanor or credentials, to discriminate between reliable and unreliable experts is likely to be quite limited indeed.11

10 Gross, supra note 2, at 1126-28.

11 Demeanor as a signal for credibility may not be particularly accurate outside of the expert context either. See generally Jeremy A. Blumenthal, A Wipe of the Hands, A Lick of the Lips: The Validity of Demeanor Evidence in Assessing Witness Credibility, 72 Neb. L. Rev. 1157 (1993); Olin Guy Wellborn III, Demeanor, 76 Cornell L. Rev. 1075 (1991). My point, however, is not to defend (or attack) demeanor as an accurate means by which to detect honesty, but simply to suggest that whatever degree of utility it has for the assessment of credibility for non-expert evidence, it is significantly diminished vis-à-vis experts themselves because parties can select them precisely for their demeanor.
Moreover, the problem of epistemic competence compounds the problem of partisanship. Because the jury does not have the expertise to evaluate the substance of expert testimony, it is unlikely that it will be an accurate evaluator of partisan bias. If the jury has epistemic competence, we may not need to be overly concerned with partisanship. The jury can independently evaluate the substance of the testimony and will often have the capacity to see through overstatements or inaccuracies that were the result of zealotry. Without epistemic competence, the jury has no choice but to rely on proxies as secondary indicia of bias, and these may often be either inaccurate or difficult to evaluate.

For example, does the $550 per hour received by the expert mean that his testimony should be discounted because he is reaping a tidy profit, or does it reflect his high stature and commensurate ability that commands an appropriately high price? Is it a sign of trouble that an expert has testified dozens of times before, and thus might be an “expert for hire,” or is it a positive sign, showing that other judges have already deemed him sufficiently expert to warrant being heard by the jury? Does the fact that the plaintiff’s expert seems to testify exclusively for plaintiffs suggest a deep-seated bias, or is the expert who testifies for plaintiffs and defendants alike a bigger concern, possibly suggesting that he will testify for anyone who meets his price? How much, if it all, should a published study in a peer-reviewed journal be discounted because it was funded by a private entity, such as a drug company, with an interest in the outcome of the research conducted? On the one hand, such a funding source could generate a bias; on the other hand, the structure of academic research and the processes of peer review and publication are designed, at least aspirationally, to check such a bias. What about a study conducted especially for this lawsuit? Is its trustworthiness diminished because it did not emerge through the typical research process, or is it the quite appropriate result of an expert, or community of experts, developing an interest in the relevant question precisely because of the lawsuit itself? The point is not that secondary indicia can never provide information relevant to an evaluation of partisanship—rather, it is that evaluating these secondary

---

12 This is precisely what Judge Kozinski suggested in his opinion on remand in Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F.3d 1311, 1317 (1995).
indicia is a murky and imperfect process given the combination of (1) adversarialism and (2) a lack of epistemic competence.

These are not new problems. They are, in fact, rather old problems. Indeed, the risks posed by expert testimony—the danger of partisanship and the problem of epistemic competence—have long been recognized, but never fully resolved, especially when considered together. And hence we continue to face these problems—in court, in our scholarship, in the jury room. Given these fundamental and seemingly irresolvable problems with experts in court, what is the law to do?

In what follows, Part I will survey the traditional, historical approaches the law has taken in attempting to resolve these central problems of partisanship and epistemic competence. Part II will then consider potential solutions to these fundamental quandaries, evaluating both the theoretical appeal and pitfalls of these approaches.

I. THE HISTORICAL PERSPECTIVE: THE SPORTING THEORY, FRYE, AND DAUBERT

First, let us look at what the law has done. The traditional common law approach to these structural dilemmas was quite simply the adversarial mechanism itself: both parties had an equal opportunity to make use of expert witnesses if they wished to do so. Yes, the witnesses for each party might tend to partisanship, but somehow from their clashing testimony, the jury in its infinite wisdom would distill the truth—or at least that was the hope. It was a “sporting theory of justice” applied to experts: so long as parties had an equal opportunity to bring forward opposing experts, under the same rules and with the same judge as umpire, then whatever the jury made of the competing experts’ stories was acceptable.13

This approach, however, was roundly criticized as early as the closing decades of the nineteenth century.14 The problem was the lack of epistemic competence: if juries could be counted on to have the ability to assess the expertise before them, then a level adversarial-playing field might indeed have been all that was needed. But given that juries often lacked the

---

14 For a detailed look at views about expert evidence at the end of the nineteenth century, see generally Mnookin, supra note 4.
competence to adequately evaluate the expertise on offer, the “sporting theory of justice” had the effect, as Roscoe Pound noted, of turning “expert witnesses[] into partisans pure and simple.”

Although critiques of expert testimony have been both frequent and strongly worded over the course of the twentieth century, the adversarial expert has remained an increasingly significant feature of the adversarial trial. And for the most part, despite criticisms, the adversarial process itself remained the dominant check, such as it was, on expert testimony until quite recently.

There were, to be sure, some earlier fledgling efforts to regulate the content of expert testimony through limitations on admissibility. The Frye rule, articulated in 1923, made explicit that for novel kinds of expert testimony, courts could require the knowledge to have gained “general acceptance” in the relevant expert community before permitting it before a jury. But Frye, though it became important in the 1970s, was not much noticed at the time it was decided. Through 1970, it was cited only fifty-eight times, and the bulk of those cases involved the lie detector, the same technology at issue in Frye.

Instead, the main vehicle for such regulation as the courts wished to exercise was qualifications: in order to testify, the expert had to have qualifications that were adequate to support his claim of expertise. Just how qualified was qualified enough? No doctrinal framework emerged to answer this question, and in practice, most judges, most of the time, did not actually interrogate a proposed experts’ bona fides in a detailed or rigorous way. In addition, trial judges’ determinations about qualifications were generally viewed as so much a matter of the trial court’s discretion as to be virtually unreviewable on

---

15 Pound, supra note 13, at 448.

16 Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923). Frye’s now classic and oft-quoted key language says:

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle of discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field to which it belongs.

Id.

17 This is based on a search for cases reported by Lexis decided prior to 1970 that cite Frye.
appeal. In practice, then, the only significant check on partisanship or even outright charlatanry was the power of the opposing attorney to cross-examine the expert. Even if one did agree with John Henry Wigmore that, as a general matter, cross-examination counted as “the greatest legal engine ever invented for the discovery of truth,” it was often not terribly effective when applied to expert witnesses.

The net result was that although few defended it as sound, the “sporting theory” as applied to expert witnesses continued to reign even into the last quarter of the twentieth century. Few doubted that this sporting theory sometimes led to embarrassing results. Some critics began suggesting that the inadequate regulation of expert witnesses was even contributing to a liability crisis, in which socially valuable products were being forced off the market because of the cost of defending against baseless tort suits. These critics' idea was that such lawsuits often lacked scientific merit, but plaintiffs’ lawyers were nonetheless able to hoodwink the jury into granting sizeable verdicts. And, according to the critics, these suits were able to get past summary judgment and reach trial precisely because of the presence of hubristic experts prepared to testify to causation with little or no basis in fact.

This, then, was the historical backdrop for the Supreme Court’s 1993 pronouncement on expert evidence, *Daubert v. Merrell Dow Pharmaceuticals, Inc.* As a formal matter, *Daubert* held that the Federal Rules of Evidence, which became effective in 1975, did not incorporate the *Frye* principle of

---

18 1 Wigmore, supra note 1, §§ 561, 670.
19 Id. vol. 3, § 1367.
20 For a satirical look from the late nineteenth century making fun of several kinds of expert witnesses (and illustrating the ineffectiveness of efforts to cross-examine them), see Richard Harris, *Hints on Advocacy* (St. Louis, Central Law Journal Co., 9th ed. 1892) (1880).
22 The classic account making this argument is Huber, supra note 5.
“general acceptance.” While rejecting the allegedly “austere standard” of Frye, Daubert parsed the language of Rule 702 of the Federal Rules of Evidence, and interpreted the rule’s use of the words “scientific . . . knowledge” to mean that federal judges have an obligation to serve as “gatekeepers,” who make sure that scientific evidence offered in court is, in fact, “scientific knowledge.” By emphasizing a judicial gatekeeping role for the assessment of expert evidence, Daubert was a meaningful move away from a pure “sporting theory” with respect to expert evidence.

Tracing the history of expert evidence, Daubert completes a shift in perspective whose outlines were already implicitly visible in Frye. Prior to Frye (and to a great extent afterwards as well), the key concern was qualifications: Is this “expert” an appropriate person to speak to the issue at hand? Does this person, through training, experience, or education have the right kind of expertise? Do his credentials qualify him to give the jury an opinion on this matter? The courts, to be sure, were not typically extremely strict about qualifications—physicians, for example, were often permitted to testify outside their primary area of expertise, and somewhat weak credentials were often seen as going to evidentiary weight rather than admissibility—but still, qualifications were the primary hook by which judges evaluated an expert’s authority. The Frye test, by contrast, was less concerned with the speaker and his qualifications and was more concerned with the substance being asserted: had the speaker’s claims reached general acceptance in the relevant field? At least in theory, an indisputably qualified expert, even though testifying to matters within his sphere of expertise, could still be excluded under Frye because his conclusions lacked general acceptance within the appropriate expert community.

But well after 1923 many courts ignored Frye, and even those courts that followed it often applied it in a form that reverted back to a qualifications test. They found that the test was satisfied so long as the properly qualified expert asserted that the substance of his claims was, in fact, generally accepted. Why should a court believe an expert’s say-so about general acceptance? Why, precisely because the expert had the appropriate qualifications. Not only does such logic have a

---

24 Id. at 588.
25 Id. at 589-91.
certain unavoidable circularity, but to the extent a credentialed expert’s own say-so is deemed adequate to establish general acceptance, a test that seems superficially different in form amounts to nothing but another version of a qualifications test. *Daubert*, by contrast, makes clear that for scientific evidence, qualifications alone ought not to suffice. The speaker’s individual expertise is still necessary, but it is no longer sufficient. Rather, the key question is whether the substance offered by the speaker has adequate indicia of reliability.26 *Daubert* thus marks a doctrinal shift away from a focus on the speaker as a person and toward a focus on the validity of the claims made.

More generally, *Daubert* was certainly a response to the twin concerns about partisanship and the epistemic competence of juries. As always, adequate qualifications remain a necessary prerequisite, but after *Daubert*, it is crystal clear that an expert’s qualifications are not, themselves, a sufficient condition for admissibility. Instead, the judge must examine the substance of the expert testimony to see whether it is adequately reliable. Counter-expertise introduced by an opposing party can, after *Daubert*, no longer be said to be an adequate cure for the problem of partisanship; rather, judges must themselves establish that the expert evidence has adequate indicia of reliability.

But even though *Daubert* can be seen as something of a response to the twin concerns of partisanship and epistemic competence, it can hardly be said to resolve these issues. *Daubert’s* approach is to have a non-scientist judge make a preliminary determination about reliability in order to limit what the jury can consider. Nothing in *Daubert* explicitly addresses partisanship (though to be sure, part of the purpose of gate-keeping for reliability is to endeavor to distinguish partisan excess from legitimate expertise). And, as many have suggested, beginning with Rehnquist in his *Daubert* dissent and continuing as a leitmotif, it is far from clear that judges have the epistemic competence to make legitimate decisions about what expert evidence is adequately valid and what is not.27

---

26 Id. at 589-90.
27 See id. at 600-01 (Rehnquist, C.J., dissenting).
II. PARTISANSHIP AND EPISTEMIC COMPETENCE: CONSIDERING ALTERNATIVE SOLUTIONS

If Daubert represents, at best, an indirect and partial effort to address the issues of partisanship and epistemic competence, what would more direct efforts look like? In what follows, I want to suggest, unfortunately, just how difficult it is to solve these problems. I will briefly address the two most obvious and often-proposed solutions for the twin difficulties of partisanship and epistemic competence, and will suggest, alas, that that these solutions offer less than meets the eye—that upon reflection, they are not in fact satisfying solutions to the problems besetting expert testimony.

First, let us look at solutions that face up to these fundamental tensions directly. If the problems with expert evidence in the adversary system are partisanship and epistemic competence, it might seem like the obvious solution is to take on one, the other, or both. How might this be done? For partisanship, it would seem that the obvious solution is to make experts non-partisan—to make them neutral or court appointed, answering to someone other than one of the parties. For epistemic competence, it would seem like the obvious solution would be to employ decision-makers or arbiters with epistemic competence: in other words, expert judges or expert juries.

Each of these solutions has been suggested before. Each will no doubt be suggested again. If history is any guide, we are unlikely to head terribly far down either path within the trial process itself. I want to suggest, however, that while each of these potential solutions does address one piece of the underlying structural dilemma, neither offers a sound alternative for resolving the fundamental structural tensions as a whole.

A. Partisanship, Neutral Experts, and Evidence Synthesis

Back in 1901, the young Learned Hand penned a still-cited article examining the difficulties posed by the use of expert evidence. His proposed solution was to create a system for neutral, court-appointed experts, unbiased advisors who would be able to deliver to the jury “those general truths, applicable to the issue, which they may treat as final and
decisive.” Hand was far from the first to think that neutral experts were a promising way to solve the problems associated with experts in court, and throughout the century that followed a great number of subsequent critics proposed variations on the same solution. There have been proposals for court-appointed experts, for government-appointed experts; for neutrals in lieu of party-controlled experts, for neutrals as a supplement to party-controlled experts, for incentive schemes to encourage opposing parties to agree on a neutral expert, and so on. For our purposes, the great variety of possible institutional designs for the use of neutral experts is not the point.

The idea of neutrals is obviously an attractive corrective to the problem of partisanship. If one of the significant problems with the use of expertise in court is that that experts tend to be biased in favor of the party that hired them, then creating a category of expert that lacks this bias would seem like a natural way to improve the information available to the factfinder. The main advantage neutral experts would offer is precisely that they would not be (metaphorically) in bed with one of the parties.

To the extent experts are corrupted by their association with the parties, the use of neutral experts could indeed be beneficial. But to see them as a panacea misunderstands the nature of most scientific disputes that arise at trial. Neutrals will only offer an adequate solution to the problems besetting the use of experts in court when, in fact, there is a reasonably high degree of consensus within the scientific community on the scientific question at issue in the case.

29 For still earlier calls for neutral or court appointed experts, see, e.g., Emory Washburn, Testimony of Experts, 1 AM. L. REV. 45, 61-62 (1867); J.B. Ransom, Medical Expert Testimony, 16 MEDICO-LEG. J. 30, 31-34 (1899); Henry Mott, Expert Testimony, 11 MEDICO-LEG. J. 44, 45 (1893); Clemens Herschel, Services of Experts in the Conduct of Judicial Inquiries, 21 AM. L. REV. 571, 577 (1887).
30 For a sampling of these more contemporary calls for the greater use of neutral or court appointed experts, see MARCIA ANGELL, SCIENCE ON TRIAL: THE CLASH OF MEDICAL EVIDENCE AND THE LAW IN THE BREAST IMPLANT CASE (1996); Gross, supra note 2; Daniel W. Shuman & Bruce D. Sales, The Impact of Daubert and Its Progeny on the Admissibility of Behavioral and Social Science Evidence, 5 PSYCH. PUB. POLY & L. 3 (1999).
31 Note, however, that under current law, courts do have the power to appoint experts when they deem it necessary. See FED. R. EVID. 706. This power, however, is rarely exercised.
Some of the time, this is no doubt the case. By the time *Daubert* went to trial, for example, the question of whether the drug Bendectin, frequently prescribed to treat acute morning sickness during pregnancy, was teratogenic had received an enormous amount of scientific attention (in significant part due to the wave of litigation surrounding it). Though there were still credentialed scientists who disagreed (some of whom were hired by the plaintiffs), it is fair to say that the great weight of scientific opinion interpreted the existing evidence as sufficient in quantity and quality to strongly support the inference that Bendectin was not teratogenic.32

But in many cases, there may be genuine disagreement across the scientific community about how to interpret the existing evidence on causation. Take a toxic torts case in which the plaintiff claims that her harm resulted from exposure to a substance produced by the defendant. Often the key issue in such cases is causation; there may be no doubt that the plaintiff was harmed, but the question is whether it was the defendant’s product that caused the harm. Frequently, when plaintiffs bring suit, there will not be as much direct evidence on the question of causation as practicing scientists would hope to see before rendering a judgment about causation, because the studies that could, in theory, provide this information have quite simply never been conducted. The plaintiff might have a variety of suggestive pieces of data from a variety of fields—perhaps a mixture of animal studies, chemical structure evidence, toxicology, and epidemiology, though perhaps the epidemiological studies are based on populations dissimilar to the plaintiff, or exposure rates that differ dramatically, or perhaps look at different, but chemically related, substances to the one at issue in the case. This was the case, at least arguably, in *General Electric v. Joiner*, the second of the Supreme Court’s trilogy on expert evidence.33 It is quite often the case in toxic tort litigation that the quantum of data investigating the question of causation is simply less voluminous than one would like. Such was the case with Parlodel, a lactation-suppressing drug removed from the

---


33 See General Electric Co. v. Joiner, 522 U.S. 136, 143-45 (1997). The Court upheld the trial court’s decision to exclude the plaintiff’s expert evidence and grant summary judgment with respect to PCB exposure. Id. at 146-47.
market by the Food and Drug Administration when concerns emerged about whether it might be spurring strokes in some women who took it after childbirth.\textsuperscript{34}

Moreover, when a toxic tort claim involves a rare event rather than a potential mass tort, research on causation will almost certainly be scarce or even non-existent. In \textit{Zuchowicz v. United States}, for example, the plaintiff claimed that the negligent misprescription of an overdose of the drug Danocrine caused her to develop an extremely rare and often fatal illness, primary pulmonary hypertension (“PPH”).\textsuperscript{35} Quite apart from the fact that there were, for obvious reasons, no studies of whether overdoses of that drug caused PPH, there was quite scant evidence on the question of whether Danocrine was capable in the first place of causing PPH: nothing more than a set of speculative claims by a pharmacologist who had a theory about how Danocrine could cause a series of hormonal imbalances that could, in combination, cause physical problems leading to PPH, coupled with testimony by the plaintiff’s treating physician, a leading pulmonologist, whose conclusions about causation were based in significant part on expertise in \textit{other} drugs that cause pulmonary disease.\textsuperscript{36} The lack of other available evidence was in no way the plaintiff’s fault. The likelihood that extremely rare events will have been carefully studied is, well, extremely low. And even for toxic tort claims that affect a significant number of people, there may be quite limited evidence available when the first lawsuits are brought; indeed, it may often be the litigation and the ensuing publicity that spurs scientific interest in studying the question of causation more carefully.

In cases where the evidence supporting causation is more limited than one would wish it to be, the questions for the scientist are particularly difficult: How do you aggregate the variety of imperfect evidence into a conclusion about general causation? How do you assess the disparate items and make a judgment about the probability that the substance is capable of causing the harm at issue? Evidence synthesis is an especially


\textsuperscript{35} \textit{Zuchowicz v. United States}, 140 F.3d 381, 383 (2d Cir. 1998).

\textsuperscript{36} See id. at 385-86.
difficult and methodologically fraught area.\textsuperscript{37} There are certainly field norms about “hierarchies of evidence” (for example, randomly controlled trials, or even better, a meta-analysis of a number of different randomly controlled studies are typically thought of as the “gold standard,” and anecdotal case observations the weakest forms of evidence\textsuperscript{38}); and there are rules of thumb about how to assess the likelihood of causation by looking at a variety of factors. (The well-known Hill’s criteria, for example, in epidemiology, direct attention to matters like temporal relationship, dose-response relationship, biological plausibility, consistency of the observed correlation, and a variety of other factors.\textsuperscript{39}) But there are not, for the most part, terribly well-developed methodologies for quantitative synthesizing of disparate kinds and categories of evidence. Engaging in evidence synthesis, many scientists would agree, is as much an art as a science, inevitably involving methods that are not fully specified and the exercise of experience-based—and somewhat subjective—judgment.\textsuperscript{40}

Even when the evidence is of a similar kind (perhaps a variety of different epidemiological studies), and meta-analytic techniques for formally synthesizing the evidence might be possible, there must first be a determination about which evidence is valid enough to be worth considering and which is not. Reputable, talented scientists may well disagree in good faith about what evidence is worth counting and what evidence ought to be dismissed from consideration altogether for methodological flaws.

\textsuperscript{37} Although scientists in a variety of fields concern themselves with the dilemmas of evidence synthesis on a regular basis, there has not been a great deal of discussion of this issue in the legal literature. \textit{See generally COMMITTEE ON DAUBERT STANDARDS & COMMITTEE ON SCIENCE, TECHNOLOGY, AND LAW, NAT’L ACAD. SCI., DISCUSSION OF THE COMMITTEE ON DAUBERT STANDARDS: SUMMARY OF MEETINGS 11-16 (Kathi E. Hanna & Anne-Marie Mazza, rapporteurs, 2006).}

\textsuperscript{38} \textit{See, e.g.,} ROBERT J. LEVINE, ETHICS AND REGULATION OF CLINICAL RESEARCH 211 (2d ed. 1986); Robin Harbour & Juliet Miller, \textit{A New System for Grading Recommendations in Evidence Based Guidelines}, 323 BRIT. MED. J. 334, 334-36 (2001).

\textsuperscript{39} Austin Bradford Hill, \textit{The Environment and Disease: Association or Causation?}, 58 PROCE. ROYAL SOC’Y MED. 295, 295-305 (1965).

\textsuperscript{40} To recognize that complex tasks of evidence synthesis have an inevitable subjective component using current methodologies is not to disparage the efforts by scientists to engage in such synthesis. There are, to be sure, significant efforts of this kind, especially in medicine, efforts to put together systematic reviews of all that is known and to draw conclusions from them in order to influence clinical practice. The Cochrane Collaboration is one of the best known and most respected of such efforts. \textit{See generally} http://www.cochrane.org/docs/descrip.htm (last visited Feb. 6, 2007).
Take, for example, the important public health question of whether regular mammograms for all women over the age of forty can help to reduce breast cancer mortality rates. This question arose not in the context of litigation, but as a critical public health issue with implications for what advice doctors should give to their female patients. Two groups of scientists carefully studied all the (considerable) available data on the subject. But each group’s decisions about what data warranted consideration differed. A review by Danish researchers decided that many of the studies had been too methodologically flawed to warrant consideration, and thus they based their analysis on a more limited number of studies that were deemed adequate; the other review, completed by the United States Preventative Task Force, agreed that many of the studies were flawed, but determined that the studies it deemed only “fair” were not so poor in quality that they should be altogether excluded from consideration. Because of these divergences in what evidence was deemed worthy of consideration, the two studies reached quite disparate conclusions. The first analysis found that women in their forties do not, in the aggregate, benefit from mammograms and in fact have increased risk of harms because of unnecessary treatments and surgeries that mammogram results generate. The second study, by contrast, found that mammography did reduce mortality and was, on balance, beneficial. How could qualified scientists disagree about which studies were even worthy of consideration? As epidemiologist Steve Goodman wrote in an editorial on the controversy:

Judgment determines what evidence is admissible and how strongly to weigh different forms of admissible evidence. When there is consensus on these judgments and the data are strong, an illusion is created that the evidence is speaking for itself and that the methods are objective. But this episode should raise awareness that judgment cannot be exercised from the process of evidence synthesis and that the variation of this judgment among experts generates uncertainty.

---

42 Id. at 363.
43 Id.
44 See id.; see also Gina Kolata, New Mammogram Studies Divided on Benefits, N.Y. TIMES, Sept. 3, 2002.
just as real as the probabilistic uncertainty of statistical calculations.45

What does this public health debate have to do with experts in court and the limitations of neutral experts? It provides a dramatic illustration of a generalizeable point: while the partisanship of experts may create the illusion of disagreement even when little exists in the broader community, even in situations far removed from the use of “hired guns” in court, significant interpretive disagreements can occur among scientists operating in good faith—and this may be so even in instances when the available quantity of data is unusually substantial. Evidence synthesis is an especially complex and fraught area, one in which reputable scientists may simply disagree about the extent to which an imperfect body of data justifies an inference of causation. While partisanship may exacerbate these differences, and the lure of high pay may risk creating the appearance of disagreement when it would be unlikely to exist outside of the courtroom, the converse is simply not true: interpretive disagreement is not necessarily the result of partisanship. It may well be the product of genuine methodological and interpretive differences, not only across scientific disciplines, but even within them.

Evidence synthesis, to be sure, is simply one salient illustration of a still more general point: scientific disagreements are not, in and of themselves, a sign that something is amiss, nor do they necessarily suggest that one or both parties to the dispute are misbehaving partisans. Quite the contrary; disagreement is an integral part of scientific processes. What this means, however, is that the use of neutral experts may bring with it significant risks.46

To be sure, in those cases in which one side’s experts are truly charlatans or have been led by partisan zeal to dramatically overstate some aspects of their testimony, neutrals could offer an effective and welcome check. But what would the use of neutrals on a more regular basis mean in those cases in which the disagreements among experts reflect legitimate differences, differences that the parties’ experts would hold equally fast to outside of the context of litigation or

45 Goodman, supra note 41, at 364.
even without the incentive of receiving fees? In such an instance, a neutral expert would end up doing one of two things. Perhaps she would support one or the other of the party’s positions. This might create for the factfinder the appearance of a consensus view, but this appearance would be illusory. In such circumstances, the jury would potentially be misled into being unduly influenced by the structurally unbiased expert. The jury in such cases might struggle less with the diverging views of the parties’ experts and simply choose to rely on the “neutral” expert; while its job might therefore be perceived as easier, it is not at all clear that the quality of its decisions would be meaningfully improved.

Alternatively, the court-appointed expert might lay out the scientific terrain for the factfinder and situate the dispute, without taking sides at all. This could potentially be educational for the jury, and perhaps a court appointed expert would be better suited to elucidating the contours of the debate than the party-selected witnesses. But in the end, the jury would be left in virtually the same place it was before the court-appointed expert assisted it: needing to decide which expert to believe while lacking the epistemic qualifications to assess the merits of the testimony.

Those who call for neutral experts, then, at least partly misunderstand the nature of scientific disputes. For whenever there is a legitimate scientific disagreement at issue in a legal case, a neutral expert would either mask a legitimate dispute or else be unable to offer “those general truths, applicable to the issue, which they may treat as final and decisive,” for which Learned Hand and others have long craved. In other words, while neutrals might indeed offer a useful, strong check on extreme partisanship, this would often be an insufficient solution precisely because of the factfinder’s continued lack of epistemic competence.


\[48\] Hand, supra note 28, at 55; see Allen & Miller, supra note 47, at 1133.
B. The Problem of Epistemic Competence

Let us turn, then, directly to the question of epistemic competence. Might there be a way to solve this problem? My focus, again, shall be on the most obvious potential solution: attempting to make use of decision-makers who themselves have epistemic competence. How might we do this? One option would of course be juries made up of experts.49

Even putting aside any potential constitutional objections (for example, would a jury of experts still be a jury of one’s peers? would it meet the requirement that a jury come from “a fair cross section” of the community?), specialized juries would quickly run up against new difficulties. Just consider trying to operationalize a system of special juries designed to deal with concerns about epistemic competence. Who would be on such a jury? Imagine a toxic torts case involving a plaintiff alleging harm resulting from the ingestion of a pharmaceutical. Imagine that the key legal question in the case is causation, as it so often is in such cases, and imagine further that the plaintiff has evidence relating to causation from a variety of sources and scientific disciplines. Let us posit that she has two epidemiological studies, some evidence from chemical studies of the drug and related substances, an animal study or two, and extensive evaluations by several physicians who endeavored to find the cause of her ailments through “differential diagnosis.”

If our goal is a decision-maker with epistemic competence, who should be eligible to sit on the jury for such a case? Just how much should a potential juror have to know about the scientific disciplines from which the evidence will come? Should the jury be limited to physicians and professional research scientists? If so, scientists from what disciplines?

Should a paleontologist be permitted, or a theoretical physicist? What about social scientists? Clearly, the boundary-drawing issues would become immediately significant. An empirical social scientist might understand the epidemiological evidence at least as well as many physicians. Moreover, the evidence presented in the case could be disparate enough that in many instances no one at all would be a true epistemic insider to all of the scientific evidence offered. Depending on the matters at issue, the pool of people truly expert in any of the relevant areas might be rather small, and the pool of people expert in all of them might not even exist. Perhaps an epistemically competent jury need not mean that every juror has epistemic competence in every expert area at issue; we might be satisfied with a jury made up of several leading members of each of the subfields in which significant expert evidence was expected.

As a thought experiment, imagining such a jury is an interesting prospect. But, in reality, it would raise enormous practical hurdles. It is simply not realistic to bring the leading experts in as jurors time after time. Certain kinds of expertise arise in trials over and over again, and it is likely that the segment of the population with these forms of expertise would become massively and unequally burdened by their jury obligations. Just how often could we ask busy epidemiologists and toxicologists, for example, to serve on juries? Meeting their civic duty too often could have devastating effects on both their income and their career! In addition, scientific sub-communities can be small and professionally interconnected, so it is likely that some of the most epistemically qualified non-testifying experts would know personally, and have views about, some of the testifying experts, or even have a prior opinion about the particular matter at issue. While these prior views are a direct consequence of the fact of their expertise, they might generate serious problems vis-à-vis our expectations of jurors: an expert with a significant degree of prior relevant knowledge might be unable to hear the evidence presented with a fresh and open mind. Epistemic competence might go hand in hand with preexisting judgments about the merits.

One might argue that the preceding discussion unreasonably overstates the necessary degree of epistemic competence. Perhaps it is not necessary to be a scientist within the relevant field in order to have enough knowledge and training to evaluate the claims made in the courtroom in a rational manner. Perhaps the preceding discussion fails to
recognize the distinction between an intelligent producer and an intelligent consumer of knowledge—the skills necessary to evaluate a claim might be significantly lower than those needed to be a substantive contributor to the debates of the field. Even if we assume this to be true, now who could be included in our epistemically competent jury? If the case involves a good deal of sophisticated statistical analysis, how much familiarity should the jury have to have with statistical thinking in order to be epistemically competent? Some graduate training? A college major in statistics, math, accounting, or some related quantitative field? Successful completion of a college course on the subject? Successful completion of an in-court quiz testing statistical knowledge, or basic numeracy?

Our problem here is a classic dilemma of boundary-drawing. Where do we draw the line between those deemed adequately epistemically competent and those who are not? If epistemic competence is tied with a reasonably high degree of precision to the matters at issue in the case, our jury system would become literally unworkable. By contrast, one could probably implement a system that required jurors to have some undergraduate level science training for cases involving expert scientific evidence. But this would create other problems. No longer would there be a direct tie between the factfinder’s knowledge and experience and the central issues in the case. Once that direct tie is broken, a system of this sort both smacks of elitism and begins to look distressingly anti-democratic. How confident are we that those with a minimum of one science class in college would actually be better, as a group, than those who did not meet the standard? How would they compare to those who went to college but did not take science? To those who excelled in some high school science class? To those without a college education but whose present employment relates to science or technology? To those with a particular I.Q., regardless of education?

Unless we have (and perhaps even if we did have) a well-grounded empirical basis for believing that jurors who met a particular standard for prior experience with scientific matters would reach significantly better decisions than those who lacked the relevant experience, we might well think that other public values and commitments should prevent us from
heading down that path. In addition, those jurors who met our epistemic criteria (whatever they were) might be, as a group, demographically different from those who did not, and in ways that might be troubling, not only along gender, race, or ethnicity lines, but also in terms of beliefs relevant to the case—ideas about politics, notions of fault and liability, or other less obvious dimensions. We might be consciously giving up diversity on one dimension (for example, eliminating all those without a certain degree of scientific background), and simultaneously giving up diversity along dimensions about which we were not even aware.

If expert juries may raise difficult concerns, expert judges might offer a less troubling method to ratchet up the legal system's degree of epistemic competence. Judges who develop a specialty might be better positioned to assess the expert evidence adduced by parties and to guide both the parties and the jury through the trial process. Although no state, to my knowledge, has a court dedicated to cases involving complex expert evidence in particular, several states are experimenting with specialty courts or specialized judges within general courts devoted to business disputes or to complex civil litigation, and this category of case does typically involve a good deal of expert testimony.

The early response to these innovations appears to be generally positive. However, they cannot be seen as a particularly robust response to the issue of epistemic competence. Even if we imagine (though it is far from certain)

50 Delaware has, in fact, a rarely used statute permitting the use of "special juries" in cases involving complex litigation. See 10 Del. Code Ann. tit. 10, § 4506 (2007). The statute is not specific about who counts as a "special juror" or what cases may qualify for this provision. Id. For discussion of this statute, see Oldham, History, supra note 49.


52 See, e.g., Kiesel & Borys, supra note 51, at 20-21; Ronald M. George, Complex Civil Litigation Pilot a Success, Court News (Judicial Council of Calif.), Sept. 2001, at 2.
that the judges in a court devoted to complex civil litigation increase their sophistication with respect to certain kinds of repeat-play evidence, the fact remains that in a jury trial, the judge does no more than rule on admissibility. The jury still must evaluate the evidence, and the fact that the judge may have grown to be a more sophisticated consumer of the expert evidence at issue than the typical judge will not necessarily translate in any obvious way to increased juror sophistication. To put it differently, at best these courts might lead to an improved set of evidentiary inputs for the jury’s consideration—improved in the sense of being more likely to be epistemically valid. But the extent to which that would translate into better outputs is not obvious, given the jury’s own lack of epistemic competence. It probably cannot hurt, but it might not help much either. Bench trials with specialized judges would take this a step further—many have suggested, for example, that a “complexity exception” to the constitutional right to the jury trial ought to be permitted.53

Could we go one step further and imagine a neutral, epistemically qualified decision-maker? Could we imagine a procedure akin to a bench trial, but in which the adjudicator was not simply a repeat-player judge in a specialized court, but was in fact an epistemic expert in the matters at issue? Of course we can imagine it, but at this point, for better or for worse, we are describing an adjudicatory regime that looks very little like our jury system. As it happens, however, we do have examples of precisely such a procedure in actual use: Some arbitration proceedings make use of an industry expert as an arbitrator in lieu of someone with legal expertise.54 For disputes with a high degree of technical complexity, or where industry norms are explicitly at issue, it is not uncommon for parties to elect to make use of an epistemically qualified decision-maker.55 The parties might choose someone with the

53 See sources cited supra note 51.
55 Note, however, that Bernstein found that at least in some industries the expert arbitrators made surprisingly little use of their insider knowledge of business norms within the field. Bernstein, Merchant Law, supra note 54, at 1771-87.
appropriate scientific or technical background rather than someone with a legal background. In a sense, then, for commercial and contracts disputes, or any other kinds of disputes for which arbitration is a viable alternative, there already exists an “opt-in” approach when parties deem an epistemically-competent evaluator to be an especially high priority. But while it would be interesting to know more about how often, and in what circumstances, parties select scientific or technical know-how over a legal background, this approach is likely to be attractive to both parties in only a limited number of cases.

CONCLUSION

Where, then, are we left? Not with solutions, to be sure, but perhaps with a clearer diagnosis of the dilemmas surrounding the use of expert evidence within an adversary system, and their tradeoffs. And perhaps we are left, too, with a bit more sympathy for Daubert, or at least a recognition that so long as we have our adversarial system in much its present form, we are inevitably going to be stuck with approaches to expert evidence that are imperfect, conceptually unsatisfying, and awkward. It may well be that the real lesson is this: those who believe that we might ever fully resolve—rather than imperfectly manage—the deep structural tensions surrounding both partisanship and epistemic competence that permeate the use of scientific evidence within our legal system are almost certainly destined for disappointment. This ought not to lead us to quiescence. It ought instead to guide us to a certain degree of realism and modesty about how much we can change about the use of expert evidence, unless we are prepared to make fundamental modifications to our adversarial system.
Getting to the Truth

GROUNDING INCOMPLETE KNOWLEDGE

Frank C. Keil†

One aspect of truth concerns knowing when to trust others when one’s own knowledge is inadequate. This is an ever more common problem in societies where technological and scientific change seems to be constantly accelerating. There is an increasing need to rely on the expertise of others and consequently to know when others are more likely to be offering an objective opinion as opposed to a biased one. Here, I argue that there are systematic and early emerging cognitive heuristics and biases that profoundly influence our patterns of deference, our ways of assessing expertise, and our sense of when testimony is to be trusted. For the most part, the power and pervasiveness of these biases are ignored or greatly underappreciated. These biases and heuristics can both mislead and inform our understanding and use of others’ expertise; it is therefore critical that we acknowledge their presence and know how to work with them.

As will be seen shortly, people tend to make serious mistakes in their evaluations of both their own knowledge and the complexity of systems. Even worse, when people do recognize that their understanding is inadequate, they can make surprising mistakes in guessing who the right expert is to fill out the details. In other cases, however, adults and even young children can accurately figure out where appropriate expertise lies. The details of this story, as described later in this paper, are central to any full account of how we know when to trust others. We have many tools that can be used to help decide when it is appropriate to defer to another’s area of expertise and when it is better to have serious doubts; unfortunately, we often do not use these tools effectively.

† Professor of Psychology and Linguistics and Master of Morse College at Yale University. Much of the research described in this paper was funded by NIH grant- R37- HD023922 to Frank Keil.
The issue of establishing trust will be considered in a series of five sections. Part I will explore the problem of unbounded causal complexity, namely, the need for ways to construct explanatory gists of causal systems that are far too complex for any one person to fully understand. Part II will ask about the ability to assess the quality of one’s own knowledge and will argue that there is a strong tendency for people to overestimate the depth and quality of their explanatory understanding of a wide range of devices and natural phenomena. Part III will consider a related phenomenon in which people are shown to have misleading illusions of insight when explanations are supplemented with certain kinds of irrelevant, but nonetheless compelling, support. Part IV will suggest that illusions of explanatory depth and illusions of insight may be related to errors in underestimating just how complicated various phenomena are, especially those related to the social sciences. Part V will consider how people manage to get by with incomplete knowledge, suggesting that there are several cognitive tools and heuristics that are used to help fill in the gaps in one’s own knowledge. Finally, Part VI will discuss how all these factors converge to explain how people are able to establish trust and what errors are likely to occur.

I. THE PROBLEM OF UNBOUNDED CAUSAL COMPLEXITY

Virtually any phenomenon in the natural or artificial world has seemingly unbounded levels of causal complexity. Ask a simple question, such as what stem cells are and how they work, and the answer can be expanded on in ever deeper and more complex ways. Stem cells may initially be described as cells that have the potential to become any one of the many cell types within an organism. A request for more detail may reveal broad classes of cell types that can be created (for example, ectodermal, mesodermal, or endodermal). Deeper answers can reveal ever more fine-grained descriptions of the different cell types. Still deeper answers reveal how all these cell types are related in a hierarchical structure. All of this information can then be elaborated on in enormous detail with respect to the signaling mechanisms that cause cells to differentiate in one manner as opposed to another, mechanisms that can involve intricate accounts of molecular pathways that regulate various sets of genes. If one is driven to gain the deepest explanation possible, one gradually gets exposed to much of developmental biology and molecular biology, as well
as areas of chemistry, physics, and even other more distant fields.

The same holds for questions about the world of human inventions, such as how helicopters fly, how drugs influence the body, or how a resolution of patent dispute might influence the economies of several nations. In all such cases it is quickly apparent that no one person can possibly know the full answer to each question. One must know how to seek out legitimate experts and how to avoid illegitimate ones. One must also know when experts are making a statement within their range of competence as opposed to outside that range. All of us have huge gaps in our explanatory understandings that we tolerate because we think they are firmly grounded in legitimate understandings in other minds, understandings on which we can rely when necessary.

These issues are highly salient in the law. The courts, for example, frequently allow the testimony of expert witnesses, assuming that there are reliable ways to identify such experts and use them to improve the quality of legal proceedings and decisions. The complexity and distributed nature of knowledge requires that we have well-worked-out and reliable mechanisms for grasping the division of cognitive labor and that we know how to identify appropriate areas of expertise and appropriate experts within those areas. In addition, we need to know when experts are likely to be more or less trustworthy. The cognitive science of such processes is critically relevant to evaluating how well such patterns of deference, consultation, and trust work; yet it has been largely neglected in the law.

II. HOW WELL DO WE KNOW THAT WE DON’T UNDERSTAND?

To assess how well people deal with causal complexity and the need to consult others, one must first know when one is in over one’s head. One must have a way of sensing when there are gaps in one’s knowledge that make one’s understanding so flawed that it is inadequate for use in a task. This problem of knowing how little one knows occurs in every facet of life. A student may not realize that she does not really understand an explanation offered by a teacher; a businessman may not know that he has misunderstood a critical clause in a contract; and members of a jury may think they understand the details of a
complex patent dispute when in fact they have overlooked a critical underlying principle.

Most people are quite inept at estimating how well they understand various everyday phenomena, showing a strong tendency to assume they understand how the world works in far more detail than they really do. We have called this effect an “illusion of explanatory depth” (“IOED”). The IOED can be quickly revealed in tasks where people are taught how to rate causal understanding on a scale that ranges from one (virtually no understanding beyond a vague sense of what a phenomenon appears like) to seven (a full mechanistic understanding of exactly how a device works or how a natural phenomenon occurs). Even though people can be shown to clearly understand the scale and use it correctly when rating sample texts of explanations, they give much higher scores than they should when asked to estimate their own explanatory understandings.

The effect is the strongest for estimates of one’s causal understanding. For example, when people are asked to rate how well they think they understand how a helicopter flies, they might give a rating of five or six. Yet we can show that this rating is far too high by simply asking them to actually write out, or verbally report, everything they know about how a helicopter flies. After giving the fullest explanation they can, people are again asked to rate their knowledge in light of that explanation. Their ratings on this second occasion show a sharp drop relative to their first ratings, with people routinely evincing great surprise at how little they actually knew. Matters get worse if they are asked a simple diagnostic question (for example, how does a helicopter go from hovering to moving forward?) and are then asked to re-rate their knowledge in light of their ability, or inability, to answer that question. People again show a considerable drop relative to their initial rating and even to their second rating. Their ratings stay low if they are shown a concise but detailed explanation and asked to compare it to their own initial understanding and then use that comparison as a basis for assigning a score to their initial rating. In other words, people consistently rate their initial guesses of how well they

understood various phenomena as being far too high when their real knowledge is put to the test.

These effects are substantial and happen for ratings of understandings of both artifacts, such as helicopters, and natural phenomena and systems, such as the workings of the heart. In contrast, when asked to rate self-knowledge of facts (for example, the capitals of relatively obscure countries) or of procedures (such as how to make an international phone call) or even of narratives (such as the plot of a well-known movie), people tend to be much better calibrated, either being fully accurate or overestimating the depth of their knowledge to a much lesser extent.²

The selectivity of the IOED has strong implications for drawing conclusions about when a person is likely to be accurate or inaccurate with respect to claims about the quality of their own knowledge. In some domains that do not require much underlying causal structure, assuming that people are acting in good faith, they are likely to be quite accurate. In others, however, such as ratings of causal explanatory understanding, even the most sincere and trustworthy people are likely to systematically overestimate what they really know. The extent of the IOED can be quite remarkable even for extremely common everyday objects. In one series of recent studies people were shown sets of simple drawings of bicycles. One drawing was correct while others had the chain connected to sprockets on both wheels in a manner that would make steering impossible or had the chain disconnected from the drive wheel. Adults had great difficulty picking out the correct drawing from among three incorrect ones, even though they all said they were familiar with bicycles and often rode them. They showed the same patterns of major errors when shown bicycle frames of which only one could functionally work. Even members of an active bicycle club made substantial numbers of errors.³ The commonplace nature of some devices does not assure that people will know how well they understand them. Indeed, in at least some cases, high familiarity may breed an especially high overconfidence. It may be that when someone can easily use something, he or she confuses that sense of ease with a sense of causal understanding.

² Rozenblit & Keil, supra note 1, at 533.
Experts may also fall prey to the IOED in their highly refined areas of expertise. They may be somewhat better calibrated, but, especially in areas that are at the boundaries or “leading edge” of their own understanding, they may sincerely believe that their understanding is deeper than it is. This seems to be particularly true in cases where experts are assessing how well they understand something, as opposed to assessing how well they know how to do a procedure. In such rarified areas, novices who rely on these experts may have an especially difficult time evaluating an expert’s sincere, but misguided, claims to having deep explanatory understanding. As will be shown, however, even when a layperson is at such a disadvantage in the relative size of the knowledge base, there still are ways to adjust confidence in an expert’s testimony.

The IOED holds not only for assessments of one’s knowledge, but also for assessments of knowledge in other minds. Thus, recent work in our laboratory indicates that ratings of how well others understand various phenomena or devices are equally likely to be inflated. This has important implications for knowing when and to whom one should defer. Thus, even if one comes to a realization that one does not understand, such a person may systematically overrate the likelihood that another person does. The IOED also tends to get stronger the less one knows or the younger one is. Thus, if a person has a very high level of ignorance in an area, that person is especially liable to not know when intellectual outsourcing, namely, the reliance on others to provide complex explanations, is needed. The greater extent of the phenomenon in children raises questions about how to evaluate their claims that they understand the reason for an action or event. Similarly, children may be especially vulnerable to knowing how well they have understood instructions or explanations that are provided to them in a courtroom.

Across numerous studies, we see several converging cognitive mechanisms that seem to set up and maintain an

---

4 Rozenblit & Keil, supra note 1, at 554-55.
especially strong IOED for explanatory understanding. People tend to confuse higher-level functional glosses on a system (such as knowing that turning a key starts a car) with lower-level mechanistic understanding (such as understanding the complex starting circuitry in modern vehicles). They also tend to confuse explanations that they are able to piece together in an ad hoc manner with pre-stored explanations that they bring to a situation, underestimating the degree to which they generate explanations “on the fly” as opposed to bringing them preformed to a situation. In addition, because people rarely give exhaustive explanations, they have little practice in evaluating explanations for completeness. These and other factors help make the IOED robust and difficult to attenuate.

III. FALSE EXPLANATORY INSIGHTS

There is a related cognitive bias to the IOED that documents ways in which people can have a false rush of explanatory insight when in fact none was actually achieved. Certain kinds of information can be associated with explanations that make them much more appealing than is appropriate. One example with increasing relevance for the law is the use of functional magnetic resonance imaging (“fMRI”) data to make claims about neural functions, or about disruptions in neural function that might arise from various forms of brain damage. Many court cases have allowed extensive testimony in which fMRI findings are alleged to explain why a person behaved as he or she did or why an individual clearly has suffered brain damage arising from the negligence of others. In many cases, fMRI findings might well provide useful additional information in legal procedures, but in many others they can have a powerful and often unrecognized ability to mislead.

In particular, fMRI results can create a false sense of insight when they are in fact completely irrelevant to the

---

7 See Frank C. Keil, Explanation and Understanding, 57 ANN. REV. PSYCHOL 227 (2006); Frank C. Keil, Doubt, Deference and Deliberation: Understanding the Division of Cognitive Labor, in 1 OXFORD STUDIES IN EPISTEMOLOGY 143, 163-64 (Tamar Szabo Gendler & John Hawthorne eds., 2006); Rozenblit & Keil, supra note 1, at 552-56.

quality of an explanation in which they are embedded. Thus, it can be more difficult to detect weak or flawed research when it is accompanied by uninformative fMRI results than it is to detect equally weak or flawed research without such results. To demonstrate such an effect experimentally, one can present adult experimental participants with one of four explanations, which are created by varying the explanations along two dimensions: whether the explanation is good or empty and circular, and whether or not it contains irrelevant fMRI results. Student participants clearly preferred the good explanations to the empty/circular ones when there were no fMRI results accompanying those explanations. In contrast, they found it much more difficult to tell the good from the bad explanations when they also contained the fMRI results, even though the fMRI results were completely noninformative.9 Only highly trained experts in cognitive neuroscience showed the same abilities to discriminate good from bad explanations when they were accompanied by neuroimaging results.10 The neuroimaging results were not particularly complicated; however, it appeared that phrases such as “brain scans showed that” made it much harder for most people to then evaluate that the brain scans added no new information. Because the experts were easily able to see the good from the bad experiments with the neuroimaging results, we know that the neuroimaging really did not add any new information of value.

It seems that we can gain misleading senses of insight when studies are made more concrete through such means as brain imaging, even when the imaging is irrelevant. In deciding whether fMRI results should be admitted into evidence, it is critical to carefully look at what additional insights they really offer, as opposed to relying on simple intuitions that the fMRI supplemented explanations seem more “solid.” More broadly, neuroimaging certainly has an important role to play in understanding behavior, but the potential for abuse needs to be recognized.11 When people try to understand the many invisible factors leading to behaviors,

9 Deena Skolnick-Weisberg et al., The Seductive Allure of Neuroscience Explanations, J. COGNITIVE NEUROSCI. (in press).
10 Id.
such as a criminal act, there is a tendency to seize on any claims about how the basis for such behaviors might be physically instantiated. It seems likely that comparable effects may be found for other alleged physical bases for behavior, such as claims that there is “a gene” causing a certain behavior.

In short, people can have rushes of insight as they learn more about a phenomenon, but sometimes these rushes do not correspond to real increases in understanding. In particular, when people are offered information that is phrased in terms of a lower, more concrete, level of analysis, they often will take that information quite uncritically and think that it adds value because it is lower level and not because it actually provides any new explanatory information. Laypeople should be especially vigilant when provided with explanations that shift down levels in some kind of reductionist hierarchy, keeping in mind that things often sound better than they really are when such a shift occurs.

IV. DO WE KNOW WHERE COMPLEXITY LIES BENEATH?

Another problem related to assessing truth concerns estimating the complexity of various systems and phenomena. Even if people were fairly accurate at sensing the degree of comprehensiveness of their own knowledge, they might get into considerable trouble because they do not understand the level of complexity of a phenomenon; they assume that their relatively simple understanding is all that is needed because they underestimate the actual complexity of the phenomenon. Difficulties here lead directly to problems of trust and deference.

Knowing when we need to defer to another’s expertise is intimately related to how complex we think the phenomenon being explained is. If the phenomenon seems trivial and relatively self-evident to any reasonably observant and thoughtful person, then it may be inappropriate to bring in an expert who might only muddy the waters rather than shed insight. We can all envision such trivial cases. Suppose, for example, a defendant drove through a school zone at 100 mph and asked for clemency on the grounds that he didn’t know that such a high speed posed a risk for school children. It is reasonable to argue that there is no need for traffic experts to be brought into such a case to testify that the risk level was increased. Moreover, it certainly makes sense for the courts to
exercise discretion in allowing people to testify as expert witnesses. Without such discretion, it is easy to see how the introduction of experts could lead to delaying tactics and obfuscation in a courtroom.

Other cases, however, may seem far simpler than they really are. In particular, experts on psychological mechanisms have often been disallowed in the courts on the grounds the expertise is simply “common knowledge” and that the alleged expert has nothing to offer that a reasonable layperson does not already know. For several decades psychological experts have been excluded in cases abroad and in the United States. One especially prominent recent case concerned the perjury accusations against I. “Scooter” Libby. Libby’s defense team wanted to bring in experts on memory to testify that it was quite plausible that Libby could have misremembered past events, rather than deliberately committing perjury. Judge Reggie B. Walton, however, disallowed such experts partly on the grounds that laypeople had accurate and detailed understandings of the fallibility of memory; it was not clear what additional insights could be added by “experts.”

It is beyond the scope of this paper to document the many complex features of the human memory system and how they can cause behaviors that may surprise the layperson and even be quite counterintuitive to lay theories of how memory works. To use just one example, many people think that memory is laid down almost like a videotape recording that may become buried but is always present. Such a view grossly underestimates the extent to which memories are constantly being revised and reconstructed in ways that can completely overwrite the earlier version. The public fascination with “recovered memories” is one example of this bias. Although

---


15 Seema L. Clifasefi et al., Setting the Record (or Video Camera) Straight on Memory: The Video Camera Model of Memory and Other Memory Myths, in TALL TALES ABOUT THE MIND AND BRAIN 60, 60-61 (Sergio Della Sala ed., 2007).
laypeople often believe that there exist exact traces of memories of childhood experiences that can be brought to light with appropriate “excavations” by therapists, the experimental evidence for such recoveries is minimal. Putting aside the merits of Judge Walton’s decision, the question arises as to whether judges might often be prejudiced against the possibility of psychological explanations having a complexity comparable to that which one might find in areas such as biology or physics.

In adults, it is very difficult to know how to measure the relative complexity of two domains, as complexity itself can vary in so many ways. For example, if asked whether a full explanation of superconductivity or face recognition is more complicated, one could easily pick either alternative by emphasizing different facets of the phenomena. Perhaps superconductivity is more complicated because it requires a particular kind of mathematics, or perhaps face recognition is more complicated because it requires integration across several different forms of mathematics as well as anatomy. Thus, it may seem impossible to distinguish between claims that people are biased to think of psychology as simpler than other sciences and claims that psychology really is simpler. Recently, however, studies have been conducted that suggest a cognitive bias is at work.¹⁶

These studies rely on the assumption that such biases might arise from very basic intuitions that emerge early in childhood, intuitions that might gradually become weaker with age by compensating knowledge. Thus, there might be a rough and ready intuition that psychological phenomena are simpler, an intuition that gets tempered with age through experience with the complexity of actual psychological situations. To examine this possibility, one study took a large number of everyday phenomena in domains such as physics, biology and psychology. The phenomena were pre-tested to find a subset that adults judged as all about equal in complexity. For example, the following “why” questions were judged to be equally complex: How does a top stay spinning upright? How does your skin heal after it has been cut? Why is it hard to understand two people talking at once? Children, ranging in age from five to fourteen years, were then asked to make

judgments about the complexity of these same selected phenomena.\(^{17}\) As the age of children decreased, a stronger bias emerged in which psychological phenomena were judged as much simpler than those in the biological and physical sciences.\(^{18}\) This pattern strongly suggests that the bias may persist in adults, but in a reduced form. Moreover, it may re-emerge in stronger terms in adults when they must also evaluate several other complex factors at the same time, such as in a legal case. Similarly, this pattern may emerge in time pressured situations, rapid verbal exchanges, or under pressing questions. Thus, even though the stimuli questions were judged by adults to be of equal complexity, the psychological ones may actually be more complex, with the difference being masked by a latent bias that still persists in a weaker form from childhood. It might be possible to show this remaining difference in complexity by contrasting detailed expert explanations of psychological and natural science phenomena that are judged to be of equal complexity by laypeople.

Children in these studies often attempted to explain their intuitions by referring to the immediacy of psychological states. They experienced them automatically and effortlessly, and that feeling may be confused with a sense that the phenomena are therefore simple. They also often said that everyone could understand such phenomena, suggesting something quite close to the “common knowledge” claims made by judges. Common knowledge may, in fact, not be well understood.

In short, it is not safe to assume that one’s novice intuitions about the complexity of phenomena are always accurate. There may be systematic biases that heavily distort one’s intuitions into thinking some classes of phenomena are much simpler than they really are. There is also the possibility that other biases may create intuitions of greater complexity than is warranted, although such patterns still remain to be experimentally demonstrated. One such example might involve a system that has a vast number of parts and seems hopelessly complex, but which is suddenly rendered very simple when a basic repeating pattern is pointed out. Some fractal patterns can take on great simplicity when seen in such a light.

\(^{17}\) Note that all the questions were phrased in very simple terms that would be readily understandable to young children.

\(^{18}\) Id.
It would be extremely useful in legal cases to have a greater awareness of complexity biases and to factor them in when making decisions about whether to allow certain kinds of expert witnesses. It does not follow that all claims of expertise should be allowed, or even that all legitimate areas of expertise are relevant to the case at hand; but it is evident that we do not currently have in place sufficiently rigorous and systematic ways of evaluating whether a phenomenon requires testimony from experts in order for jurors and judges to act in a more informed manner.

V. GETTING BY WITH LESS

The story so far seems pretty discouraging. Normal adults, as well as attorneys and judges, labor under several cognitive biases that could powerfully influence trust and deference. They overestimate their own explanatory understanding, they get false surges of insight from irrelevant sources of seductive information, and they may have heavily distorted senses of what phenomena are truly complex and require expert assistance and deference to those experts. How might these problems be addressed? There are two answers, one arising from trying to instill a mindfulness about these biases in the legal system, and the other arising from ways in which complete novices do have tools at their disposal that enable them to evaluate expert claims, even when they understand very little about the details of those claims.

A. Mindfulness

In the case of mindfulness, legal cases might well be helped by a simple awareness not only of the fallibility of our knowledge related judgments, but also of the details of how those judgments are distorted. All parties in a legal proceeding should constantly be asking if they might be systematically overestimating not only their own understandings, but also those of others. This might entail techniques to carefully examine levels of understanding in ways that are thorough without being badgering or intimidating. There may be a set of heuristics that one could apply before making statements about the depth of one's own understanding or of that held by others. For example, one might be required to write out complete explanations of some facet of a phenomenon or to answer
certain critical diagnostic questions posed by top experts in the field.

Similarly, one should not assume that it is an easy matter to ascertain whether an area of expertise is legitimate or merely common knowledge. It would be a mistake to clutter the courts with an endless parade of experts on every possible topic, but it would be an equally large mistake to not try to develop explicit sets of procedures for evaluating alleged areas of expertise. If, for example, people have a tendency to succumb to reductionist explanations, even when they are empty, it might be possible to present them with examples of noninformative reductionist explanations to illustrate common pitfalls. Alternatively, one could suggest that people specifically ask what value is added when a new piece of information is introduced at a different level of analysis; for example, asking what predictions the new information now makes possible that were previously impossible or more difficult before receiving the new information. It is unclear what mindfulness strategies may be most effective simply because research in this area is relatively new. There is a great deal of cognitive science to be done in this area as well, but the knowledge gained so far can certainly suggest some guidelines.

B. Implicit Tools for Evaluating Expertise

Everyone shares certain heuristics that can be brought to bear to assess both the credibility and relevance of experts and the appropriateness of judgments, heuristics that have roots in early childhood. For example, young school children are more likely to doubt the testimony of people whose statements are self-serving. Thus, a person who claims to have won a close race is more likely to be doubted than one who claims to have lost the same race.19 This may seem blindingly obvious, but it is not clear how often people are provided with adequate information about the potential links between a person’s testimony and their own motivational states. For example, only recently have several major newspapers adopted the practice of disclosing the funding sources behind new discoveries in biomedical research. This practice undoubtedly was influenced by studies showing, for example, that scientists funded by pharmaceutical companies have quite different, and

19 Mills & Keil, The Development of Cynicism, supra note 5, at 389.
usually much more company-self-serving, results than scientists funded by government grants. Indeed, drugs have been reported as twenty times more effective in supposedly objective trials when the studies are funded by drug companies as opposed to when they are funded by federal agencies. One cannot rely on the researchers conducting such studies to monitor themselves, as even the most ethical may unknowingly introduce bias. Whenever information that has been gathered by others is presented, it is important to know how that information relates to the motivational states of others.

A second form of implicit knowledge relative to the evaluation of expertise concerns having a sense that a domain of knowledge is appropriate for expertise even when one does not grasp most of the details of that domain. Here, we can consider some forms of implicit knowledge that are very early emerging in normal cognitive development and therefore should be considered available to virtually all adult jurors. There are, in fact, several distinct ways to enhance our judgments of the legitimacy of expertise: whether the properties being stressed are likely to co-occur in a domain, whether a coherent domain is being discussed, whether there are signs of deeper causal structure, and whether an appropriate gist has been constructed. Each of these ways is actively being pursued in research, but there are already indications that laypeople also use these to evaluate expertise in their everyday lives.

People will look at which property types are emphasized in an explanation to support hunches about whether an avowed expert is likely to be making sense. For example, all things equal, someone who explains how a novel machine works by stressing its color and the precise number of internal parts is less likely to be a legitimate expert than one who stresses the shape of the machine and the strength or fragility of specific parts. In contrast, someone who is trying to explain the nature of a novel plant might well focus on color and precise numbers of parts of its anatomy. Even five-year-olds have some sense of this difference and can use it to choose between experts.22

22 Id.
Similarly, even preschoolers have a sense that it is more sensible to talk about the overall function of novel artifacts (such as machines and tools) than it is to talk in the same manner about novel animals (for example, it makes more sense to say what a new tool “is for” than it does for a new mammal).

Even children may have a sense to doubt expertise that spans too diverse or broad areas of knowledge. Thus, if one person claims to be an expert on ducks and swans, that person is more credible as an expert than one who claims to be an expert on ducks and lasers. All laypeople, as well as young children, share a sense that it is more likely that one will be an expert on areas of knowledge that are closer together in some taxonomy of knowledge. Similarly, even children are sensitive to the fact that it is more plausible for someone to be an expert on a smaller class of entities at a lower level in a taxonomy (for example, all ducks) than it is to be an expert on a larger class at a higher level of a taxonomy (for example, all animals).

These intuitions require some sense of the division of cognitive labor in a person’s community and the idea that knowledge clusters can be understood to form a kind of hierarchical structure similar to that found in the academic organizational charts of universities. Quite surprisingly, children as young as five do have some sense of such a structure, which they extract by looking at meaningful causal clusters in the world.

In addition, people can also sense when there are signs of deeper causal structure and use those to guess whether one is discussing a legitimate area of expertise. It is, for example, quite obvious to young children that it makes much more sense for there to be an expert on hunting dogs than on dogs with red collars. Even though a child may know virtually nothing about hunting dogs, he or she seems to engage in causal conjectures that reveal a likely causal structure. For example, a child might speculate that dogs that hunt would have better vision and smell and that an expert might know all about how that happened. For dogs with red collars, however, the child would see that there were no comparable plausible speculations about

23 Keil et al., supra note 16.
24 Other work shows that children do have some sense of such taxonomies. Frank C. Keil et al., Discerning the Division of Cognitive Labor: An Emerging Understanding of How Knowledge Is Clustered in Other Minds, COGNITIVE SCI. (forthcoming); Judith Danovitch & Frank C. Keil. Should You Ask a Fisherman Or a Biologist?: Developmental Shifts in Ways of Clustering Knowledge, 75 CHILD DEVELOPMENT 918, 919 (2004).
25 Keil et al., supra note 24; Danovitch & Keil, supra note 24, at 927-28.
causal generalizations that would follow from having a red collar. Thus, even when one does not really understand a domain, one can often use simple heuristics, such as causal counterfactuals, to get a sense of whether there is likely to be causal depth present worthy of expertise. These heuristics are not perfect however and can fail when rating the relative complexity of physical and psychological phenomena. A better understanding is needed of when they can work and when their usefulness is more limited.

Finally, there are indications that people understand the difference between reasonable and unreasonable gists of complex explanations and can use the difference to evaluate experts as well as an area of expertise. They may be able to do so without having much understanding at all of the details of the explanation. For example, ongoing research in our laboratory suggests that people can look at very general structural principles of an explanation, such as how some details are elaborated on by others, to guess at which is a good gist. A good expert should be focusing on core concepts and not on peripheral elaborations of a specific point. Someone who does not really grasp a domain well may not know enough to say much about the core concepts and may try to feign expertise by going on about subdetails of one facet of a phenomenon. Even laypeople, however, can sense or can be trained to sense when this sort of excessive detail is occurring and to start questioning expertise. There are some relatively simple clues as to when someone is going into irrelevant details as a way of trying to cover up ignorance of a central issue, and these clues may be available to a wide range of observers with quite modest knowledge of the area of alleged expertise.

VI. CONCLUSIONS

Human cognition can cause both pitfalls and opportunities in our efforts to get at the truth in a causally complex world in which deference and trust are essential. The pitfalls revolve around the ways in which individuals can be quite poor at recognizing their own areas of weak understanding. We live under illusions of explanatory depth and we have the same illusions about explanatory understanding in others. We are further hindered by a tendency to be seduced by a sense of false insight when we are presented with certain ways of making phenomena more concrete, even when such concreteness is nothing more than uninformative fMRI results.
We may also introduce systematic distortions into our sense of where the deepest causal complexities in the world arise, with the result that we tend to underestimate the complexity of psychological phenomena relative to most physical ones.

The opportunities to overcome these predispositions arise from the surprisingly sophisticated ways in which all of us, even young children, can use our intuitive senses of real world causal structure and of the nature and purpose of explanations to assess experts even when we have very modest knowledge of an area of expertise. We can evaluate experts (and judges) in terms of their self-interest. We can also evaluate experts in terms of the reasonableness of the avowed area of expertise, using such factors as the breadth and depth of what they talk about and the ways in which they summarize complex bits of information. All of these are quite recent areas of research, but every sign is that cognitive science will soon tell us a great deal more about both the illusions we labor under and should be mindful of and about the ways in which we use heuristics and implicit knowledge to have a good sense of when and where to place our trust.

These new developments create a burden for cognitive scientists to communicate the current state of this research more clearly with those in the law, as well as a burden on those in the law to recognize both our cognitive biases and our surprising evaluative abilities. At present, we both overestimate and underestimate different aspects of people’s cognitive capacities, and we do so in ways that may well impair the manner in which trust should optimally function in the courtroom.
On the Conceptual and the Empirical

A CRITIQUE OF JOHN MIKHAIL’S COGNITIVISM

Dennis Patterson†

INTRODUCTION

This symposium was convened to consider the question of scientific truth. Of course, there are many questions one might ask about scientific truth. The most obvious question is whether “truth” names a property and, if so, what sort of property is it? If truth is not a property, then how are we to conceive of it? Is it a relation? If it is, between what things does the relation hold? Sentences or states-of-affairs are possible candidates. While important, these questions are best approached after one has addressed a more fundamental issue, that of the distinction between the conceptual and the empirical.

In this Article, I will argue for two claims: First, there is in fact a distinction between conceptual and empirical questions. Second, conceptual questions are prior to (that is, they antecede) matters of truth and falsehood.¹ The relation-

---

† Board of Governors Professor of Law and Philosophy, Co-Director, Institute for Law and Philosophy, Rutgers University, School of Law (Camden); Professor of Jurisprudence and International Trade, Swansea University, School of Law, Wales, U.K.

¹ Bennett and Hacker explain the empirical/conceptual distinction and its importance for empirical inquiry in the specific context of neuroscience.

Conceptual questions antecede matters of truth and falsehood. They are questions concerning our forms of representation, not questions concerning the truth and falsehood of empirical statements. These forms are presupposed by true (and false) scientific statements and by correct (and incorrect) scientific theories. They determine not what is empirically true or false, but rather what does and what does not make sense. Hence conceptual questions are not amenable to scientific investigation and experimentation or to scientific theorizing. For the concepts and conceptual relationships in question are presupposed by any such investigations and theorizings. Our concern here is not with trade union demarcation lines, but with distinctions between logically different kinds of intellectual inquiry. . . .

Distinguishing conceptual questions from empirical ones is of the first importance. When a conceptual question is confused with a scientific one, it is
ship between the conceptual and the empirical is important because any empirical inquiry that proceeds from conceptual confusion cannot yield satisfactory results.

After setting out the distinctions just outlined in Part I, I will illustrate my claims in Parts II-V with reference to recent work by John Mikhail on moral cognition. Just as Noam Chomsky hypothesizes a universal linguistic grammar to explain speech behavior, so too, Mikhail argues, we can explain the moral behavior of persons in terms of universal moral grammar” ("UMG"). I will argue that Mikhail’s claims on behalf of UMG suffer from conceptual confusions that are not amenable to empirical resolution.

I. THE EMPirical AND THE CONCEPTUAL

There is an important distinction between conceptual and empirical questions. Empirical assertions are claims of fact. They are tested by the methodology of science, that is, through experimentation. In the realm of cognition, neuro- and cognitive scientists test hypotheses about brain functions. To bound to appear singularly refractory. It seems in such cases as if science should be able to discover the truth of the matter under investigation by theory and experiment—yet it persistently fails to do so. That is not surprising, since conceptual questions are no more amenable to empirical methods of investigation than problems in pure mathematics are solvable by the methods of physics. Furthermore, when empirical problems are addressed without adequate conceptual clarity, misconceived questions are bound to be raised, and misdirected research is likely to ensue. For any unclarity regarding the relevant concepts will be reflected in corresponding unclarity in the questions, and hence in the design of experiments intended to answer them. And any incoherence in the grasp of the relevant conceptual structure is likely to be manifest in incoherences in the interpretation of the results of experiments.


2 John Mikhail, Universal Moral Grammar: Theory, Evidence and the Future, 11 TRENDS IN COGNITIVE SCI. 143, 143 (2007). A central claim Mikhail makes is that “the mind contains a moral grammar.” Id. at 144. Mikhail makes clear that his account of moral grammar is a “computational” theory. Id. at 143.

3 Mikhail claims that “further research is needed to clarify the relevant conceptual and evidentiary issues.” Id. at 148. I contend that more research will clarify nothing because the research program is fundamentally misconceived. I discuss these issues more broadly in Dennis Patterson, Fashionable Nonsense, 81 TEX. L. REV. 841 (2003) (review essay discussing ANTHONY G. AMSTERDAM & JEROME E. BRUNER, MINDING THE LAW (2000); STEVEN L. WINTER, A CLEARING IN THE FOREST: LAW, LIFE AND MIND (2001); VINCENT DESCOMBES, THE MIND’S PROVISIONS: A CRITIQUE OF COGNITIVISM (Stephen Adam Schwartz, trans., 2001)).
take just one example, neuroscientists are particularly interested in correlating brain function with emotional and behavioral responses.4 Such work is thought to provide insight into the connection between the brain and behavior.

Conceptual questions, on the other hand, involve matters of sense.5 They are not amenable to empirical assessment, confirmation, or analysis. Conceptual relationships are presupposed by empirical claims. If there is conceptual confusion, then nothing of empirical value can obtain. The success of empirical inquiry depends upon conceptual clarity (that is, the absence of conceptual error or confusion).

Some neuroscientists are not content with limiting their work to brain functions and have given in to speculation about the relationship between the mind and the brain. Consider the concept of “mind.” The question “what is mind?” or “what is a mind?” implicates a panoply of other concepts like “vision,” “understanding,” and “thought.” The question “what is mind?” is conceptual and not empirical because, among other reasons, no experiment could answer the question. An answer to the question “what is mind?” requires a different sort of reply than we give to the question “where in the brain does one find the medulla?”


5 Bennett and Hacker explain the relationship of sense to truth thus:

Cognitive neuroscience is an experimental investigation that aims to discover empirical truths concerning the neural foundations of human faculties and the neural processes that accompany their exercise. A precondition of truth is sense. If a form of words makes no sense, then it won’t express a truth. If it does not express a truth, then it can’t explain anything. Philosophical investigation into the conceptual foundations of neuroscience aims to disclose and clarify conceptual truths that are presupposed by, and are conditions of the sense of, cogent descriptions of cognitive neuroscientific discoveries and theories. If conducted correctly, it will illuminate neuroscientific experiments and their description as well as the inferences that can be drawn from them. In Philosophical Foundations of Neuroscience we delineated the conceptual network formed by families of psychological concepts. These concepts are presupposed by cognitive neuroscientific research into the neural basis of human cognitive, cogitative, affective, and volitional powers. If the logical relations of implication, exclusion, compatibility, and presupposition that characterize the use of these concepts are not respected, invalid inferences are likely to be drawn, valid inferences are likely to be overlooked, and nonsensical combinations of words are likely to be treated as making sense.

Conceptual confusions arise in a variety of ways and can take several forms. Fundamentally, error can arise from failure to employ words in accordance with the rules for their use. But confusion can also arise in more complex ways. Professor Mikhail errs when he tries to locate moral knowledge in a “place,” that place being the mind. This form of conceptual error, I will argue, undermines Mikhail’s arguments for the explanatory power of UMG.

II. MIND AND MORAL GRAMMAR

Now to the more complicated subject of “mind.” What are the proper forms of expression for referring to “the mind”? To this point, I have endeavored only to make the point that the question “what is mind?” is not amenable to an empirical answer and that the answer it requires is bound up with rules for the use of words and expressions associated with our neural capacities. Now, I shall detail the central claims made by Mikhail on behalf of the theory of UMG before suggesting why these claims are conceptually confused.

I start with what Mikhail identifies as the core questions for UMG before considering Mikhail’s key claims. These questions frame Mikhail’s inquiry into the nature of mind and delineate the central focus of his research into the relationship between mind and moral knowledge. He asks:

1. What constitutes moral knowledge?
2. How is moral knowledge acquired?
3. How is moral knowledge put to use?
4. How is moral knowledge physically realized in the brain?
5. How did moral knowledge evolve in the species?

In describing the main features of our innate moral capacity, Mikhail makes a number of key claims about the nature of mind and moral grammar. These are:

1. “[T]he mind contains a moral grammar . . . .”

---

6 I discuss this topic in connection with philosophical naturalism in Dennis Patterson & John Oberdiek, Moral Evaluation and Conceptual Analysis in Jurisprudential Methodology, in LAW AND PHILOSOPHY 60-75 (Michael Freeman & Ross Harrison eds., 2007).

7 Mikhail, supra note 2, at 144.

8 Id.
2. “[T]he manner in which this grammar is acquired implies that at least some of its core attributes are innate, where ‘innate’ is used in a dispositional sense to refer to cognitive systems whose essential properties are largely pre-determined by the inherent structure of the mind . . . .”

3. Moral intuitions “are best explained by assuming [that individuals] possess tacit knowledge of specific rules, concepts or principles.”

I shall accept Mikhail’s claims about the nature of mind and, further, assume arguendo the truth of his claims about the role and function of moral grammar.

Nevertheless, Mikhail makes at least two philosophically suspect claims about the relationship between mind and moral grammar. These claims hint at larger, more structural problems with UMG. These are the two claims:

1. Moral knowledge is in the mind (brain); and
2. Moral reasoning is a matter of unconscious application/interpretation of rules, principles, and “domain specific algorithms.”

The first claim involves the locus of the mind’s tool for solving ethical problems. In the course of solving these problems, Mikhail argues, the mind accomplishes its moral computational tasks unconsciously. The second claim asserts that the mind’s methodology for tackling ethical problems is “interpretation.”

There are two problems with the idea that we can explain moral cognition with the claim that the mind follows rules by unconsciously interpreting their requirements in particular cases. The first involves the claim that to understand what a rule requires we need to interpret it. The second

---

9 Id.
10 Id.
11 Id. (“[T]he mind contains a moral grammar.”). This grammar is “innate” in that its “essential properties are largely pre-determined by the inherent structure of the mind.” Id.
12 Id. at 148. Knowledge of these moral rules is “tacit.” Id. at 145. In solving ethical problems, “a pattern of organization . . . is imposed on the stimulus by the mind itself.” Id. The process of computation is “unconscious.” Id.
13 Mikhail puts it this way: “[H]ow the mind goes about interpreting these novel fact patterns, and assigning a deontic status to the acts they depict, is not revealed in any obvious way by the scenarios themselves.” Id.
involves the claim that “following a rule” is something a person does “unconsciously.” Neither of these claims makes sense; and, lacking sense, neither is amenable to empirical testing (that is, experiment).

III. UNDERSTANDING AS INTERPRETATION

Mikhail’s claim that we follow rules by interpreting them is one made in a wide variety of humanistic and social-scientific disciplines. The problem with the claim is structural and conceptual. If understanding a rule first requires interpretation of it, then there is no reason why the interpretation itself does not similarly stand in need of interpretation. This process of interpretive regression can go on infinitely. Hence, the term “infinite regress” has been used to describe the argument against the idea that to be understood, rules must first be “interpreted.” This argument, however, is

14 For detailed discussion, see generally Dennis Patterson, The Poverty of Interpretive Universalism: Toward the Reconstruction of Legal Theory, 72 TEX. L. REV. 1 (1993).

15 Wittgenstein made the point this way:

“But how can a rule shew me what I have to do at this point? Whatever I do is, on some interpretation, in accord with the rule.”—That is not what we ought to say, but rather: any interpretation still hangs in the air along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meaning.

LUDWIG WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS § 198 (G.E.M. Anscome trans., 1958); see also ROBERT B. BRANDOM, MAKING IT EXPLICIT: REASONING, REPRESENTING, AND DISCURSIVE COMMITMENT 508-09 (1994) (“Linguistic understanding depends on interpretation . . . only in extraordinary situations—where different languages are involved, or where ordinary communication has broken down.”); Jeff Coulter, Is Contextualising Necessarily Interpretive?, 21 J. PRAGMATICS 689, 692 (1994) (“Understanding is not an activity: it is akin to an ability. To understand is to have achieved knowledge of some kind, whilst interpreting is an activity which is akin to hypothesis formation or, in a different sense, to the assignment of significance (explanatory or otherwise) broader than the determination of intelligibility.” (footnote omitted)).

16 Peter Hacker explains:

[It is a grievous error to think that in understanding an utterance one always or even usually engages in interpretation. To interpret an utterance is to explain it, typically to paraphrase it in the same language or to translate it into a different language. . . . Obscurities, ambiguities or complexities may call out for an interpretation, but it would be wholly incoherent to think that all understanding is interpreting. For then the interpretation given, i.e. the paraphrase, would itself stand in need of an interpretation in order to be understood; and a vicious regress would be generated. This misconception has manifold roots. One is the bizarre idea that what we hear or utter are mere sounds which have to be correlated with or mapped on to meanings in order to be understood. But we no more hear or utter mere sounds than we
compressed: I shall spell it out in more detail, using Wittgenstein’s arguments regarding understanding and interpretation to support my case.\footnote{For detailed discussion of the understanding/interpretation distinction and its relevance for law, see Dennis Patterson, Law and Truth 86-89 (1996).}

Wittgenstein’s basic claim is that “understanding” is primary and “interpretation” a second-order or “parasitic” activity.\footnote{WITTGENSTEIN, supra note 15, §§ 139-242.} Interpretation is parasitic in the sense that interpretation only arises where understanding is already in place. Understanding, according to Wittgenstein, is unreflective action. When we follow rules, we do so without second-guessing ourselves and without reflection on what the rule requires.

Wittgenstein begins his argument for the primacy of understanding by presenting us with a paradox. He writes:

> This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here.\footnote{Id. § 201.}

Why does Wittgenstein question the importance of interpretation as an explanation of meaning (that is, as an explanation of what it is to grasp the meaning of a rule or what the rule requires by way of correct action)? His point is that if the understanding of an utterance or sign were a matter of advancing an interpretation (which is just another utterance or sign), then the interpretation itself would require its own interpretation, and so on, infinitely. This argument—the infinite regress argument—is meant to inspire us to question the idea of understanding as interpretation. Wittgenstein urges us to rethink the notion that before we can understand an utterance we must first interpret it. According to him, understanding a rule is fundamental to our role as participants in practice. Interpretation, by contrast, is an activity we engage in when our understanding breaks down.

Wittgenstein’s insight is that rule-following is not a mental phenomenon. Succinctly stated, Wittgenstein relocates normativity in action, specifically in social action. The

---

normativity of rule-following—the ground of correctness and incorrectness—is not to be found in the agreement of others as such. Rather, the agreement of rule-followers over time is the ground of understanding. Agreement is a necessary feature of the normativity of our practices, but the agreement must be a regularity in reaction to use. In short, when we say there must be “agreement in actions” what we are really saying is that there must be harmony in application over time. This harmony in reaction and application is constitutive of all practices, including legal practice. It is the basis of our legal judgments.

The distinction between correct and incorrect rule-following is a matter of community agreement in judgments over time. If Wittgenstein is correct, then the idea of unconscious rule-following is nonsensical. Following a rule, making judgments about what a rule requires, and the very idea of normativity itself require a role for others in the intersubjective constitution of norms of correctness. The Mikhail/Chomsky view of rule-following never gains traction because it never moves beyond the ground of the internal constitution of mind.

20 For discussion, see Meredith Williams, Wittgenstein, Mind and Meaning: Toward a Social Conception of Mind 176 (1999).

It is in this sense that community agreement is constitutive of practices, and that agreement must be displayed in action. There are two important features about this account that need to be highlighted. First, it is the social practice that provides the structure within which individual understanding can obtain or individual judgement be made. Central to Wittgenstein’s thought is the claim, repeatedly argued for, that no isolated event or behavior can correctly be described as naming or obeying or understanding. The rule as formula, the standard as chart, or the paradigm as an instance have no normative or representational status in their own right. They have this status only in virtue of the way the formula or the chart or the instance is used. It is the use that creates the structured context within which sign-posts point, series can be continued, orders be obeyed and paradigms be exemplary. Only then can we see a particular action as embodying or instancing a grammatical structure. In short, the mandatory stage setting is social practice.

Second, community agreement does not constitute a justification for particular judgements. What is indispensable for correct, or appropriate, judgement and action is that there is concord, not that each individual justifies his (or anyone else’s) judgement and action by appeal to its harmony with the judgement of others.

Id. (footnote omitted).

21 See id. at 169.
IV. UNCONSCIOUS RULE-FOLLOWING

Like Chomsky’s rules of generative grammar, Mikhail’s UMG postulates innate knowledge of a moral grammar. This claim makes no sense. How can a child be said to “know” moral norms without ever being conscious of them? In other words, before a child even learns a syllable of language, how can she be said to possess moral knowledge?

The problem posed by this question cannot be avoided by asserting that we “follow” rules “unconsciously.” Again, the problem is conceptual. “Rule-following” includes a panoply of normative activities. When we follow rules we do the following things:

1. Justify our behavior by reference to the rule;
2. Consult the rule in deciding on a course of conduct;
3. Correct our behavior and that of others by reference to the rule; and
4. Interpret the rule when we fail to understand what it requires.

It is difficult to see how these normative activities are possible when we are unconscious of the existence of the rule. Of course, we may act in a manner consistent with a rule. But that is not to say that we are following the rule, for to do that would require that we do all the things I just mentioned. Thus, Mikhail’s claim that a person follows a rule unconsciously is untenable.

V. MIND, MORAL GRAMMAR, AND KNOWLEDGE

Finally, I come to Mikhail’s most fundamental claim, that is, that “the mind contains a moral grammar” and that

---

22 Generative grammar is a theory of syntax. The grammar takes the form of a system of formalized rules which mechanically generate all and only the grammatical sentences of a language. See generally Noam Chomsky, Aspects of the Theory of Syntax 3-10 (1965).

23 In the opinion of one careful reader, Chomsky—the inspiration for Professor Mikhail’s model of unconscious rule-following—has abandoned the idea. See John Searle, End of the Revolution, N.Y. Rev. Books, Feb. 28, 2002, at 36 (reviewing Noam Chomsky, New Horizons in the Study of Language and Mind (2000)) (“Chomsky has now given up on the idea that Universal Grammar is a matter of unconscious rule-following. But he also dismisses the idea that real human languages are governed by rules. That, I believe, cannot be right.”).

24 Mikhail, supra note 2, at 144.
this grammar is part of “the inherent structure of the mind.”\textsuperscript{25} Like all forms of rationalism, Mikhail’s premise is that the object of explanation (in this case, moral knowledge) is located in a place. In Mikhail’s view, moral knowledge is in the mind.

This claim is confused because “knowledge” is an ability and not a thing. “Know” is a success verb, not a referent.\textsuperscript{26} To “know” something is neither to be in a certain state nor is it to be in possession of a particular structure of mind or brain.\textsuperscript{27} Like all abilities, the ability to know moral rules is exhibited in behavior. Criteria for the ascription of knowledge consist of correct performances. “Knowledge” includes, among other things, being able to spot error, explain the error, and correct it. In doing these things, one demonstrates that one has mastered rules—not that one’s mind or brain “contains” the rules.\textsuperscript{28}

Consider: when we say “Jones knows the train schedule from Warsaw to Krakow,” we are not saying that Jones has the schedule hard-wired into his mind. Even if he did, that would still not be sufficient to say that he “knows” the schedule because to know the schedule means knowing how to read the schedule correctly. To do this, Jones needs to be able to do things with the schedule. It is that doing that is the ground of the ascription “Jones knows.” Since knowledge is an ability, rather than a thing, it cannot be located in the brain, or anywhere else for that matter.

There is also a more fundamental problem with Mikhail’s account of UMG, one that goes beyond the confusion of an ability with a thing. The problem is that Mikhail thinks that the mind is a “place.” In this place, Mikhail locates the moral grammar whose “inherent structure” explains our behavior. But there is no such place as “the mind.” John Searle recently compared thinking with digestion.\textsuperscript{29} He postulated that just as digestion occurs in the stomach, so does consciousness

\textsuperscript{25} Id.

\textsuperscript{26} Of course, when we make a decision, many things may cross our mind or come to mind. The causal processes for these are varied and many. But none of these accompaniments constitutes “thinking” or “deciding.”

\textsuperscript{27} ANTHONY KENNY, THE LEGACY OF WITTGENSTEIN 129 (1984) (“To contain information is to be in a certain state, while to know something is to possess a certain capacity.”).

\textsuperscript{28} See P.M.S. Hacker, Chomsky’s Problems, 10 LANGUAGE & COMM. 127, 128-29 (1990).

\textsuperscript{29} See John Searle, Putting Consciousness Back in the Brain: Reply to Bennett and Hacker, Philosophical Foundations in Neuroscience, in BENNETT ET AL., supra note 5, at 97, 108-09.
Because we predicate both thinking and digesting to persons, these important capacities could each be located in a person. But the analogy does not hold. If we open someone’s stomach, we can see the process of digestion occurring. But if we open someone’s brain or mind, we do not find anything we might call “thinking.” Of course, Mikhail’s point could be that in order to have knowledge we must first have the necessary equipment to make judgments. This is granted but handled by the distinction between having an ability and exercising it. The normative (“what is thinking?”) cannot be reduced to the causal (“what enables us to think?”).

The most fundamental problem with Mikhail’s view of the nature of mind is that the idea of a moral grammar hard-wired into the mind is a posit that can never be shown to be true. Deeming this posit a hypothesis is to use the language of experimentation to support a thesis the truth of which could never be shown to be true or false. Mikhail’s claims for the

30 Id.

31 Query how one would “open a mind”?

32 Of course, an MRI scanner will show that certain areas of the brain are actuated when a person is thinking. While the brain is necessary for one to have thoughts, the thoughts are not “located” in the brain. See Bennett & Hacker, supra note 5, at 143.

33 See Hacker, supra note 28, at 134.

Neurophysiologists may discover that certain neural configurations are causally necessary conditions for having the ability to speak a language. But they will never find any knowledge in the brain. Neither what one knows, namely truths, facts, or propositions, nor abilities to say, show or tell what one knows (i.e. abilities constitutive of knowing) can (logically) be found in the brain. For truths, facts, and propositions, although they can be recorded on paper or on a computer disc, cannot be recorded on or in the brain. For whereas we record facts by means of a symbolism, a language, and write the information down in a notebook or store it on a computer disc, there is no such thing as the brain’s employing a language, understanding or speaking a language, nor is there any such thing as human beings employing brains as repositories for written records, let alone as computer discs. To say that truths, facts, or propositions are stored, filed away, or retained in one’s mind is just to say that they are known and not forgotten.

34 Richard Rorty makes the same point with respect to Chomsky:

Consider, for example, Chomsky’s claim that there is ‘a fixed biologically determined function that maps evidence available into acquired knowledge, uniformly for all languages’. It is hard to see this as an empirical result, since it is hard to think what could disconfirm it. It is uncontroversial that organisms that can learn languages have this ability because they have different neural layouts than other organisms. The layouts, to be sure, are biologically determined. But in what sense can a function be so determined?
explanatory efficacy of UMG trade on the language of science without delivering scientific results. In short, the arguments for UMG confuse two different discourses, those of scientific proof and those of speculative metaphysics.

So, what is “mind”? Mind is the ability to engage in linguistic behavior. “Having a mind” is just having a set of social skills. Talk of moral grammar is simply an elliptical reference to the language of morals, not a state of the mind or brain. This is not to minimize moral problems—far from it. It is, however, to locate them in their proper place—that is, “outside” the mind or brain.

CONCLUSION

In this Article, I have defended the claim that there is a distinction between the empirical and the conceptual. I have made the case that conceptual questions precede empirical investigations because without conceptual clarity, nothing of sense can follow from experiment. I have used John Mikhail’s arguments on behalf of UMG to make the case for the distinction between the conceptual and the empirical. The issues Mikhail has taken up implicate some of the most basic debates in contemporary philosophy. But these contemporary debates are a reprise of a much older debate, that between rationalism and its critics. The primary issues presented by

To say that a mechanism embodies a function is just to say that its behavior can usefully be described in terms of a certain specifiable relation between input and output. Nobody can specify any such relation between the inputs provided by language-teaching adults and the outputs provided by a language-learning child, because they are too various. It would be like trying to specify a relation between the events that occur in the course of learning to ride a bicycle and those that are the actions of the accomplished bicyclist.

But, Chomsky tells us, there is a function that, rather than mapping inputs onto outputs, maps inputs into something called ‘acquired knowledge’. Well, the bicyclist too has acquired some knowledge. Should we say that he has acquired it thanks to a biologically determined function that maps the events of his early, tentative, abortive rides onto a set of internal representations whose possession is a necessary condition of his newly acquired ability? We could, but what would count as confirming the existence of such a mediating entity, in between the learning events and the actions which produce successful bicycle rides?

UMG cannot be resolved through further empirical research.35 These issues can never be resolved by empirical methods because they are, at bottom, philosophical.

35 Mikhail seems to think otherwise. Mikhail, supra note 2, at 148 (“[F]urther research is needed to clarify the relevant conceptual and evidentiary issues.”).
Scientific Realism in Constitutional Law

David L. Faigman†

I. INTRODUCTION

In scholarly circles a debate rages over whether scientific research describes true underlying realities of the natural world or merely represents constructed accounts of observed events.1 Much of this debate involves natural science and reaches such fundamental issues as whether, for example, we can positively conclude that electrons exist or must be limited to statements about the observed effects of electrons, since we cannot observe those subatomic particles directly. Scientific realists argue that science is not simply a collection of hypotheses supported by empirical tests, but actually describes an underlying reality that exists outside of human observers. The world, according to this view, is “mind independent.” The core philosophical disagreement found in the natural sciences can also be found in the many other disciplines in which knowledge about the empirical world is essential, from the social sciences to the humanities, only more so. Any mind-dependence infecting physics would impair the social sciences and humanities at many times the rate. Not surprisingly, therefore, the law, as a consumer of these empirically conscious disciplines, is deeply affected by these debates and, to some extent, must choose sides. In the law, the question whether reality is real or not is more than an academic debate. The law’s response to this debate has, not to put too fine an ironic point on it, substantial real-world consequences.

† John F. Digardi Distinguished Professor of Law, University of California, Hastings College of the Law.

1 See generally JAMES ROBERT BROWN, WHO RULES IN SCIENCE: AN OPINIONATED GUIDE TO THE WARS (2001); STATHIS PSILLOS, SCIENTIFIC REALISM: HOW SCIENCE TRACKS TRUTH (1999); MICHAEL DEVITT, REALISM AND TRUTH (2d ed. 1991).
A diversity of views is represented by the label “scientific realism.” Among the choices of philosophical starting premises, however, realism most centrally embraces the notion that science discovers “truth.” Truth, albeit with a lowercase t, does not necessarily imply that scientists can say unambiguously or with certainty that the world operates in a particular fashion. Rather, the world exists in particular ways, and science more or less—or with greater or lesser precision—endeavors to describe that world. But science is a human enterprise and a community effort. The real truth, therefore, may be known only rarely and, even then, only after extensive study and many missteps. Still, its existence largely makes the scientific effort worthwhile. Scientific methods permit the development of a body of knowledge about the world that does not depend on the cultural backgrounds or values of its originators. In common parlance, science can be “objective,” in that it can be tested “inter-subjectively” by different people in different places having different values.\(^2\) Realists believe that scientific methods provide an objective lens through which the world can be described, if only imprecisely.\(^3\)

Challenges to realism come from a wide assortment of disciplines, including, among others, philosophy, sociology, and literary theory. Critical scholars in these fields\(^4\) contest the objectivity of knowledge and dispute the claim of mind-independence that realists believe is possible.\(^5\) While there are indeed widely ranging views regarding the inscrutable issue of the reality of “truth,” necessity requires a simpler presentation of the debate in this essay. From the law’s perspective—and, more particularly, from the perspective of constitutional adjudication—the matter comes down to either believing that science can describe the empirical world largely free of bias or that it cannot. If facts having relevance to constitutional lawmaking do not exist—or cannot be described—separately from the values endemic in that lawmaking, then it is

\(^2\) Karl Popper, The Logic of Scientific Discovery 44 (Harper & Row 1968) (1959) (“The objectivity of scientific statements lies in the fact that they can be inter-subjectively tested.” (emphasis in original)).


\(^4\) See generally Brown, supra note 1.

\(^5\) Professor Susan Haack refers to them as the “New Cynics.” She notes that they disagree among themselves “on the finer points,” but generally agree that “concern for truth, is a kind of illusion, a smokescreen disguising the operations of power, politics, and rhetoric.” Susan Haack, Defending Science—Within Reason 20-21 (2003).
incumbent on courts not to pretend that they do. Facts and values (or biases), under this view, may not be one, but they are inextricably bound. If this is so, anti-realism is the more rational choice to provide the philosophical basis for constitutional adjudication. But if facts can exist independently of biasing influences, as I believe they can, then courts should fully account for them in their decisions. In short, scientific realism obligates courts to take facts seriously.6

The anti-realist claim that must be rebutted in order to substantiate my argument that facts ought to be taken seriously in constitutional cases is associated with the belief that scientific knowledge is largely socially constructed. Adherents of social constructionism fall along a wide spectrum of beliefs, with some subscribing to more or less extreme versions. Indeed, many realists share the concerns that lie at the core of anti-realist critiques of science. Scientific realists well appreciate, for example, the effects a researcher’s values might have on how hypotheses are formed or what methods are selected to test them. Similarly, realists generally accept, at least in principle, Thomas Kuhn’s basic claim that theoretical paradigms affect the problems scientists study and the answers they obtain.7 To a large extent, the difference between sober realists and sensible anti-realists is one of degree or emphasis. Anti-realists generally hold the view that scientific statements are so imbued with the values and social and historical contexts of their declarants that they are effectively normative in scope. They deny any special claim of “objectivity” to scientific facts and, in effect, deny the fact-value distinction altogether.

6 Although the perspective to be defended here is the realist one, as opposed to anti-realist and social constructivist alternatives, I need not defend a strong version of this perspective. For example, scientific realism is sometimes juxtaposed to empiricism. Whereas scientific realists posit the true existence of unobservable entities, empiricists are content to be agnostic about underlying realities, though still committed to the rationality of hypothesis formation and test. The empiricist tradition thus seeks to demonstrate “that theoretical discourse may be so construed that it does not commit to the existence of unobservable entities.” Psillos, supra note 1, at 3. In contrast, the realist tradition “aims to show that a full and just explication of theoretical discourse in science requires commitment to the existence of unobservable entities.” Id. From the law’s perspective, however, this particular debate is academic, since empiricists and realists agree on the virtue of rigorous hypothesis testing. Empiricism and realism share the attribute that I refer to as taking facts seriously. Compared to strong anti-realist views, therefore, realism and empiricism are close cousins.

7 See Thomas Kuhn, The Structure of Scientific Revolutions 43-51 (1970); see also Brown, supra note 1, at 63-71 (discussing Kuhn and realists’ responses to him).
Realists, in contrast, believe that the world exists independently of the minds of its explorers and that the methods of science are largely effective in discovering the mechanics of that world. While biases can—and too often do—infect the explorations of scientists, scientific methods are designed and employed to limit that bias as much as possible. Hence, when failures occur, as they have and inevitably will, they are attributable to the scientists, not science. The solution is to strive for better scientific research, not abandon the enterprise.

Although the United States Supreme Court is an eminently realist institution in that the justices almost certainly see themselves as situated in a mind-independent real world, the Court tends to employ facts as though they were subjects of social construction. The Court insistently employs factual arguments rhetorically, as premises that can be manipulated or massaged in the service of one or another legal outcome. The Court has largely constructed an empirical world that serves the normative vision it holds for the Constitution. For example, it may be that the Fourth Amendment right to be free from unreasonable search and seizure is defined in light of an “objective” person’s “reasonable expectations of privacy,” but the justices make no attempt to match the empirical reality of such expectations with constitutional outcomes. The Court’s subjective construction of those expectations establishes the contours of the Fourth Amendment right. The justices, therefore, stand in the untenable position of subscribing to scientific realism as a foundational philosophy, but act as anti-realists in crafting constitutional outcomes. To vary an old saying, they want to have their cake and make us eat it too.

In this Essay, I examine whether facts can be treated realistically in constitutional decision making. In particular, I consider two potentially insurmountable challenges to a scientifically realist constitutional jurisprudence. The first is the question of whether the sorts of facts the Constitution makes relevant—primarily behavioral and societal facts studied by social scientists—can be studied relatively objectively. The second is whether the constitutional inquiries in which facts play a part are hopeless conglomerations, so that

---


the empirical and the normative cannot be separated. I conclude that while both of these issues present formidable challenges, neither is fatal to the development of a scientifically realist constitutional jurisprudence.

II. REALIZING A SCIENTIFICALLY REALISTIC JURISPRUDENCE

It must be emphasized at the outset that believing that scientists study true underlying realities says little about the value of a particular research program for the law. Although some fact may be “true,” this does not mean that particular legal consequences flow from that “truth.” The law is an applied discipline and so the issue of whether genes truly exist, for example, does not answer the question of how, or even whether, research indicating some genetic predisposition is legally cognizable. In the simplest of terms, “is” does not entail “ought.” At its best, science has no particular political agenda. For instance, discovering a genetic basis for pedophilia might have multiple legal impacts, variously having “liberal” or “conservative” consequences. For instance, this genetic evidence may be used by defendants to support an insanity plea or by prosecutors to establish guilt; both defense lawyers and prosecutors may use this proof at sentencing; and the state will undoubtedly seek to use this sort of evidence in commitment hearings of alleged sexually violent predators. Good science is neither inherently liberal nor conservative.

It is also important to emphasize that applied science is invariably variable. What is “true” generally will be true only some of the time in practice. The “truth” of the empirical connection between genes and pedophilia is not the ultimate question in most legal disputes. While the general truth is certainly pertinent, the operative issue typically will be whether some particular person has acted (or will act in the

---

10 It is worth noting that even general “truths” in science are provisional, or uncertain, in a variety of ways. General findings are only as good as the research supporting them and, especially in the behavioral sciences, are usually described probabilistically. For example, psychologists have found that cross-racial identifications are less reliable than same-race identifications. See Christian A. Meissner & John C. Brigham, Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review, 7 PSYCHOL. PUB. POLY & L. 3, 4 (2001). Although this is a well-researched phenomenon, it is not invariable. See Stephanie J. Platz & Harmon M. Hosch, Cross-Racial/Ethnic Eyewitness Identification: A Field Study, 18 J. APPLIED SOC. PSYCH. 972, 972-73 (1988). In addition, even the best believed scientific truths might someday be overturned by new discoveries or more inclusive theories.
future) in accordance with this genetic predisposition. However true the underlying reality, this knowledge will be probabilistic in application. As with virtually all applied science, research will at best illuminate the statistical relations between factors. Moreover, in most cases, the hypothesized empirical connection, itself probabilistically described, will be clouded by the possibility of factors never studied and systematic and random error endemic to any research program. Research, for example, might indicate that forty-five percent of men with a specific set of, say, thirteen genes will be sexual aggressors. This, of course, would be highly relevant information. Yet, fifty-five percent of men with these genes will not be sexual aggressors. It may be that environmental factors or other variables partly explain which men will act on their seeming predispositions and which will not. Whatever the case, when it comes to individual statements of fact, the best that scientists can do is speak in terms of probabilities and statistics.

The probabilistic character of applied science is an inherent limitation of the discipline. The tools of science, therefore, are limited in their capacity to describe the world that the law regulates. Importantly, however, the uncertainty, or error, associated with scientific tools is primarily random rather than systematic. In other words, the error is randomly distributed among political outcomes and does not systematically prefer the conclusions particular researchers might favor.

The anti-realist critique is directed at the prospect of systematic error. Anti-realists believe that researchers’ subjective biases infuse the design and interpretation of their work. They do not believe reality exists separately from researchers’ statements. The reality is in the words, not the world. To many in the law, however, asking whether constitutional facts are mind-independent will strike them as patently absurd. Indeed, even among anti-realists, the strong version of the claim is truly endorsed by only a small group of skeptics, and one might wonder how strongly even they believe it. Even the most ardent anti-realists look both ways when they cross the street. But a somewhat weaker form of anti-realism might have a place in constitutional cases. This is so for two independent reasons.

First, many of the facts having constitutional significance come from the so-called softer disciplines of psychiatry, psychology, economics, political science, and sociology. These fields have a history of producing socially dependent knowledge and employ methods that limit their power to
transcend time, place, and setting. Unlike many of the natural sciences, these fields tend to produce results that are weak in explanatory power, methodological rigor, and demonstrated reproducibility.

The second reason to believe that a weak form of anti-realism has a place in constitutional cases is that, in practice, the Court tends to amalgamate constitutional facts and constitutional norms. For example, in Planned Parenthood of Southeastern Pennsylvania v. Casey, the Court held that pre-viability abortion regulations violate the Constitution if they “unduly burden” the right of reproductive choice. The Court defined as unduly burdensome laws that create a “substantial obstacle” to the exercise of the right. Whether a law places obstacles in the path of a woman’s exercise of her right to an abortion is an empirical question. Whether these obstacles qualify as substantial obstacles—enough to create an undue burden—is a question that contains a strong normative component. The undue burden standard, therefore, creates a constitutional problem that is an admixture of fact and value, and thus arguably contemplates a socially constructed answer.

This section considers the two basic forms of anti-realism as they might be manifested in constitutional cases. Due to space considerations, the following discussion is limited to social science, which is the predominant form of science found in constitutional cases. Also, if the case is made successfully with social science, its more muscular cousins should pass philosophical muster easily. Part A first considers whether the social sciences can achieve some measure of objectivity. Part B then considers whether constitutional standards, such as the undue burden test of Casey, can be untangled so that the factual elements can be examined independently of the normative insights that inform them.

---

12 Id. at 846.
13 This essay focuses on science; space constraints preclude my consideration of history, the penultimate (and possibly ultimate) source of factual information in constitutional cases. In the admittedly misleading hierarchy of objectivity, if the methods of natural science allow it to be more objective than social science, the methods of social science allow it to be more objective than history. Hence, on balance, the anti-realist critique resonates somewhat more with historical “truth” than scientific “truth”—though good history is much more than mere social constructions.
A. **Soft Science and Social Construction**

Implicit or explicit in virtually all discussions of the significance of social science findings to legal decision-making lies the question of whether social inquiry can be scientific. Perhaps the most often repeated criticism of social science is that it is inherently value laden. A component of this argument is based on the truism that, as human beings, social scientists study themselves. Social scientists bring too much baggage of their own to the laboratory, the argument goes, to be able to study other people’s behavior objectively. Without question, social scientists’ values affect the kinds of research they do and, at least indirectly, their findings. This is true of the natural sciences as well. The topics selected for study, the variables identified as worthy of measurement, and, to some extent, the interpretation of findings, depend on the values, interests, and intentions of the scientist and the times in which he or she lives. The principal advantage of scientific methods is not that they eliminate researchers’ biases, only that they help to control and reveal the biases that do exist.

Essentially six basic sources of bias in social inquiry can be readily identified: (1) the selection of problems, (2) the definition of the subject of study, (3) the methodological choices made, (4) the determination of the contents of conclusions, (5) the division of fact from value, and (6) the assessment of evidence. Although these six sources of bias constitute significant challenges to social scientific inquiry, they do not doom the project.

1. **Selecting Problems**

Social scientists have long been criticized for spending a disproportionate amount of time on law-related issues about which the law cares relatively little. For example, researchers’ efforts to study juries is out of all proportion either to the

---


15 This section owes a considerable intellectual debt to Professor Ernst Nagel, who posited—and refuted—several of these sources of bias in his extraordinary book *The Structure of Science*. See generally Ernst Nagel, *The Structure of Science: Problems in the Logic of Scientific Explanation* (1961).
number of trials or to the number of trials involving juries. Additionally, a brief perusal of the social science and law literature would suggest that eyewitness identification is the most important empirical issue facing the legal system.\textsuperscript{16} Although social scientists have increasingly expanded their research focus to new areas of the law,\textsuperscript{17} it remains fair to complain that they concentrate inordinately on juries, witnesses, and criminals. Behavioral issues in other areas, such as constitutional law, torts, and property, are largely ignored.

Not surprisingly, social scientists tend to select problems on the basis of their interests, their understanding of the law, and the amenability of the problems to scientific study. Thus, the proliferation of studies on eyewitness identification is understandable, in that it flows naturally from a long history of research on human perception and memory. Also, a non-lawyer can easily understand the danger of eyewitness misidentification and its importance to the law.\textsuperscript{18} Thus, unlike complex legal and psychological issues, such as the coercive impact of religiously inspired prayer at graduation ceremonies,\textsuperscript{19} eyewitness perception requires little legal sophistication and is relatively easy to research. Moreover, the eyewitness research literature has been an influential component of public policy debates and has led to a variety of contemporary reforms.\textsuperscript{20} Psychologists interested in having an impact on public policy, therefore, have naturally focused on a subject as readily amenable to study as eyewitness identification.

Criticism of problem selection in the social sciences should be directed more at the possible lack of relevance of the research and less at the inherent value bias of the researchers. In general, scientists select problems on the basis of what seems important, and to this extent all science is culture-


\textsuperscript{17} Outside of jury and eyewitness work, promising areas of study include children’s memory, predictions of future violence, judgment and decision-making, and fMRI brain research.

\textsuperscript{18} See Elizabeth F. Loftus, \textit{Reconstructing Memory: The Incredible Eyewitness}, 15 JURIMETRICS J. 188, 190 (1975) (“Since eyewitness testimony carries so much weight, it is important to find out why distortion occurs in a witness’ memory.”).

\textsuperscript{19} See Lee v. Weisman, 505 U.S. 577, 592-93 (1992) (applying what Justice Scalia called the “psychological coercion” test to measure Establishment Clause violations, id. at 632 (Scalia, J., dissenting)).

bound. But in the context of science and law, the criticism that scientists’ biases influence the hypotheses they test is particularly misplaced. In the end, it is the law that dictates which hypotheses merit study.

In *Witherspoon v. Illinois*, for example, the Court considered the question of the constitutionality of an Illinois statute providing that “[i]n trials for murder it shall be a cause for challenge of any juror who shall, on being examined, state that he has conscientious scruples against capital punishment, or that he is opposed to the same.”21 The challenger argued that common sense and the research available indicated that excluding jurors who oppose capital punishment (called “*Witherspoon*-excludables”) would result in a jury biased in favor of conviction. Justice Stewart, writing for the Court, agreed that this empirical question was constitutionally relevant and deplored the lack of data to answer it:

The data adduced by the petitioner . . . are too tentative and fragmentary to establish that jurors not opposed to the death penalty tend to favor the prosecution in the determination of guilt. We simply cannot conclude, either on the basis of the record now before us or as a matter of judicial notice, that the exclusion of jurors opposed to capital punishment results in an unrepresentative jury on the issue of guilt or substantially increases the risk of conviction.22

The *Witherspoon* Court, therefore, left open the question whether research might yet demonstrate that excluding those opposed to capital punishment from the guilt-phase of capital trials might produce panels that have a propensity for finding defendants guilty: “[A] defendant convicted by such a jury in some future case might still attempt to establish that the jury was less than neutral with respect to *guilt*.”23

---


23 *Id.* at 520 n.18 (emphasis in original). The Court continued as follows: “If he were to succeed in that effort, the question would then arise whether the State’s interest in submitting the penalty issue to a jury capable of imposing capital punishment may be vindicated at the expense of the defendant’s interest in a completely fair determination of guilt or innocence . . . .” *Id.*
The social science community’s response to the Court’s entreaty was extraordinary. Social scientists conducted more than a dozen reported studies on the effects of excluding jurors opposed to capital punishment. The near-consensus of the investigators and reviewers of this research corroborated the intuitive judgment of the petitioner in Witherspoon that excluding death-qualified jurors would result in conviction prone juries. The courts, therefore, have the power to influence the social science agenda. There is no question that an explicit, or even a veiled, call for data will cause social scientists to come to the Court’s assistance.

Whether such assistance will be heeded, or heeded well, is something that, history suggests, is questionable at best. Social scientists’ Witherspoon experience well illustrates the dangers associated with taking seriously the Court’s expressions of interest in data. In Lockhart v. McCree, the Court rejected both the validity and the relevance of the many studies done in response to the Witherspoon Court’s call for research. On the one hand, Chief Justice (then Justice) Rehnquist repudiated the validity of the fifteen studies McCree had introduced because of “several serious flaws” Rehnquist found in the research. On the other hand, Rehnquist stated that even assuming the validity of this research, “the Constitution does not prohibit the States from ‘death qualifying’ juries in capital cases.”

---

24 See Michael Finch & Mark Ferraro, The Empirical Challenge to Death-Qualified Juries: On Further Examination, 65 Neb. L. Rev. 21, 24 (1986) (“In the seventeen years following Witherspoon, death qualification has been one of the most studied subjects in the area of sociological jurisprudence.”). See generally William C. Thompson, Death Qualification After Wainwright v. Witt and Lockhart v. McCree, 13 Law & Hum. Behav. 185 (1989) (analyzing the Court’s treatment of social science in constitutional litigation concerning death qualification and discussing the future role of such research).

25 Finch & Ferraro, supra note 24, at 24-25.

26 Finch and Ferraro reported that the data supported three hypotheses:

(1) jurors excluded because of their inability to impose the death penalty are more attitudinally disposed to favor the accused than are non-excluded jurors; (2) excluded jurors are more likely to be black or female than non-excluded jurors; and (3) excluded jurors are more likely to actually acquit the accused than are non-excluded jurors.

Id. at 25.


28 Id. at 168-69.

29 For an in-depth discussion of these “flaws,” see Faigman, supra note 8, at 590-92.

30 Lockhart, 476 U.S. at 173.
Lockhart decision repudiated Stevens’s legal analysis in Witherspoon, and found that the research was irrelevant to the applicable constitutional provisions—the Sixth and Fourteenth Amendments—that applied in the case.\(^3\)

Although it is true that studying a phenomenon gives it status, lawmakers remain the ultimate arbiters of a phenomenon’s importance. In most cases, researchers take their cue from the agenda set by lawmakers. Even when social scientists are ahead of the law in identifying and studying factors of possible importance, lawmakers must independently assess the legal relevance of the factors identified. Certainly, policymakers should never defer to social scientists’ ordering of phenomena, just as they must guard against singling out for reliance certain factors simply because these factors have been the subject of scientific testing. As long as lawmakers are deciding the areas of importance, however, they have no ground to criticize the researchers’ fidelity.

2. Defining the Subject of Study

Underlying the realist perspective is the key methodological tool of replication. If cold fusion exists, for example, it should be demonstrable whether the researchers are in Provo, Palo Alto, or Princeton. The first lesson of scientific publication is that enough detail must be provided so that a reader could replicate the study. This requirement serves two essential functions. First, it permits what Karl Popper called inter-subjective testability.\(^3\) If the research findings have merit, other researchers in other settings should be able to obtain substantially the same results. Second, and of special concern to the law, this requirement informs readers regarding how the researchers concretely defined the object of their study. An essential step in science, therefore, is to make amorphous concepts concrete for the purpose of study by defining them operationally.\(^3\)

The need to operationally define terms is pervasive in science, whether it is physics or psychology. Consider concepts such as persistent vegetative state, intelligence, deterrence,

---

\(^3\) See Faigman, supra note 8, at 594-95 (discussing the Sixth and Fourteenth Amendment standards set forth in Lockhart).
\(^3\) See Fineman & Opie, supra note 14, at 125 n.50.
\(^3\) POPPER, supra note 2, at 44.
\(^3\) See PSILLOS, supra note 1, at 5.
bio-diversity, competence, violence, prurient interest, viability, involuntary euthanasia, reasonable expectations of privacy, and so forth. These terms are not self-defining. Scientists must somehow make concrete, for purposes of measurement, the vague and indefinite concepts of the law. Notions of justice, fairness, and equality are hardly self-defining. They must be operationally defined. Herbert Feigl nicely explained the matter as follows:

To put it briefly, if crudely, operational analysis is to enable us to decide whether a given term in the way it is used, has a “cash value,” i.e., factual reference. If it does have factual reference, operational analysis is to show us precisely what that factual reference is, in terms, ultimately, of the data of direct observation.35

The issue of defining abstract concepts operationally is not unique to social science. Moreover, usually there is some choice involved in how something should be operationalized, and a researcher should be obligated to explain why he or she made one decision rather than another. Consider the simple example of temperature. Temperature cannot be directly observed, but might be operationally defined as “the linear expansion of a mercury column in a glass tube of even width.”36 But this time-honored mode is not the only way to measure how hot or cold it is outside. For example, meteorologists might measure temperature by “windchill.” Windchill combines thermometer readings with wind speed and takes into account physiological factors, such as heat loss from the body (i.e., modern heat transfer theory).37

An indispensable part of evaluating any scientific research program ostensibly relevant to a legal matter, therefore, requires that lawyers ensure that the researchers studied the phenomenon that the law is interested in having studied. Consider, for example, the issue of children’s competency to make complex decisions, an issue that arises in a multitude of constitutional contexts.38 How can we be sure that the “competence” the courts speak about is the same

35 Herbert Feigl, Operationism and Scientific Method, 52 PSYCHOL. REV. 250, 252-53 (1945) (emphasis in original).
36 Id. at 254.
“competence” the social scientists measured in their research? The short answer to this question is that we cannot be sure; but a court can compare what the social scientists did with its own conception of competence. For example, in evaluating juveniles’ “competency” to waive their *Miranda* rights, Professor Thomas Grisso identified primarily three components of competency reflected in the legal literature: (1) comprehension of rights, (2) beliefs about legal context, and (3) problem solving style. Since Professor Grisso’s tests of competency are based on courts’ explanations of the concept, courts might be expected to find his results to be of some assistance to their original inquiry. The important point is that such a comparison can be made. Whether the psychological measure of competence adequately meets the legal conception of competence, therefore, can be evaluated by lawmakers who want to rely on the science. Hence, the law must, initially, identify the concept of interest and, in the end, decide whether scientists who have studied the concept of interest have done so adequately.

3. Methodological Choices Made

Science does not exist as a separate repository into which all well-founded knowledge is poured. Science is a dynamic enterprise that spans subject areas ranging from the lowly microbe to the grand universe. The scientific method, therefore, is not one method. It is an orientation or approach to empirical exploration. Different subjects demand different modes of analysis. Both electrons and electricians can be topics of scientific inquiry, but the particle physicists and industrial-organizational psychologists who study these respective subjects necessarily use very different techniques. But within areas of study, not all methods are equal, and they are not all employed equally well. Some methods provide brilliant probes into the operation of phenomena and others offer little more than dim glimpses of fleeting truths. Science, across the disciplinary landscape, from acoustics to zoology, is marked by


40 See HAACK, supra note 5, at 10 (“There is no distinctive, timeless ‘scientific method,’ only the modes of inference and procedures common to all serious inquiry.”).
methodological variability. An approach that is appropriate, or even possible, for a problem in celestial mechanics may be entirely inapplicable for a problem in cellular biology. Indeed, very often in science, a single paradigm will not be sufficient to study any particular phenomenon. New drugs are tested first in laboratory animals and second on humans, and both methods are tools of science.

Since scientific knowledge—or “the truth”—is only as good as the methods that researchers bring to bear to discover it, it behooves judges and lawyers to have some sophistication about those methods. The law, of course, relies on a wide assortment of scientific expertise, so it might be unrealistic to expect that judges will be able to develop proficiency in all of them. For example, the American Psychological Association’s amicus brief in the juvenile death penalty case of *Roper v. Simmons* advanced data from behavioral studies conducted by psychologists and brain-imaging studies done by neuroscientists. How can judges be expected to be critical consumers of such disparate forms of science?

The short answer is that they have no choice. The real question is not whether they can do it, but how they should go about doing it. The science exists and judicial decisions that ignore the empirical implications of the decision still have real-world consequences. A judge’s ignorance of causes might make him or her ignorant of the consequences of a particular decision, but the consequences still occur. Hence, for example, if research indicates a high false positive rate for predictions of violence in the civil commitment of sexually violent predators, ignoring this research, as the Court has done, does not alter the fact that many people are wrongly deprived of their liberty.

Although the task for judges appears daunting, it is not as difficult as it might first appear. First of all, most of the research that is introduced in constitutional cases is not rocket science. An elementary understanding of basic statistics and research methods will suffice in many cases to reveal the benefits and limitations associated with much of the research.

---


courts must consider. Second, especially in high profile constitutional cases, the Court will have considerable help in understanding the state of the art of the science. The Court suffers no shortage of amici when it comes to answering empirical questions about issues such as the onset of viability, whether a health exception is necessary to a ban on partial birth abortions, the developmental capacities of juveniles, the effects of physician-assisted suicide, the effects of virtual child pornography, and similar factual questions. Finally, although the Court is reluctant to employ this aid, all courts have the inherent authority to appoint experts to assist them with complex technical subjects.

4. Determining the Contents of Conclusions

A pervasive and troubling concern present in all scientific research, but particularly in the social sphere, is the danger that researchers will graft their values onto their conclusions. It may be assumed that, in many cases, what initially attracts researchers to legal problems is the hope to reform legal rules they view as “substantively” wrong. In researching the factual context of a legal rule with which social scientists disagree, they may unwittingly (or wittingly) interpret their data as more supportive of a particular normative position than the data actually compel. Although natural scientists share this source of difficulty, they do so to a lesser degree because a natural scientist’s inquiry tends to be less inherently value-laden. All scientists, whether natural or social, whose work potentially impacts public policy formation confront this issue. See, e.g., Jocelyn Kaiser, Taking a Stand: Ecologists on a Mission to Save the World, 287 Science 1188 (2000).

47 See Washington v. Glucksberg, 521 U.S. 702 (1997); see also infra notes 88-104 and accompanying text.
49 See David L. Faigman, Laboratory of Justice: The Supreme Court’s 200-Year Struggle to Integrate Science and the Law 360 (2004) (reporting interviews with Justices Stevens, O’Connor, and Breyer, who commented on “how unusual” the Court’s appointment of a technical adviser would be).
50 Nagel, supra note 15, at 488-89; see also John Passmore, Can the Social Sciences Be Value-Free?, in Readings in the Philosophy of Science 674 (H. Feigl & M. Brodbeck eds., 1953).
51 All scientists, whether natural or social, whose work potentially impacts public policy formation confront this issue. See, e.g., Jocelyn Kaiser, Taking a Stand: Ecologists on a Mission to Save the World, 287 Science 1188 (2000).
scientific social inquiry. At the science and law intersection, lawyers and social scientists share the burden of identifying and reducing the bias from research findings.

For their part, researchers must be forthcoming, possibly by stating explicitly their substantive biases entirely separately from their scientific findings. Of course, such a practice will never be totally effective because many value preferences are not fully known to the scientist, or their effect on the analysis is not fully understood. Nonetheless, a greater recognition of the problem will likely mitigate its effect. Additionally, social scientists should display more modesty when evaluating the significance of their findings. Sometimes researchers exude the confidence in their conclusions that their one study has settled the matter for the law. Rarely, if ever, is one study so conclusive that a legal rule can rest solely upon it.

Lawyers must also take responsibility for identifying bias where it occurs in empirical research. This means that lawyers must understand more than the conclusions advanced; they must also consider how the findings were obtained. For this purpose, the most important section for lawyers to read and understand in a scientific paper is the methods section. There, the researcher explains the design of the study, describes the sample population, defines—concretely (that is, operationally)—the question addressed, and describes the statistics used to measure subjects’ responses. The worth, or worthlessness, of a study can almost always be discerned from the methods section. Only if one understands how the study was conducted can one evaluate the soundness of the researcher’s conclusions.

The ability of readers of the scientific literature to identify errors due to extraneous factors should not be overstated, because some errant variables will not be observable in the methods, or any other, section. A multitude of unanticipated factors could influence the findings of a particular study or series of studies. But in the long-term, the ordinary checks inherent in the scientific enterprise can be relied upon to expose the biases, unconscious or conscious, of the researchers.

---

52 NAGEL, supra note 15, at 489.
53 Id.
54 See Michael J. Mahoney, Experimental Methods and Outcome Evaluation, 46 J. CONSULTING & CLINICAL PSYCHOL. 660, 660 (1978) (“The perfect experiment has yet to be designed and is, in some sense, inconceivable.” (citation omitted)).
Professor Nagel explained the dynamics of this system as follows:

[M]odern science encourages the invention, the mutual exchange, and the free but responsible criticisms of ideas; it welcomes competition in the quest for knowledge between independent investigators, even when their intellectual orientations are different; and it progressively diminishes the effects of bias by retaining only those proposed conclusions of its inquiries that survive critical examination by an indefinitely large community of students, whatever be their value preferences or doctrinal commitments.55

The conclusion that value biases influence the lessons researchers draw from their data is less surprising than the suggestion that lawmakers can be so easily misled by that bias. The methods and conclusions of social science research are like the premises and conclusions of legal argument: the validity of the premises must be determined in order to assess the soundness of the conclusions that the premises purportedly compel. Just as no good lawyer would accept a legal conclusion without examining the validity of the premises, no good lawmaker should accept research findings without examining how they were obtained.

5. Dividing Fact from Value

A fundamental criticism of scientific social inquiry concerns the assertion that fact and value are distinguishable in social inquiry. Critics argue that in studying purposive human behavior, value judgments invariably become intertwined with the descriptions of that behavior.56 Specifically, in the ordinary course of describing and categorizing events, social scientists cannot help but make evaluative judgments. The alternative of describing discrete factual events would be cumbersome, simplistic, and probably misleading. An arguable instance of this criticism is the psychological study of children’s competence, an issue discussed above. It might be argued that an unavoidable consequence of studying competence is the

56 See Fineman & Opie, supra note 14, at 130 (“The data can never be totally separated from the political, personal, and professional opinions of the person manipulating them.”); William C. Whitford, Critical Empiricism, 14 LAW & SOC. INQUIRY 61, 62 (1989) (asserting the impossibility of fully separating description from evaluation); see also Howard T. Markey, Jurisprudence or "Juriscience"?, 25 WM. & MARY L. REV. 525, 529 (1984).
inevitable value judgment required by that categorization. This criticism, however, misconstrues the evaluative role of social scientific inquiry.

Without denying that many researchers blur factual judgments and value judgments in the course of scientific inquiry, in principle these judgments can be kept distinct. Professor Nagel noted that confusion often arises from the failure to distinguish between “characterizing value judgments” and “appraising value judgments.”57 He provided the following example of a characterizing value judgment from biology:

Animals with blood streams sometimes exhibit the condition known as “anemia.” An anemic animal has a reduced number of red blood corpuscles, so that, among other things, it is less able to maintain a constant internal temperature than are members of its species with a “normal” supply of such blood cells. However, although the meaning of the term “anemia” can be made quite clear, it is not in fact defined with complete precision. . . . [T]o decide whether a given animal is anemic, an investigator must judge whether the available evidence warrants the conclusion that the specimen is anemic. . . . When the investigator reaches a conclusion, he can therefore be said to be making a “value judgment,” in the sense that he has in mind some standardized type of physiological condition designated as “anemia” and he assesses what he knows about his specimen with the measure provided by this assumed standard.58

In addition to the assessment that the animal is anemic, a biologist might assert that this condition is undesirable because of the animal’s inability to maintain itself. Professor Nagel referred to such expressions of approval or disapproval as “appraising value judgments.”59 To be sure, at times the terminology of social inquiry make fact/value distinctions difficult, with characterizing value judgments often implying appraising value judgments. But this point counsels caution. It does not contravene the capacity of social scientists to make the distinction.

By distinguishing characterizing value judgments from appraising value judgments, one can understand the factual nature of social scientific inquiry into children’s competence. For example, psychologists interested in children’s competence have gleaned certain characteristics from case law associated

57 NAGEL, supra note 15, at 493.
58 Id. at 492-93 (emphasis in original).
59 Id. at 493.
with various areas of legal competence. In studying this question, researchers typically compare children of different ages to adults (who are presumed competent by the law) on these factors. Subsequent characterizations of certain children as competent is analogous to the biologist’s characterizations concerning anemia in animals. The researcher classifies the subject group either within or outside the category of competence based on the identified characterizing criteria. Whether the psychological measure of competence adequately meets the legal conception of competence, therefore, remains challengeable separately in the same way that the factors characterizing anemia may be challenged. But in no respect does the characterizing value judgment that children of a certain age are “competent” entail a corresponding appraising value judgment.

In fact, psychological studies that find children as young as fifteen comparable to adults in their competency to make important decisions may be cited to support widely divergent legal conclusions. Whereas this research may support children’s participation in decisions of commitment to mental hospitals and autonomous abortion decisions, it may also support juveniles’ waivers of Miranda rights. Indeed, children’s competencies was a hotly disputed issue in Roper v. Simmons, in which Justice Kennedy cited social science research in support of exempting minors from capital punishment, while Justice Scalia, dissenting, decried the disingenuity of social scientists who proclaimed minors’ capacities in abortion cases but disavowed those capacities in capital cases. Although the cognitive ability of children is a

---

60 See, e.g., GRISSO, supra note 39, at 41-58.
61 See, e.g., id. at 95-97; Weithorn & Campbell, supra note 39, at 1591.
62 See Weithorn & Campbell, supra note 39, at 1596.
64 GRISSO, supra note 39, at 194 (noting research finding that juveniles between the ages of 15 and 16 with I.Q. scores above 80 “demonstrate[] a level of understanding and perception similar to that of 17- to 21-year-old adults for whom the competence to waive rights is presumed in law.”).
65 Compare Roper v. Simmons, 543 U.S. 551, 569 (2005) (“Three general differences between juveniles under 18 and adults demonstrate that juvenile offenders cannot with reliability be classified among the worst offenders.”) with Roper v. Simmons, 543 U.S. 551, 616-17 (2005) (Scalia, J., dissenting) (“[T]he American Psychological Association (APA), which claims in this case that scientific evidence shows persons under 18 lack the ability to take moral responsibility for their decisions, has previously taken precisely the opposite position before this very Court.”). See
scientific question, the legal consequences that befall competent and incompetent children remain policy choices. Psychologists and economists may be able to identify some of the consequences of choosing one course over another, but they can never offer scientific judgments on what effects are better avoided. Therefore, a court may continue to hold that fifteen-year-old children should not be consulted when committed to mental hospitals, even though they generally may be as competent to make important decisions as adults. This result can be justified by a concern for family autonomy or a recognition of competing parental rights. When confronted by conflicting value choices, courts must exercise their best judgment in light of all of the information available. Where relevant and valid, social science research can help clarify the available choices.

6. Assessing the Evidence

In addition to accusations that a researcher’s values affect her conclusions, critics claim that bias may enter into the very assessment of data. There are at least three distinct variants of this claim: First, a researcher’s social position and educational training influence the kinds of evidence deemed important. Second, the statistical decision rules employed by researchers mask important value choices. And third, a researcher’s relative “social perspective” impedes attempts to identify “universal” principles.

Whether a researcher’s social status affects the kinds of evidence he or she deems relevant to social inquiry is an empirical question. Some support may be expected for the assertion that a researcher’s socioeconomic, religious, and political views play some part in the assessment of data. But, as the previous discussion indicates, manifestations of such bias are recognizable by careful review of the measures applied
68 See NAGEL, supra note 15, at 495.
70 NAGEL, supra note 15, at 498-99.
71 Id. at 495-96.
in the research. Once prejudice is identified, a study’s findings may be discounted or dismissed accordingly.

A more technical objection to the problem of assessing evidence concerns the statistical rules that researchers use to decide if any effect has occurred. When comparing sample populations in order to determine whether some variable had an effect, two types of error are possible. A researcher may conclude that the factor of interest did have an effect when it did not (type I error, or “false positive”); or the researcher may conclude that the factor of interest had no effect when it did (type II error, or “false negative”). Social scientists are well acquainted with these sources of error and have devised various strategies to avoid them. The present criticism, however, is directed not at the possibility of error, but instead at the values employed when deciding to avoid one error at the expense of possibly committing the other.

Consider, for example, a group of hypothetical researchers who are interested in whether the death penalty is a deterrent. Hypothesizing that capital punishment lowers murder rates, they might compare states with capital punishment to a comparison group of states that do not. Upon comparison, the researchers find different murder rates between the two groups, but must decide whether they are “significant” enough to conclude that the death penalty made the difference; after all, some differences should be expected as a matter of chance. In assessing the data, the researchers must be cognizant of the possibility of committing one of the two types of error mentioned above. If they make a type I error, they will erroneously conclude that the death penalty had a deterrent effect when it did not. Alternatively, they might make a type II error, erroneously concluding that the death penalty had no deterrent effect when it did. Unfortunately, the researchers cannot eliminate or fully minimize the chance of making both types of error at the same time, and therefore must decide which error is more important to avoid. It appears, therefore, that researchers cannot avoid importing their values into the assessment of data.

Although this example illustrates a valid source of concern, the magnitude of the problem is not as great as it might first appear. Within the social sciences, certain conventions have arisen that minimize an experimenter’s independent judgment regarding drawing statistical conclusions from data. In particular, the much discussed convention of a .05 confidence level restricts researchers to the relatively conservative
risk of a five percent chance of making a type I error. Specifically, in the example above, this means that the researchers will mistakenly conclude that the death penalty has a deterrent effect five or fewer times out of a hundred if the death penalty is not a deterrent.\footnote{Put succinctly, the p-value is the probability of incorrectly rejecting the null hypothesis when the null hypothesis is true.} On some occasions, researchers might wish to lessen the risk of making a type I error by adopting a more conservative level of confidence, say one out of one hundred (.01). Similarly, less concern with making a type I error may lead a researcher to adopt a less conservative level of confidence, possibly ten out of one hundred (.10). Without question, scientists’ value preferences can affect the setting of confidence levels in a way that makes drawing a particular conclusion more or less difficult. Ideally, these judgments should be the responsibility of lawmakers. In any case, the standard selected should always be made explicit so that readers understand the decision rule the scientist applied in stating the conclusion. Departures from .05, and indeed even the decision to use a value of .05, should be scrutinized independently by anyone relying on a researcher’s findings. Once again, the important lesson is that a review of a researcher’s methodological discretion illuminates biases potentially affecting the reported findings.

The third and most “radical” claim that values influence the assessment of data maintains that a “necessary logical connection” exists between the researcher’s social perspective and the method and understanding of what is studied, rendering lessons from one time or place of no relevance to another time or place. Knowledge of societal or cultural facts, according to this view, is context specific. Therefore, the factual validity of a social finding can only be understood by knowing the society from which it emerged. As Professor Nagel explained the criticism, “there is no analysis of social phenomena which is not the expression of some special social standpoint, or which does not reflect the interests and values dominant in some sector of the human scene at a certain stage of its history.”\footnote{NAGEL, supra note 15, at 498.}

Although the claim typically excludes the natural sciences from its critical gaze, natural scientists also must state conclusions in a manner dependent on context. For instance, simply measuring the velocity of a stone dropped...
from a fixed point requires specification of the system of measurement used as well as a statement of the experimental conditions under which the measurement is taken. The situational dependence of this example is complicated further by adding the perspective of the observer. Albert Einstein provided the paradigmatic illustration of this complication:

I stand at the window of a railway carriage which is traveling uniformly, and drop a stone on the embankment, without throwing it. Then, disregarding the influence of the air resistance, I see the stone descend in a straight line. A pedestrian who observes the misdeed from the footpath notices that the stone falls to earth in a parabolic curve. I now ask: Do the “positions” traversed by the stone lie “in reality” on a straight line or on a parabola?\(^7^4\)

Absolute objectivity, it would seem, is unattainable even in natural science. There is thus no “God’s-eye-view” of the world that is discoverable by science, at least not without specifying the mountaintop on which He stands. Yet, substantial objectivity, or what Professor Nagel refers to as “relational objectivity,” is achieved when natural scientists identify invariant connections between factors. As a matter of logic, natural science can, and often does, identify relations which are demonstrable within the specifications established by experiment and which transcend particular value orientations or social perspectives.

The social sciences also operate in relationally specific contexts. To the extent that objectivity in the natural sciences depends on identifying and then transcending specific relational contexts, the social sciences, in principle, can do the same. Even though two sets of experimental results may be the product of separate social perspectives or value orientations, additional research may seek out “common denominators” from which results may be formulated, irrespective of the researcher’s initial perspectives.\(^7^5\) The complaint that the new synthesis suffers from a similar perspective-myopia can be admitted, though it is hardly the fatal flaw social science’s critics suppose. The goal for social science, as well as natural science, is relative objectivity, not absolute objectivity.

A researcher’s values and social perspective inevitably intrude into the identification of problems, the analysis of data,

\(^7^4\) _Albert Einstein, Relativity; The Special and the General Theory_ 9 (Robert W. Lawson trans., Holt 1920).

\(^7^5\) _Nagel, supra_ note 15, at 501.
and the conclusions drawn from the inquiry into social facts. Adhering to the scientific method in such studies perhaps provides only a limited, and not entirely satisfying, check on the interference of researchers’ biases. But, however imperfect the process might be, the benefits of a scientific social inquiry are worth the effort.

B. **Amalgamation of Facts and Norms in Constitutional Doctrine**

In addition to the complicating reality that research on social facts can be imbued with value preferences, courts and scholars regularly conflate facts and values in constitutional discourse. This is so in respect to both the language of the Constitution itself and the rules and standards that give the Constitution effect. The Constitution, for example, guarantees the people the right to peaceably assemble and to be secure against unreasonable searches and seizures. But what conduct passes as peaceable and what actions are unreasonable are not specified, though assemblages and searches and seizures are easily enough imagined empirically. In applying the necessarily imprecise words of the Constitution, the Court also regularly fashions tests that are composites of fact and value. The Constitution, for instance, limits congressional power to regulate commerce to “interstate commerce.” In one application of this doctrine, the Court asks whether the subject of regulation “substantially affects interstate commerce.” When the substantiability threshold has been crossed is a value judgment that, surprisingly, has garnered little serious scholarly attention.

In constitutional cases, therefore, the line dividing law and fact is not a bright one. Indeed, it is so dim that courts and commentators regularly fail to notice it. As a consequence, normative and empirical arguments in constitutional litigation tend to meld into one another and clarity is the primary victim. For instance, *Casey’s* undue burden standard, mentioned above, is explicitly contemplated as an empirically conscious test of the constitutionality of pre-viability abortion regulations. Yet in neither the case itself nor in subsequent case law has the Court adequately defined the normative component represented by the term “undue” or the empirical component represented by the term “burden.” Instead, the undue burden
standard, in practice, is ill-defined both normatively and empirically.76

In constitutional litigation, as is the case in most litigation, the division of responsibility between judge and trier of fact (jury or judge) is allocated on the basis of the nature of the issue presented. Matters of pure law are the exclusive responsibility of judges and matters of pure fact are resolved by triers of fact. The third category, and the one in which most constitutional facts fall, are mixed questions of fact and law. In ordinary litigation, these categories tend to define the line between judge and jury. In constitutional cases, in contrast, judges often—albeit not always—act in the dual capacity of determiner of law and finder of fact. Nonetheless, for a variety of reasons, the categories of pure law, pure fact, and mixed questions of fact and law, are an important component of constitutional adjudication.

Questions of pure law involve the interpretation of the Constitution and the setting forth of doctrine. In practice, purely legal questions thus concern the definition of the rules and standards that are applied in constitutional adjudication. Examples range from whether “fighting words” sometimes qualify as “speech” within the First Amendment77 to whether congressional actions must “affect” or “substantially affect” interstate commerce.78 In the Casey example from above, it was a purely legal question whether the Fourteenth Amendment’s protection of reproductive liberty should be implemented pursuant to the “undue burden” standard or the more traditional strict scrutiny test. This issue is reserved exclusively for judges to decide. This category of purely legal questions, therefore, encompasses all matters of doctrinal definition. It is under this doctrinal edifice that pure facts and mixed questions of fact and law are ultimately decided.

In non-constitutional cases, “case-specific facts” (also known as “pure facts” or “historical facts”)79 fuel the engine of

---

79 If forced to choose between these two terms, I prefer “pure facts,” though like the term “truth,” it promises more than it can possibly deliver. In constitutional
litigation. Such facts play a more modest role in constitutional cases. Indeed, although case-specific facts occur throughout constitutional law, they typically must be digested by applicable doctrine and come to be treated as mixed questions of fact and law. For example, in *Scott v. Harris*, the plaintiff-motorist claimed that the defendant police officer acted unreasonably and violated his Fourth Amendment rights when the officer forced the motorist off the road during a high-speed chase.\(^80\) As the Court explained, the circumstances of the car chase raised factual issues that had to be resolved in light of the evidence available and pursuant to ordinary rules of procedure.\(^81\) Once these case-specific facts were determined, the question whether the officer acted in an objectively reasonable fashion became a legal question. The process of evaluating the facts against the applicable legal standard is designated as a mixed question of law and fact. According to the Court, “[i]n determining the reasonableness of the manner in which a seizure is effected, ‘[w]e must balance the nature and quality of the intrusion in the individual’s Fourth Amendment interests against the importance of the governmental interests alleged to justify the intrusion.’”\(^82\)

As a practical matter, therefore, although case-specific facts can be readily identified in constitutional cases, they virtually always must be evaluated under some constitutional rule of decision. Whether this or that fact is constitutionally protected or unprotected is a legal question. In *Scott*, for instance, the plaintiff-motorist argued that the constitutionality of the officer’s action in forcing him off the road during a high-speed car chase should have been controlled by the outcome in *Tennessee v. Garner*.\(^83\) In *Garner*, the Court found that an officer had used excessive force when he shot a fleeing unarmed burglary suspect in the back of the head.\(^84\) The constitutional issue presented in *Scott* and *Garner* was the

---


81 In *Scott*, the plaintiff had lost at summary judgment, so the relevant facts had to be determined with “all inferences in favor of the nonmoving party to the extent supportable by the record.” *Id.* at 1776 (emphasis omitted).

82 *Id.* at 1778 (quoting United States v. Place, 462 U.S. 696, 703 (1983)).


same. Given the facts presented,\textsuperscript{85} did the police act in an objectively reasonable way? The \textit{Scott} Court concluded, however, that the same rule-of-decision dictated different outcomes. The Court stated that “the threat posed by the flight on foot of an unarmed suspect [was not] even remotely comparable to the extreme danger to human life posed by [the fleeing motorist] in this case.”\textsuperscript{86}

As \textit{Scott} illustrates, ordinary case-specific facts can be incorporated into constitutional lawmaking through the decisional rules established pursuant to the Constitution. Hence, the plain facts of the car chase were evaluated under the normative reasonableness standard contained in the Fourth Amendment. But constitutional facts are rarely as plain or unambiguous as \textit{Scott}'s car chase, which was captured on video. Indeed, most facts having constitutional relevance are not simple case-specific facts, like \textit{Scott}'s car chase, but instead are complex scientific or general historical facts, like viability or the intentions of the drafters of the Fourteenth Amendment or the effects of violent television on young viewers.

Although they might be the subject of numerous and intricate scientific studies, facts such as “viability” or the “effects of violent television” can still be identified separately and then integrated into constitutional lawmaking.\textsuperscript{87} Consider, as a case study, the complex empirical issues surrounding physician-assisted suicide that were presented in \textit{Washington v. Glucksberg}.\textsuperscript{88} The plaintiffs in \textit{Glucksberg} claimed that a person who is terminally ill and mentally competent should have the right to choose what form his or her death would take, and have the right to a physician’s assistance in exercising that right.\textsuperscript{89} The plaintiffs claimed that the Fourteenth Amendment’s Due Process Clause extended to “the liberty to choose

\begin{itemize}
  \item \textsuperscript{85} The division of the issue into “pure facts” and the legal resolution of those facts raises the question of the respective responsibilities between trier of fact and judge as a practical matter. Space precludes consideration of this issue in the essay.
  \item \textsuperscript{86} \textit{Scott}, 127 S. Ct. at 1777.
  \item \textsuperscript{87} The complexities surrounding historical facts, such as the original intentions of the drafters of the Second Amendment, are beyond the scope of this essay. As a general matter, however, the methods of historical exploration are sufficiently similar to the methods many social scientists use in that most of the lessons drawn here regarding science and scientists should apply to history and historians as well. \textit{See HAACK, supra note 5, at 24 (comparing the methods of science to “what historians or detectives or investigative journalists or the rest of us do when we really want to find something out”).}
  \item \textsuperscript{88} 521 U.S. 702 (1997).
  \item \textsuperscript{89} \textit{Id}. at 722.
\end{itemize}
how to die,” or “the right to choose a humane, dignified death.”\footnote{Id. at 703.} The Court, however, with Chief Justice Rehnquist writing, defined the asserted right as the “right to commit suicide which itself includes a right to assistance in doing so.”\footnote{Id. at 723.}

It is, of course, a basic principle of hornbook law that the existence or non-existence of a fundamental right ordinarily dictates the level of judicial review that is accorded a disputed state action. Hence, if the plaintiffs’ view prevailed, so that the right to physician-assisted death was deemed a fundamental right, the state would be obligated to justify infringements or limitations of the exercise of that right by demonstrating that its action was narrowly tailored to advance a compelling government interest. If the right was deemed a liberty interest but not fundamental, the state’s burden would be substantially lighter, and would require only that the action was rationally related to a legitimate government interest. The Court had little difficulty in finding that the right, given its definition as the right to commit suicide, was not fundamental. “[F]or over 700 years,” Rehnquist explained, “the Anglo-American common law tradition has punished or otherwise disapproved of both suicide and assisting suicide.”\footnote{Id. at 711.} Under the Court’s interpretation, therefore, the state did not have to advance and prove a strong justification for its prohibition of assisted suicide. But the legal posture of \textit{Glucksberg} turns out to be more complicated, as the concurring opinions of Justices Stevens and Souter make clear.

Justice Stevens wrote separately to emphasize that the Court’s holding did not preclude later protection of a terminally ill patient’s right to assistance in hastening death. The Court, he stated, merely found that the Washington statute prohibiting suicide was not invalid “on its face.”\footnote{Id. at 739 (Stevens, J., concurring).} The Court’s decision “does not foreclose the possibility that some applications of the statute might well be invalid.”\footnote{Id. at 739 (Stevens, J., concurring).} In particular, Stevens contemplated a case in which a mentally competent person who is terminally ill and suffering excruciating pain seeks a physician’s help to facilitate the end. “The liberty interest at stake in a case like this,” Stevens said, “is an interest in

\footnote{Id. at 703.} \footnote{Id. at 723.} \footnote{Id. at 711.} \footnote{Id. at 739 (Stevens, J., concurring).} \footnote{Id.}
deciding how, rather than whether, a critical threshold shall be crossed.”95

Justice Souter also wrote separately to point out that Glucksberg sits atop a constitutional fault-line that might shift as our understanding of the empirical landscape changes. Souter sought to reconcile Rehnquist’s majority opinion finding no right to assisted suicide and Stevens’s presaging the next case down the line, which presents the sympathetic situation of the competent terminally ill patient in debilitating pain who wants to choose a dignified end to a dignified life. Souter wrote that the core concern in these cases was fact-based.

Souter asserted that the state has no interest in denying a competent terminally ill patient in debilitating pain his or her choice of how to die. According to Souter, the state’s legitimate interest lies in averting mistakes, in precluding assisted suicide from becoming directed suicide. The state thus rationally fears the slippery slope that once a procedure is set in place that permits some to freely choose death, others will be encouraged or even forced into this choice. “The nub of this part of the State’s argument is not that such patients are constitutionally undeserving of relief on their own account, but that any attempt to confine a right of physician assistance to the circumstances presented by these doctors is likely to fail.”96

Whether the state is correct that compassionate assistance in dying ineluctably leads to involuntary euthanasia is an empirical question. In the majority opinion, Chief Justice Rehnquist cited and discussed at length a study from the Netherlands designed to test this hypothesis. According to Rehnquist, the 1990 Dutch study reported “2,300 cases of voluntary euthanasia (defined as ‘the deliberate termination of another’s life at his request’), 400 cases of assisted suicide, and more than 1,000 cases of euthanasia without an explicit request.”97 More profoundly disturbing, Rehnquist reported that, in addition to those 1,000 cases, “the study found an additional 4,941 cases where physicians administered lethal morphine overdoses without the patients’ explicit consent.”98

Rehnquist concluded that the Dutch study

---

95 Glucksberg, 521 U.S. at 745.
96 Id. at 754 (Souter, J., concurring).
97 Id. at 754 (Souter, J., concurring).
98 Id. at 734 (majority opinion).
suggests that, despite the existence of various reporting procedures, euthanasia in the Netherlands has not been limited to competent, terminally ill adults who are enduring physical suffering, and that regulation of the practice may not have prevented abuses in cases involving vulnerable persons, including severely disabled neonates and elderly persons suffering from dementia.99

Justice Souter, however, found the empirical record more mixed. On the one hand, some commentators found that Dutch guidelines have “proved signally ineffectual; non-voluntary euthanasia is now widely practised and increasingly condoned in the Netherlands.”100 On the other hand, some researchers have found the opposite, that “Dutch physicians are not euthanasia enthusiasts and they are slow to practice it in individual cases.”101 Souter concluded that he could not say with any “assurance which side is right.”102

Ordinarily, when the fact of the matter is uncertain, procedural burdens of proof allocate the risks of error and guide decision making. Standards of proof should be based on underlying normative considerations associated with the costs of making a mistake. This is why the burden of proof in civil cases is the preponderance standard, but is beyond a reasonable doubt in criminal cases. In constitutional cases, such allocation ought to depend on the constitutional values found to be implicit in the text. In Glucksberg, this issue depended on the Court’s interpretation of the due process clause. The main area of focus, as set forth in the Stevens and Souter opinions, was whether the due process clause might extend protection to a competent, terminally ill patient, who was in debilitating pain. According to Stevens, “[a]voiding intolerable pain and the indignity of living one’s final days incapacitated and in agony is certainly ‘at the heart of [the] liberty . . . to define one’s own concept of existence, of meaning, of the universe, and the mystery of human life.’”103 If Stevens’s position is correct, then the state should have the burden to demonstrate that procedural protections cannot be enacted to

99 Id.
100 Id. at 786 (Souter, J., concurring) (internal quotation marks omitted) (citing and quoting John Keown, Euthanasia in the Netherlands: Sliding Down the Slippery Slope?, in EUTHANASIA EXAMINED 261, 289 (John Keown ed., 1995)).
101 Id. (internal quotation marks omitted) (citing and quoting RICHARD A. EPSTEIN, MORTAL PERIL: OUR INALIENABLE RIGHT TO HEALTH CARE? 322 (1997)).
102 Id.
103 Id. at 745 (1997) (Stevens, J., concurring) (alterations in original) (citing and quoting Planned Parenthood of Se. Pa. v. Casey, 505 U.S. 833, 851 (1992)).
avoid involuntary euthanasia if it seeks to entirely proscribe assisted suicide for the terminally ill, competent patient suffering intolerable pain. Accordingly, the Constitution would guarantee the right of physician-assisted suicide in a select group of cases until states adequately demonstrated that procedural mechanisms were unavailable to avoid the slide into involuntary euthanasia.

Oddly, Justice Souter failed to follow this basic logic when he reached the issue of how the empirical question might be resolved in future cases. According to him, the Court’s decision regarding whether the right to die is constitutionally based should await state experimentation to determine whether there is a workable stopping point between assisted suicide and involuntary euthanasia. He stated that the Court should “stay its hand” until the state legislatures had ample opportunity to study the question. But Souter essentially put the horse behind the cart. The existence of a constitutional right to assisted suicide should not depend on whether procedural protections can be constructed to avoid having the right to die turn into the duty to die. This factual issue, the subject of the Dutch research and Souter’s hoped-for subject of future American research, concerns the government’s interest in curtailing the claimed right to die. The fundamental right to autonomy over death, if it exists, exists separately from the state’s claimed reasons for regulating or prohibiting it.104

By analogy, states have often sought to regulate violent pornography on the basis that it makes consumers of it more prone to be violent.105 Violent pornography falls within the protection of the First Amendment’s guarantee of free speech. This means that violent pornography cannot be prohibited until the state demonstrates empirically that it causes violence. Thus, the right is protected first, and government claims of compelling reasons to permit regulation of it must be proved—a demand that might take considerable time and effort on the

104 This practice of using a state’s asserted justification for infringing a right as a basis for finding that no right exists in the first instance is not unique to Justice Souter. It can be found throughout the Court’s cases. See Lawrence H. Tribe and Michael C. Dorf, Levels of Generality in the Definition of Rights, 57 U. CHI. L. Rev. 1057, 1096-97 (1990). Its frequency, however, does not render it legitimate. See David L. Faigman, Reconciling Individual Rights and Government Interests: Madisonian Principles Versus Supreme Court Practice, 78 Va. L. Rev. 1521, 1539 (1992).

105 See, e.g., Am. Booksellers Ass’n, Inc. v. Hudnut, 771 F.2d 323 (7th Cir. 1985) (invalidating Indiana’s anti-pornography statute, which, among other things, proscribed violent pornography).
part of legislatures. Contrary to Souter’s argument, this has not interfered with public officials studying the issue of the effects of violent pornography and, indeed, this topic has been the subject of substantial research attention as well as two presidential commissions. In fact, placing the burden on legislatures is likely to produce more research, not less, since states need to generate evidence to justify their legislation. Hence, the free speech right exists and continues to be protected until legislatures develop sufficient proof to demonstrate a compelling interest to justify infringements of that right. There is no reason why exactly the same sort of analysis should not apply to assisted dying.

Recognizing that the Constitution evolves as society—and, more particularly, our factual understanding of society—changes does not make the Constitution any less “durable” than Souter’s institutional deference to legislatures in Glucksberg would make it. In his universe, the Constitution “changes” if the legislative answer is that procedural protections can be instituted to ensure that assisted dying does not become forced euthanasia. At that point in time, the Glucksberg ruling would have to be “amended” to permit the right to die so long as it is accompanied by whatever procedural protections the states come up with to prevent involuntary euthanasia. In the alternative, the Constitution “changes” if assisted dying is protected today, but legislatures demonstrate tomorrow that procedural controls are ineffective. States would have demonstrated that they have a compelling interest in prohibiting all assisted suicides, because the practice cannot be limited to the small group in which it is appropriate. The only question is what is to be the default position. In the absence of sound empirical research one way or the other, does assisted suicide receive constitutional protection or does it not? The empirical question of the availability of procedural controls adequate to avoid involuntary euthanasia must be evaluated in light of the answer to this question. This is a matter of constitutional interpretation. Science cannot say what the Constitution means, but it can provide a window into the world so that constitutional values can be justly applied. Science thus informs the constitutional analysis.

---

In the example of physician-assisted suicide, science should play a pivotal role in deciding actual cases. But it does not displace classic constitutional value definition. Indeed, it should clarify it. This issue presents an archetypal clash of irreconcilable principles. On the one hand, the state declares an abiding and weighty interest in protecting life, including especially the weak and vulnerable who might be hastened toward death in a general scheme permitting assisted suicide. On the other hand, due process guarantees individuals the liberty to make decisions regarding core attributes of their lives, which, under certain circumstances, includes how their lives end. Although the weight of the state’s interest in life or the magnitude of the costs associated with an individual’s loss of liberty if forced to endure his or her final days in intolerable pain are normative judgments, the empirical realities endemic to this clash of principles are readily determinable. A fair and just determination of constitutional cases cannot be achieved without a sound and accurate understanding of the world to which those decisions apply.

When facts are relevant under particular constitutional rules or standards, courts should strive to define and understand them separate from the constitutional norms that apply. If a regulation operates as an obstacle to the exercise of a woman’s constitutionally protected right to a pre-viability abortion, for instance, that fact can be independently determined. Once it is determined—albeit with all of the limitations and caveats associated with doing this research—courts can separately resolve whether the obstacle is “substantial” or the burden it puts on the right is “undue.” Similarly, the factual question of whether physician-assisted suicide increases the incidence of involuntary euthanasia is a component of, but independent from, the constitutional norm of whether physician-assisted suicide is a protected fundamental right. Even the most fundamental of rights, such as political speech, can be regulated if the government’s interests are sufficiently compelling. Good research on physician-assisted suicide should demonstrate whether the dangers of this practice provide a compelling justification for prohibiting it. By keeping constitutional value-definition separate from constitutional fact-finding, the analytical bases for constitutional outcomes will be clearer and thereby more legitimate.
III. Conclusion

The question whether a world exists independent of our minds’ perception of it would probably appear quite absurd to the average lawyer or judge. The justices of the Supreme Court would undoubtedly be amazed to be asked such a question. They almost certainly subscribe to basic realist tenets, at least in the belief that the world is mind-independent. The justices would probably also share the realists’ belief that while researchers’ values sometimes affect the conclusions they draw about the mechanics of the “real world,” the methods of science are well designed to limit or reveal those biases.

Yet, despite its likely realist orientation, the Court repeatedly treats facts in constitutional cases in anti-realist ways. In particular, the Court describes the factual world constructively, so that the facts serve normative or interpretive ends. The Court appears largely unconcerned with the actual reality of the factual premises it relies upon. Anti-realists may believe that this is an inevitable consequence of the task the Court faces in integrating highly complex empirical information into the intricacies of constitutional doctrine. In this essay, I argue that it is possible for the Court to employ a scientifically realist constitutional jurisprudence. In so concluding, I consider two principal challenges to a realist approach in constitutional cases. First, I reject the argument that the sorts of facts the Constitution makes relevant—primarily social and behavioral facts studied by social scientists—cannot be studied objectively. Second, I describe how constitutional tests that appear to be conglomerations of facts and values can be separated into their component parts. While these challenges are formidable, they are not insurmountable.

Constitutional rulings should be defined by the real world because they define the real world. The Constitution is an eminently practical document. Although cast for immortality, it is grounded in modern times and must attend to contemporary circumstances. As John Marshall put it, the Constitution was “intended to endure for ages to come, and, consequently, to be adapted to the various crises of human affairs.” McCulloch v. Maryland, 17 U.S. (4 Wheat.) 316, 415 (1819) (emphasis in original).
The Discovery of Truth in Context

COMMENTS ON FAIGMAN, KATSKEE, AND KEIL

Sam Glucksberg†

Truth is hard to come by, even in optimal circumstances where the criteria are explicit and clear, and where it can be objectively established whether those criteria have been met (at least in principle). Consider baseball. When a batter does not swing at the ball, the umpire must make the call. Is it a strike or a ball? The criteria are explicit: the ball must be within the strike zone. The batter knows this, the pitcher knows this, the fans know this, and the umpire, of course, also knows this. The criteria are clear. In principle, one could have the decision made by machine, as it is often done in tennis. Despite the empirical clarity and explicit decision rules, there are still three ways of construing the truth every time the umpire makes the call. The first might be called objective realism: when queried, an umpire who is committed to this position would say, “I call them as they are.” The second, subjective realism: “I call them as I see them.” The third, what might be called declarative realism: asserted by one of the greatest umpires of his time, Bill Klem, “It ain’t nothin’ till I call it.”

From my layperson’s point of view, many judicial decisions, even those informed by scientific realism, seem to be somewhat akin to Klem’s decisions. They are constrained by the relevant criteria and by the “objective” facts, yet they ultimately depend on human judgment. A case that comes to

† Professor of Psychology Emeritus, Princeton University.

mind is a decision by the International Court of Justice. That
court decided that the massacre of Bosnian Muslims at
Srebrenica in 1995 was an act of genocide, but that Serbia
itself was not guilty of the crime. Judicial decisions such as
this constitute a particular type of speech act. According to the
philosopher of language, John Searle, there are roughly five
kinds of speech acts: assertives, whereby we describe a state of
the world such as “today is Friday”; directives, where we
request something, such as “could you please hand me a glass
of water”; commissives, whereby we commit ourselves to some
course of action, such as “I'll have chicken tonight”; expressives,
whereby we tell people how we feel about something, such as “I
really enjoyed today’s sessions”; and, most relevant to this
discussion, declarations, whereby we accomplish something
merely by saying something. Bill Klem’s shouting “Steerike!”
declared that pitch as a strike, just as the International Court of
Justice’s declaration that genocide had occurred did not simply
label an act, but also defined the 1995 massacre as an act of
genocide.

These two examples illustrate not only the nature of
declaratives, but also that most speech acts accomplish more
than one thing at a time. In these examples of calling a strike
and declaring an act of genocide, the speakers are not only
describing a state of the world but characterizing those states.
Calling a strike creates a strike and purports to describe a
state of the world—the ball is asserted to have been in the
strike zone. Declaring an act of genocide not only creates a new
act of genocide, but also asserts that the act did in fact violate

I do not know if Professor David Faigman would agree,
but it seems to me that the concept of scientific realism and
issues of factual “truth” share some of the properties of these
eamples. Scientific realism is the assumption that there is a
real world independent of our minds that can be studied
scientifically. At the very least, the notion of scientific truth
should encompass not just fact finding, but also fact creation—
not just asserting, but declaring that an act meets the criteria

---

2 Marlise Simons, Court Declares Bosnia Killings Were Genocide, N.Y.
3 JOHN R. SEARLE, EXPRESSION AND MEANING: STUDIES IN THE THEORY OF
4 For elaboration of this concept, see generally David L. Faigman, Scientific
for, say, genocide. This not only declares that a particular act meets those criteria, but also creates a new member of that category of acts.\(^5\)

Faigman’s discussion of scientific realism and the place of facts in the context of constitutional interpretation reminds me of the debates in economics concerning another kind of realism, “behavioral realism.” Behavioral realism refers to the use of scientific knowledge about people to evaluate assumptions (often unfounded) about human nature. The work of such scholars as Amos Tversky, Danny Kahneman, and Richard Thaler,\(^6\) among many others, raises important challenges to the concept of “rational man,” culminating only recently in the establishment of behavioral economics as a recognized discipline.\(^7\) While by no means universally accepted, more and more economists are recognizing the relevance of behavioral and social science findings to economic theory and practice. Again, from my layperson’s perspective, scientific realism seems to be a necessity not only in economics but also in legal interpretation, especially in those contexts where abstract concepts such as equality and liberty must be instantiated anew as time passes and the world changes.

From a psychological/behavioral standpoint, people’s understandings of abstract concepts are rarely, if ever, explicit. Instead, such concepts can be inferred by examining how they are instantiated in a given context. As Faigman clearly points out, as times and circumstances change, people’s constructions of concepts such as segregation also change. The notion that segregation is inherently unequal provides a good example of unanticipated instantiation. Consider the problem of warring gangs in prisons: when gang members are assigned cell blocks on the basis of their race or ethnicity, is this an instance of “segregation”? Just as declaring that the murder of Bosnian Muslims was genocide, declaring that the concept of segregation is or is not applicable to the gang situation instantiates that concept in a new way. The transcendent concepts listed by


Faigman, such as free speech, free exercise of religion, equal protection, and due process, may well be eternal truths, but those truths must necessarily be instantiated in concrete cases as people decide whether situation X is a case of free speech or due process or equal protection. Declaring or not declaring it so is analogous to calling a ball or a strike; in a very real sense, saying so makes it so. And because our world is dynamic and changing, such decisions must always be made anew (Justice Thomas notwithstanding). In short, Faigman’s arguments are remarkably consistent with what we know in cognitive science about the nature of human concept learning, concept structure, and concept instantiation and application.

Because general concepts must be instantiated anew with every new context, there is ample opportunity for biases to shape and color each new instantiation. Faigman’s analysis of biases in social science research is consistent with what we know of bias effects and how we try to minimize them. The opportunities for biases to affect the selection of problems, the definition of the subjects of study, determination of what to conclude, discriminating between fact and value, and assessment of the validity and applicability of evidence are virtually endless. The first step to minimize bias is to explicitly recognize that there is a distinction between fact and value. Further, and not incidentally, this requires explicit adherence to the proposition that there is a real world, independent of our minds, and that we can discover facts about that world. Then, as in all sciences, we must do what we can to minimize bias in every phase of our work. For example, in both behavioral and medical research, we employ double blind studies to minimize both observer and participant bias effects. We use placebos in medical research to isolate the effects of expectations from the effects of the treatment under investigation. As Faigman observes, even the “hard” sciences such as physics recognize the interactive effects of observer, observer position, and the phenomena under study. I agree with his analyses and concerns, and with his conclusion: “Adhering to the scientific method . . . perhaps provides only a limited, and not entirely satisfying, check on [the effects] of researchers’ biases. But, however imperfect the process might be, the benefits of scientific social inquiry are worth the effort.”

8 Faigman, supra note 4, at 1089-90.
9 Id. at 1091.
Faigman points out that all scientists, not just social scientists, are subject to the biases that might be generated from one or another normative positions. But I am not convinced that “a natural scientist’s inquiry tends to be less inherently value-laden.” For starters, all scientists usually start off with a favored hypothesis, and natural scientists’ commitments to their pet theories are no less strongly held than those of social scientists. When a hypothesis is firmly held, it makes not a whit of difference if the science is physical or social: observer error driven by bias must always be minimized by following appropriate procedures. A prime example is provided by Alfred Binet in his efforts to determine if intelligence is related to brain size. Binet firmly believed that children with larger brains were more successful in school than those with smaller brains, and his measurements of skull size (as a surrogate variable for brain size) seemed to confirm his hypothesis. As quoted by Stephen Jay Gould, Binet could not be more sure: “The relationship between the intelligence of subjects and the volume of their head . . . is very real and has been confirmed by all methodical investigators, without exception . . . we conclude that the . . . [correlation between head size and intelligence] must be considered as incontestable.” But he was too good a scientist to accept his first findings and conclusions. Two years later, he wrote:

I feared that in making measurements on heads with the intention of finding a difference in volume between an intelligent and a less intelligent head, I would be led to increase, unconsciously and in good faith, the cephalic volume of intelligent heads and to decrease that of unintelligent heads.

. . .

Suggestibility . . . works less on an act of which we have full consciousness, than on a half-conscious act—and this is precisely its danger.

Note that skull measurement can serve either natural or social science purposes. If one’s hypothesis concerns the relation between skull size and brain maturation, then it may qualify

---

10 *Id.* at 1082.
11 *Id.*
13 *Id.* at 177 (emphasis added) (quoting A. Binet, *Recherches sur la Technique de la Mensuration de la Tête Vivante*, 7 *L’année Psychologique* 314, 323-24 (1900)).
as natural science. If the relation is between skull size and social skills, then it may qualify as social science. The opportunity for bias in measurement and for drawing conclusions is not inherently different in these two cases. One need only turn to the controversy involving evolution and creationism to find another clash of values affecting data interpretation in a natural science. Fortunately, both natural and social science provide methodological safeguards to minimize bias in every phase of the scientific enterprise, from problem selection to measurement to drawing inferences and implications from data.

Parallel to the effects of bias in scientific research, Faigman makes a persuasive case for the effects of bias in courts’ selection and interpretation of scientific findings to justify, or perhaps to rationalize, their decisions. The use of neurological evidence that the brains of eighteen-year-olds are still developing to support a decision to exclude the death penalty for people eighteen or younger is a case in point. But if the continuing development of the brain is the definitive factor, then twenty-five-year-olds should also be spared. But of course, scientific evidence is not the only factor to be taken into account. Surely there is a place for community standards, values, and ethics, in addition to facts, in legal decision-making. Distinguishing between fact and value does not necessarily mean that we should, or even can, exclude value as an important factor. But it is important to explicitly recognize the use of values and not pretend that it’s “just the facts” that drive judicial decisions.

Professor Lawrence Solan provides revealing examples of how judges interpret language to serve their individual, value-driven positions on constitutional law. One of the most striking examples is the interpretation of the word “use” in the context of illegal drug transactions. Prison sentences for those convicted of drug offenses are often longer if a gun was “used” during a drug transaction. In one case, a convicted person appealed his sentence by claiming that his “use” of a gun did

---

14 Still, there is the lingering intuition that social science is not as “hard” as natural science, as reflected in W.H. Auden’s cynical remark, “Thou shalt not commit a social science.” W.H. AUDEN, “Under Which Lyre,” reprinted in COLLECTED POEMS 335 (Edward Mendelson ed., 1991).
15 Faigman, supra note 4, at 1084-85.
16 See id. at 1086-87.
not fall under the gun-use sentencing provision because he used the gun for barter, offering it in lieu of additional money to buy drugs during a sting operation. In effect, he did not use the gun as a weapon, but as a bargaining instrument. The United States Supreme Court in a majority decision ruled that the term “use” covered any use of a gun, whether as a weapon or not, and declined to interpret “use” in terms of an inferred legislators’ intent of use-as-a-weapon. I am confident that Faigman could cite many such examples, and that he would not be surprised by this one.

On a related issue, Richard Katskee distinguishes between revealed truth and scientific truth. I think that most of us would agree with his characterization that received truth and scientific truth are incommensurate. But that does not mean that a given individual cannot hold two sets of beliefs: one based on religion and the other on science, even when those two beliefs are glaringly contradictory. According to biblical teaching, the universe was created by God and the earth is at most 10,000 years old. Young-earth creationists believe this. According to contemporary paleontology, though, the earth is hundreds of millions of years old, and events can be dated using fossil records and various other dating techniques. According to a recent Ph.D. dissertation, a species of marine reptiles, mosasaurs, vanished at the end of the Cretaceous era about 65 million years ago. Palentologists believe this. What is intriguing about these two beliefs about the age of the earth is that one man, Marcus Ross, believes that both are true. Ross is the author of the dissertation on mosasaurs, even though he identifies himself as a young-earth creationist. How does he reconcile his two completely different sets of beliefs? These views can coexist because, as Ross put it, he is “separating the different paradigms.”

Can people in general separate their religious and scientific paradigms, and if they can, will they then rely exclusively on the scientific paradigm in the legal and political

---

19 Id. at 240-41.
20 By “revealed” or “received” truth, I refer to truth given by one or another dogma, including religion, that is not subject either to logical or empirical test.
22 See Id.
23 Id.
24 Id.
realm? When it comes to values and strongly held religious beliefs, I seriously doubt it; witness the perennial disputes about teaching evolution, providing information on sexually transmitted diseases, making abortion safe and available, even the inclusion of the word “God” in our pledge of allegiance. Separation of revealed and scientific truth is relatively easy in principle. Separation in practice may well be impossible—and for very good reasons. Scientific truth can adjudicate decisions on how to implement policies. But the policies themselves often stem from values and moral beliefs, and these are, in the end, not provided by scientific fact or theory. My guess is that the best we can do is recognize the differences between science and religion, and try to keep those distinctions in mind when evaluating conflicting views in the legal and political realms.

To whom do we turn when we are in doubt about either scientific or religious truths? Professor Frank Keil raises an important issue: when do we know that we do not know? This question is the province of the field of metacognition, the study of how we assess what and how we know things. As Keil observes, people are not very good at assessing their own states of explanatory knowledge. A few familiar household examples should suffice. What do people know of the operation of thermostats? In a study of how to get people to conserve energy in heating their homes, people were asked how they used their thermostats. More than half the people interviewed set their thermostat five degrees higher than the temperature they really wanted in the belief that the house would heat up faster that way. Then, when the house was warm enough, they would reset the thermostat to the desired resting level. This invariably wasted energy because people rarely noticed when the temperature they actually desired had been reached or exceeded. In an informal follow-up to that study, I asked several people to explain how their toilets worked. Most had only the vaguest ideas, mentioning pumps, propellers, suction devices and the like. Yet most of us, including the people in these studies, feel that we know about thermostats, toilets, and other common household devices well enough. Only when our

26 Id. at 1037-41.
27 See generally Willet Kempton, Two Theories of Home Heat Control, 10 COGNITIVE SCI. 75 (1986).
knowledge is probed beyond mere surface level is our relative state of ignorance exposed.

Keil astutely points out the relevance of this lack of knowledge, and the lack of awareness of that lack of knowledge, in legal contexts. What is common knowledge to experts and to some lay people is certainly not common knowledge to everyone. Even more important, what may seem like common knowledge to someone could actually have been learned very recently, even within the proceedings of a court case. Jurors can often learn something in the course of trial testimony and then, a few minutes later, be under the impression that they had known that “fact” all the time. This is a well-known phenomenon in cognitive and social psychology, the “hindsight bias.” Hindsight bias refers to the tendency of people who learn something new that seems commonsensical to come to believe, sincerely believe, that they had known it all along. This bias may have its roots in early childhood, where the analogous phenomenon is observed in a “false belief” context.

In this context, a child, say David, is shown a candy box and asked, “What’s in it?” David replies, of course, “Candy.” The box is then opened, and lo and behold there are crayons, not candy, in the box. David is then asked what his friend Tommy will think is in the box if he is shown the closed box. David’s answer? “Crayons!” This is hindsight attributed to another person. And, if we now ask the first child, David, what he thought was in the box before it was opened, he says, with great confidence, “Crayons!”

While not so extreme as this example, adults in the context of jury deliberations and decision-making will not only fall prey to the hindsight bias, believing that X is something they knew all along, but also that X is common knowledge, something that everyone knows or should know. For example, it is not common knowledge that using a racing (shallow) dive into a four-foot deep swimming pool can result in traumatic spinal cord injury and quadriplegia. This is because the neck can flex sharply forward if the diver’s forehead hits the bottom

---

28 See generally Keil, supra note 25.


30 See John H. Flavell et al., Development of the Appearance-Reality Distinction, 15 COGNITIVE PSYCHOL. 95 (1983). (The description in the next paragraph draws from this study.)
and does not slide. The weight of the diver’s body moving down and forward can then break the neck and crush cervical vertebrae. A steeper dive poses a risk of concussion, but less of a risk of traumatic spinal cord injury. I served as an expert witness in a case involving such an injury. After an aquatics expert had explained the potential risks of shallow racing dives, the jury members seemed to understand the mechanics of spinal cord injury in that context. After a brief recess, an opposing witness testified that it was common knowledge that shallow racing dives were dangerous, and so the quadriplegic victim knew, or should have known, the potential consequences of his (obviously foolhardy) behavior. From the jurors’ nods of agreement, it seemed to me that the hindsight bias was at work. Now that the jurors knew of the hazard, they felt that it was common knowledge and that everyone should know it. In cases like this one, expert testimony on the hazards of diving and the need for adequate warnings is not enough. Such testimony should be supplemented by a description and explanation of how the hindsight bias works and how it can lead to erroneous beliefs about what is common knowledge and what is not.

Keil nicely points out the hindsight bias in the context of how people decide whether or not something is or is not a legitimate area of expertise. The legal community should be aware of these cognitive and metacognitive biases, both in terms of what they themselves believe and how these biases can affect jurors’ beliefs and decisions. So, not only should expert witnesses be able to explain complex phenomena in their own fields of expertise to lay people, they should also be able to explain relevant cognitive and metacognitive phenomena as well.

In the end, what are we to make of Bill Klem’s declarative realism, that a pitch is nothing—neither a strike nor a ball—until the umpire calls it? There is, after all, the assumption of scientific realism, which, as Faigman persuasively argues, is a necessary assumption. Katskee adds another important element concerning one aspect of scientific realism in the context of judicial decision-making: the critical importance of decisions seeming right and justifiable in the public eye, even when such decisions create judicial “truth.”

31 Keil, supra note 26, at 1043-45.
32 Katskee, supra note 5, at 860-65.
Declarative realism, as exemplified by Bill Klem, is constrained by scientific realism—the call must appear justified by the ball’s perceived trajectory in or out of the strike zone. Similarly, judicial decisions, such as the one described by Katskee concerning the teaching of intelligent design, are instances of declarative realism. Nevertheless, like Klem’s calls, they must be justifiable by the evidence and arguments presented. More important, perhaps, they must be perceived as justifiable by the legal community and ultimately by the public. Yes, truth is hard to come by, but both the legal and scientific communities have evolved principles and procedures to minimize bias and integrate declarative truth with scientific and, I daresay, judicial realism.

33 See Katskee, supra note 5, at 873-76.
NOTES

Other People’s Money

DRAWING THE CONSTITUTIONAL LINE BETWEEN THE RIGHT TO COUNSEL AND CONSTRAINTS ON OBSTRUCTIVE FEE ADVANCEMENT IN THE WAKE OF UNITED STATES V. STEIN

I. INTRODUCTION

In 2003, the U.S. Department of Justice (“DOJ”) issued its federal prosecutors a written set of guidelines to assist them in their investigation and prosecution of white-collar corporate crime.\(^1\) Specifically, these guidelines, issued in a document commonly known as the Thompson Memorandum,\(^2\) addressed the question of whether and under what circumstances Assistant U.S. Attorneys (“AUSAs”) conducting investigations into white collar crimes committed by employees and executives should bring a formal charge against the company itself.\(^3\) The Thompson Memo listed nine separate factors for prosecutors to evaluate when making the decision of whether or not to seek an indictment.\(^4\)

In the AUSAs’ determination of whether to prosecute a company, the Thompson Memo stressed consideration of the


\(^{3}\) Thompson Memo, supra note 1, at 1-2.

\(^{4}\) Id. at 3; see also Paul J. McNulty, Deputy Attorney General, DOJ, Statement Before the Committee on the Judiciary United States Senate Concerning “The Thompson Memorandum’s Effect on the Right to Counsel in Corporate Investigations” (Sept. 12, 2006) [hereinafter McNulty Statement], available at http://judiciary.senate.gov/testimony.cfm?id=2054&wit_id=2742.
authenticity of a company’s “cooperation” with the government investigation.\(^5\) One measure of this cooperation was whether the company was advancing legal fees to its investigated employees.\(^6\) If the company possessed no legal obligation to advance legal fees to its implicated employees in connection with the investigation, then the AUSAs prosecuting the case, in accordance with the guidelines, were permitted to view the advancement of legal fees as a failure to cooperate with the government.\(^7\)

However, this practice was recently condemned in a pair of decisions issued by Judge Lewis Kaplan of the Southern District of New York in *United States v. Stein*.\(^8\) In the first decision, issued in June 2006 (“*Stein I*”), the DOJ’s practice of considering the advancement of legal fees as a failure to cooperate was challenged by former employees of accounting giant Klynveld, Peat, Marwick, Goerdeler\(^9\) (“KPMG”).\(^10\) The DOJ had begun an investigation of KPMG and its employees over the “development, marketing and implementation of abusive tax shelters.”\(^11\) The DOJ sought cooperation from the KPMG employees suspected of misconduct as well as from KPMG as a company.\(^12\) In doing so, the AUSAs leading the investigation inquired into KPMG’s legal obligation to advance fees to its employees.\(^13\) When KPMG could show a history of advancing legal fees, but could not show any clear legal obligation to do so, the prosecutors pressured KPMG into limiting such assistance to its employees to demonstrate its

\(^5\) Thompson Memo, *supra* note 1, at 6.

\(^6\) *Id.* at 7-8.

\(^7\) *Id.* However, as will be discussed *infra* in Part II, this consideration was limited by certain factors in an effort to eliminate the risk of abuse by prosecutors. See McNulty Statement, *supra* note 4.


\(^12\) *Stein I*, 435 F. Supp. 2d at 341.

\(^13\) *Id.*; Browning, *KPMG No Sway*, *supra* note 10.
cooperation with the government and possibly avoid an indictment against the company.\textsuperscript{14}

The KPMG defendants claimed, and the \textit{Stein I} court agreed, that the pressure placed on KPMG to disregard its “long-standing policy”\textsuperscript{15} of advancing legal fees to employees being investigated or indicted deprived the employees of their Sixth Amendment right to the assistance of counsel.\textsuperscript{16} Alternatively, the court held that this pressure violated the defendants’ substantive due process right to a fair criminal trial under the Fifth Amendment.\textsuperscript{17} The court noted that:

\begin{quote}
KPMG refused to pay because the government held the proverbial gun to its head. Had that pressure not been brought to bear, KPMG would have paid these defendants’ legal expenses. Those who commit crimes—regardless of whether they wear white or blue collars—must be brought to justice. The government, however, has let its zeal get in the way of its judgment. . . . Defendants had . . . an expectation that their expenses in defending any claims or charges brought against them by reason of their employment by KPMG would be paid by the firm. The law protects such interests against unjustified and improper interference.\textsuperscript{18}
\end{quote}

Because the court intended its substantive due process analysis to be used only as an alternative in the event that a reviewing court disagreed with its Sixth Amendment analysis,\textsuperscript{19} this Note will focus primarily on the protections the \textit{Stein} court recognized under the Sixth Amendment.

In its decision in \textit{Stein I}, the court first determined that KPMG would have advanced—and may even have been legally obligated to advance—legal fees to its employees.\textsuperscript{20} This determination was based on a state statute that gives companies the authority to indemnify their employees through means which include the advancement of legal fees, as well as on KPMG’s history of paying the legal expenses of its partners and employees inured as a result of their employment,

\begin{flushright}
\textsuperscript{14} \textit{Stein I}, 435 F. Supp. 2d at 342, 344. \\
\textsuperscript{15} \textit{Id.} at 352; \textit{Browning, KPMG No Sway, supra note 10}. \\
\textsuperscript{17} \textit{Stein I}, 435 F. Supp. 2d at 356, 360; \textit{see also Stein III, 495 F. Supp. 2d 390, 409 n.80 (S.D.N.Y. 2007)}. \\
\textsuperscript{18} \textit{Stein I}, 435 F. Supp. 2d at 336, 366-67. \\
\textsuperscript{19} \textit{Stein III, 495 F. Supp. 2d at 409 n.80}. \\
\textsuperscript{20} \textit{Stein I, 435 F. Supp. 2d at 353, 356; see also Stein III, 495 F. Supp. 2d at 394, 409}. \\
\end{flushright}
regardless of cost. Next, the court found that the inherent threat of an indictment based on the guidelines in the Thompson Memo and the pressure supplied by the AUSAs' reinforcement of that threat caused KPMG to depart from its usual policy of paying legal fees and expenses.

These holdings were confirmed in July 2007 when the court issued a second opinion ("Stein III") in which it dismissed the indictments of a number of the defendants and rejected arguments by the government challenging the correctness of the court's rulings in Stein I. The Stein III court reinforced its holdings in Stein I by pointing to additional facts that showed KPMG's intent and desire to cover its employees' legal fees. However, the Stein decisions are subject to a number of weaknesses that limit their effectiveness in preventing government prosecutors from infringing on the right to counsel while preserving the government's interest in prosecuting white collar crime.

First, in arriving at its conclusions, the court was forced to engage in a long, murky, and protracted analysis of the facts that may subject its conclusions to alternative interpretation by other courts facing similar claims in the future.

Second, the claim of prosecutorial misconduct presented by the

---

21 Stein I, 435 F. Supp. 2d at 355-56; Stein III, 495 F. Supp. 2d at 405-09.
22 Stein I, 435 F. Supp. 2d at 352.
23 The second decision is referred to as Stein III for purposes of consistency with the court, which referred to a July 2006 decision concerning suppression of statements made by the KPMG defendants as "Stein II." See infra note 198.
24 Stein III, 495 F. Supp. 2d at 393-94.
25 Id. at 407-09.
26 The Stein I court held that the defendants' right to "obtain and use . . . resources lawfully available to him or her" in preparing a defense "free of knowing or reckless government interference" is a fundamental constitutional right protected under notions of substantive due process. 435 F. Supp. 2d at 360-62. Any government infringement of this right is subject to a strict scrutiny standard of review—the government's actions must be "narrowly tailored to achieve a compelling government interest." Id. at 362. As the Sixth Amendment right to counsel has also repeatedly been declared "fundamental" by the Supreme Court, see Cuyler v. Sullivan, 446 U.S. 335, 343 (1980); Argersinger v. Hamlín, 407 U.S. 25, 32 (1972); Gideon v. Wainwright, 372 U.S. 335, 344 (1963), any government interest which would impede this right of a criminal defendant must also be compelling.
27 The difficulty of the analysis for the Stein court is demonstrated by its reliance in Stein III on additional pieces of evidence, at least one of which had not been discovered, when it confirmed its conclusions and holding in Stein I. See Stein III, 495 F. Supp. 2d at 407.
28 See Browning, Tactic Questioned, supra note 16 (reporting that while the district court ruling applies only to the KPMG case, it carries extra weight because of the large number of "high profile white-collar and corporate fraud cases" presented before the federal court in Manhattan).
defendants in the Stein cases still leaves courts unable to protect the right to the assistance of counsel for defendants until after that right has been violated. Finally, as in Stein, the need to hear such claims may force courts into the unwanted position of intruding on a prosecutor’s broad discretion to determine whether or not to seek an indictment. Since the decision to seek an indictment rests with the executive branch, judicial review of a prosecutor’s authority threatens to undermine the doctrine of separation of powers.

This Note will argue that while modern application of the Sixth Amendment right to counsel could cover a company’s agreement to advance attorney’s fees in certain instances, the Stein decisions are only minimally effective in ascertaining where such coverage applies. First, the Stein decisions do not clearly distinguish between conduct constituting impermissible government interference with the right to counsel and voluntary choices of companies denying advancement to employees based on decisions which serve the companies’ best interests. This requires an alternative solution to provide prosecutors with a clearer gauge of where government conduct exceeds the proper balance between prosecution of white-collar crime and individual constitutional rights. Second, such a solution must not excessively intrude on the government’s compelling interest in prosecuting white-collar crime and limiting obstructive conduct.

Part II of this Note will review the history of the DOJ’s implementation of the guidelines set forth in the Thompson Memo and recap the history of the guidelines as set forth five years before the DOJ brought action against the KPMG employees. A brief history of the KPMG investigation will be provided to re-establish the setting that brought these issues to

---

30 See infra notes 261-262 and accompanying text.
31 Much has been written regarding the intent behind the right to counsel in the years preceding and immediately following American independence. See, e.g., William M. Beaney, The Right to Counsel in American Courts 14-24 (1955); Elliott Evans Cheatham, A Lawyer When Needed 14, 49-50 (1963); William M. Beaney, The Effective Assistance of Counsel, in Arthur L. Harding et al., Fundamental Law in Criminal Prosecutions 39, 39-40 (1959). However, this Note will primarily focus on the last seventy years of Sixth Amendment jurisprudence dealing with the right to counsel.
32 See Advanced Mining Sys., Inc. v. Fricke, 623 A.2d 82, 84 (Del. Ch. 1992) (explaining that Delaware law leaves to the business judgment of the board “the task of determining whether . . . advancement of expenses would on balance be likely to promote [a] corporation’s interests”).
light, including specific incidences within the investigation that
gave rise to some of the claims of abuse and misconduct that
this Note’s proposal attempts to remedy. Part III will briefly
restate the scope, application, and limits of the Sixth Amend-
ment right to counsel to highlight the protections guaranteed
by the right. Part III will also discuss the district court’s
application of the right to the facts in the Stein cases.

Part IV will analyze the district court’s decisions in
Stein I and III. This analysis will highlight the weaknesses of
the court’s decisions and focus on balancing the need to grant
prosecutors the necessary discretion to determine whether to
bring charges while protecting individuals and companies from
potential abuse of that discretion. This section will include
analysis of state statutes addressing indemnification and
advancement of legal fees to officers and employees.

Part V will propose a bright-line alternative to aid
courts and prosecutors in determining when a defendant’s right
to the assets of a third party for the payment of legal fees is
protected under the Sixth Amendment and when (and under
what circumstances) government interference is appropriate.
The proposal will resurrect a portion of the AUSAs’ claim that
the Sixth Amendment right to counsel does not cover “other
people’s money” in the form of a legislative enactment
permitting prosecutors to use preliminary injunctive restraints
that are already in use with current federal forfeiture
provisions.

II. BACKGROUND

A. The Thompson Memorandum

Formally titled “Principles of Federal Prosecution of
Business Organizations,” the Thompson Memo was issued by
Deputy Attorney General Larry Thompson as a “revised set of principles to guide” Department Prosecutors as they make

33 While the memo explicitly states that the principles are intended as a
“guide” for prosecutors, some analysts, as well as the court in Stein have concluded
that the principles were used as hard rules in their assessments of corporate
cooperation and determinations to bring an indictment. See, e.g., Lynnley Browning,
Judges Press Companies That Cut Off Legal Fees, N.Y. Times, Apr. 17, 2006, at C1
[hereinafter Browning, Judges Press] (pointing out that while prosecutors claim that
the Thompson Memo was meant to serve as an internal guide for prosecutors, many
prosecutors use the guidelines “like a bible” when investigating a company); Stein I,
435 F. Supp. 2d at 338 (“[T]he Thompson Memorandum is binding on all federal
prosecutors. Thus, all United States Attorneys now are obliged to consider the
the decision of whether to seek charges against a business organization.”34 It was issued in the wake of a number of corporate scandals involving high-profile companies such as MCI WorldCom, HealthSouth, and Adelphia35 that followed shortly after the collapse of Enron.36 These scandals “cost investors billions of dollars and thousands of workers lost their jobs,”37 and they compelled the DOJ to take action against the corporate malfeasors.38

While guiding prosecutors in their consideration of seeking an indictment against a company was at least one purpose of the Thompson Memo,39 one of its primary purposes was to “increase[] emphasis on and scrutiny of the authenticity of a corporation’s cooperation” with the DOJ’s investigation.40 Prosecutors from the DOJ had always sought cooperation from companies during the investigation of potentially illegal conduct,41 but the content of the Thompson Memo lends itself advancing of legal fees by business entities . . . as at least possibly indicative of an attempt to protect culpable employees and as a factor weighing in favor of indictment of the entity.” (footnote omitted)).

34 Thompson Memo, supra note 1, cover page. The memo was actually a “modest revision” of a document issued in June 1999 by then-U.S. Deputy Attorney General Eric Holder entitled “Federal Prosecution of Corporations,” commonly referred to as the “Holder Memo.” Stein I, 435 F. Supp. 2d at 336; see also Wray & Hur, supra note 2, at 1099 (characterizing the Holder Memo as the first “uniform policy on corporate prosecution”).

35 Wray & Hur, supra note 2, at 1100.

36 Browning, Judges Press, supra note 33. Speaking before Congress, Deputy Attorney General Paul McNulty called the year 2002 “a time of great concern to . . . Congress and to American workers and investors” based on the reduction in the public’s trust in corporate America due to the “large-scale bankruptcies of companies like Enron.” McNulty Statement, supra note 4. “The guidance contained in the Thompson Memorandum . . . must be viewed in the context of these massive corporate scandals.” Id.

37 Browning, Judges Press, supra note 33.

38 As McNulty testified:

The American people and their representatives . . . in Congress demanded that those responsible for corporate malfeasance be brought to justice . . . . The Department of Justice responded to this crisis in corporate America with vigor and action . . . . Since 2002, the Department of Justice obtained more than 1000 corporate fraud convictions and convicted more than 160 corporate presidents and executive officers.

McNulty Statement, supra note 4.

39 The Thompson Memo also discussed the important public benefits that would flow from corporate prosecution, such as the likelihood that the company will take “immediate remedial steps,” specific deterrence in the form of a changed culture in the indicted corporation, and minimized risks of large-scale public harm, such as environmental crimes and financial fraud. Thompson Memo, supra note 1, cover page.

40 Id.; Wray & Hur, supra note 2, at 1097, 1135.

41 See McNulty Statement, supra note 4 (calling the Thompson Memo “a time-tested and fair summary of the factors a prosecutor considers in charging a
to the inference that the DOJ and Thompson had become skeptical of the cooperation they were receiving: “Too often business organizations, while purporting to cooperate with a Department investigation, in fact take steps to impede the quick and effective exposure of the complete scope of wrongdoing under investigation.”

The Thompson Memo, which acknowledged that only a minority of cases would result in a company itself being subjected to criminal charges, listed a number of factors for prosecutors to consider in determining whether to bring charges against an entity. Some of these factors were identical to those that prosecutors were already using to determine whether to bring charges against individuals, such as sufficiency of the evidence, success at trial, deterrence, rehabilitation, and alternative consequences of conviction. However, due to a company’s status as a “corporate person,” additional factors were given for consideration. These factors included, but were not limited to, the nature and seriousness of the offense, the extent of the wrongdoing within the company, the history of such conduct, voluntary disclosure of any wrongdoing by the company and willingness to cooperate, corporate compliance programs, remedial measures (including

corporate entity [that] commits to paper what good prosecutors have been doing for decades”).

42 Thompson Memo, supra note 1, cover page. This skepticism was even shared by attorneys working as in-house counsel for private companies under investigation as little as three months prior to the decision handed down in Stein I. See Nathan Koppel, U.S. Pressures Firms Not to Pay Staff Legal Fees, WALL ST. J., Mar. 28, 2006, at B1 (quoting Harvey Wolkoff, an in-house lawyer for Enterasys Networks, Inc., supporting government requests that Enterasys challenge Delaware law authorizing advancement of legal fees to the defendants: “If [the defendants] did something criminal, why should’ their legal fees get reimbursed?”).

43 Thompson Memo, supra note 1, cover page. This was in part due to the recognition that a company “can act only through individuals,” and that “imposition of individual criminal liability may provide the strongest deterrent against future corporate wrongdoing.” Id. at 1; see also Wray & Hur, supra note 2, at 1106-07.

44 Thompson Memo, supra note 1, at 3.

45 Prosecutors had already been instructed to consider these factors in seeking indictments against individual defendants, as the factors had already been listed in the United States Attorneys Manual. See U.S. DEPT OF JUSTICE, U.S. ATTORNEY’S MANUAL § 9-27.220 (2007).

46 Thompson Memo, supra note 1, at 3.

47 In evaluating cooperation, the prosecutors could also consider whether a company waived corporate attorney-client and work product protection. Id. at 3. One reason the DOJ began looking at waivers of privilege and work product is that in previous corporate criminal investigations, counsel for the companies would run all of the companies’ documentation through their in-house legal department so as to claim attorney-client or work product protection for documentation detailing routine business activity as well as accounting and financial records. McNulty Statement, supra note 4.
termination of responsible employees and making restitution), and the collateral effects of a potential indictment (including harm to shareholders, pension holders, and non-culpable employees).48

In elaborating on the consideration that should be given to a company’s level of cooperation, the Thompson Memo acknowledged the difficulties of conducting an investigation of a company, such as ascertaining who the culpable individuals are, their location, and the location of records.49 These difficulties made obtaining company cooperation so critical that the Memo referenced guidelines for prosecutors to consider in determining whether to offer the company immunity or amnesty in the form of a non-prosecution agreement.50 The Memo also stated that prosecutors should assess whether the company “appears to be protecting its culpable employees.”51 This protection referred to company conduct that included advancing attorney’s fees, retention of employees without sanction for any misconduct, and providing information to employees about the government’s investigation.52 As it pertained to legal fees, if prosecutors felt that such protection was being used to limit or prevent the flow of truthful communication from employees to the government, or to protect the culpable employees or the company, such provision of fees could be weighed in evaluating the adequacy of cooperation.53

48 Thompson Memo, supra note 1, at 3.
49 Id. at 6.

It will often be difficult to determine which individual took which action on behalf of the corporation. Lines of authority and responsibility may be shared among operating divisions . . . and records and personnel may be spread throughout the United States . . . . Where the criminal conduct continued over an extended period of time, the culpable or knowledgeable personnel may have been promoted, transferred, or fired, or they may have quit or retired. Accordingly, a corporation’s cooperation may be critical in identifying the culprits and locating relevant evidence.

Id.

50 Id.; Wray & Hur, supra note 2, at 1103-04.
51 Thompson Memo, supra note 1, at 7.
52 Id. at 7-8. The Memo also addressed other concerns regarding whether an investigated company is engaged in conduct that impedes the investigation, such as making overly broad assertions of corporate representation of employees, and issuing directions to its employees not to cooperate. Id.

53 McNulty Statement, supra note 4. The DOJ was primarily concerned about the abuse of fee advancement as a means of obstructing the government’s investigation in conjunction with other indications of non-cooperation, such as “overly broad assertions of corporate representation of its employees, a refusal to sanction wrong-doers, a failure to comply with document subpoenas and a failure to preserve documents.” Id. To the extent that these other indicators were not present, the
B. The Investigation of KPMG

The DOJ used the guidelines set out in the Thompson Memo when it began its investigation of KPMG. In October 2001, the IRS initiated an investigation into the creation of and failure to register a number of “abusive” tax shelters that KPMG had participated in forming beginning in 1997. Following the initiation of the investigation, as well as the issuance of a number of summonses for information on these abusive shelters, a Senate subcommittee launched another investigation “into the development, marketing and implementation” of these shelters. The subcommittee found that KPMG sold a number of these illegal shelters to at least 350 people in the four year span from 1997 to 2001. In addition, KPMG earned $124 million in fees, while depriving the Treasury of at least $1.4 billion in unpaid taxes. These investigations brought much negative attention to KPMG.

As concern grew regarding the fallout from the conclusions of the Senate subcommittee and IRS investigations, KPMG chair Eugene O’Kelly retained a private law firm in an effort to develop a “cooperative approach” with the government. This approach included the decision to ask some of KPMG’s senior partners to vacate their positions within the company, including deputy chair and Chief Operating Officer Jeffrey Stein, vice-chair of tax services Richard Smith, and a partner in personal financial planning, Jeffrey Eischid. However, in terminating their employments, KPMG negotiated...
severance agreements with at least two of these individuals, specifically Messrs. Stein and Smith.64

Mr. Stein, who held a senior position with the company, worked out an agreement with KPMG in which his departure would be “cushioned substantially.”65 Under this agreement, Mr. Stein would be retained as a consultant for three years with a monthly compensation of $100,000, would release all claims against the firm and its partners, and would be provided legal representation at the expense of the firm by counsel acceptable to both him and the firm in any suits brought against KPMG or its personnel and himself.66 KPMG would also continue to cover Mr. Stein under its Professional Indemnity Insurance Program against any claims arising from his role with the company.67 KPMG’s agreement with Mr. Smith contained an essentially identical clause, by which the firm agreed to pay the costs of Smith’s defense.68

In early 2004, the IRS concluded its investigation and made a criminal referral to the DOJ recommending prosecution of KPMG.69 The DOJ, in turn, referred the case to the United States Attorney’s Office (“USAO”) in February of that year.70 Upon learning of the criminal referral, but prior to any meeting with the USAO, KPMG issued a voicemail message to its partners stating that it would pay for “competent counsel” for any present or former members of the firm who were asked to appear before the USAO in relation to the investigation.71 The message made no mention of any conditions on the payment, nor did it mention any limits on the amount of legal expenses KPMG would pay.72

In the initial discussions between the USAO and the retained counsel for KPMG, the AUSAs immediately inquired whether KPMG was paying the legal fees of the investigated

64 Id.; Stein III, 495 F. Supp. 2d at 408.
65 Stein I, 435 F. Supp. 2d at 339.
66 Id. It was also agreed that in the event that only Mr. Stein was named as a party in any suit arising out of his actions with the company, the counsel need only be reasonably acceptable to him. Id.
67 Id. at 339 n.25; see also Lynnley Browning, Prosecutor Denies Pressure on KPMG to Cut Off Legal Fees, N.Y. TIMES, May 9, 2006, at C7 [hereinafter Browning, Prosecutor Denies Pressure] (reporting the value of the severance package to Mr. Stein at between $8 and $10 million).
68 Stein III, 495 F. Supp. 2d at 408.
69 Stein I, 435 F. Supp. 2d at 339.
70 Id. at 340.
71 Stein III, 495 F. Supp. 2d at 407.
72 Id.
employees and what obligations or agreements it had made to do so.\textsuperscript{73} KPMG’s counsel indicated that the company’s objective was not to protect its employees, but rather to save itself out of the fear that a formal indictment would be disastrous for the company, forcing the firm out of business.\textsuperscript{74} As such, while lawyers for KPMG told the AUSAs that paying legal fees for employees had been the firm’s “common practice,”\textsuperscript{75} the vagueness of the firm’s partnership agreement and the Delaware law governing the agreement gave KPMG the discretion to make its own determination.\textsuperscript{76} Counsel for KPMG also indicated that in spite of this “common practice,” it “still was checking on its legal obligations,” and would not pay legal fees “for employees

\textsuperscript{73} Stein I, 435 F. Supp. 2d at 346. Prior to the court’s decision in Stein I, the parties had stipulated that before February 2004, KPMG had a “longstanding voluntary practice” of advancing and paying legal fees without a preset cap or condition of cooperation with the government, for counsel for partners, principals, and employees of the firm in those situations where separate counsel was appropriate to represent the individual in any civil, criminal or regulatory proceeding involving activities arising within the scope of the individual’s duties and responsibilities as a KPMG partner, principal, or employee.

\textsuperscript{74} Id. at 340 (emphasis added, internal citations omitted).

\textsuperscript{75} Id. at 341. Stein I references the widely discussed and accepted conclusion that KPMG’s primary concern was making sure it did not suffer the same fate as its competitor, Arthur Andersen LLP, id., which “imploded shortly after its indictment in 2002 for allegedly obstructing the government’s investigation of fraud at Enron Corp.” Koppel, supra note 42; see also Browning, Judges Press, supra note 33 (calling a formal indictment “a virtual death knell for many companies, as it was for the accounting firm Arthur Andersen”); Bruce D. Fisher, Andersen v. U.S.: A Shift in the Legal Winds for Public Auditors? TENN. BAR J., Nov. 2005, at 22 (stating that “mere indictment—formal criminal accusation—proved to be the Andersen firm’s downfall” because following the indictment, “Andersen’s clients deserted it, and the firm eventually filed for bankruptcy and thousands lost their jobs and pensions”). The consequences of such negative publicity can be further seen by the fact that after Andersen was both indicted and convicted on trial, even the over-turning of the conviction by the United States Supreme Court “probably does not portend a significant change in the legal winds for either Andersen or for the thousands of former Andersen employees.” Id. at 32.

\textsuperscript{76} Stein I, 435 F. Supp. 2d at 342. KPMG claimed, and the DOJ did not dispute, that it could not recall any partner, principal, or employee who had been indicted for conduct arising within the scope of the duties of their position since two previous employees (both partners) were indicted in 1974. Id. at 340. Though the company could locate no documentation to support its claim that it paid pre- and post-indictment fees on behalf of those individuals, both sides stipulated that litigation expenses for those employees were covered by KPMG. Id.
who declined to cooperate with the government . . . as long as it had discretion to take that position.\footnote{Stein I, 435 F. Supp. 2d at 342 (emphasis added). One of the AUSAs present at the initial meeting made a statement in reference to “federal guidelines” that misconduct was not to be rewarded. Id. While intended by the AUSA as a reference to federal sentencing guidelines, it was understood by counsel for KPMG as a reference to the Thompson Memo, id. at 342 n.45, and “as a reminder that payment of legal fees by KPMG, beyond any that it might legally be obligated to pay, could well count against KPMG in the government’s decision whether to indict the firm,” id. at 344. The court also determined that another AUSA present at the meeting made a comment that discretion regarding payment of legal fees would be looked at “under a microscope,” even though the comment appeared only in the notes of one attorney for KPMG, and no witness who testified at the hearing recalled the statement being made. Id. at 344, 344 n.52.}

Sensing the government’s displeasure at the idea of severance packages for suspected individuals, KPMG chose not to sign the agreement it had negotiated with Richard Smith, which was still pending at the time of the initial meeting with the AUSAs.\footnote{Stein III, 495 F. Supp. 2d at 408-09.}

Shortly thereafter, counsel for KPMG reported to the USAO that the firm did not believe any “binding legal obligation to pay legal fees” existed, “but that ‘it would be a big problem’ not to do so” due to the firm's structure as a partnership.\footnote{Stein I, 435 F. Supp. 2d at 344, 345; see also Browning, Prosecutor Denies Pressure, supra note 67 (“Partnerships like KPMG typically pay an employees legal fees. Nonetheless, KPMG had a choice.”).}

KPMG announced in a form letter to its employees that it would advance the legal fees but would limit the amount it paid to up to $400,000 per individual and that payment of the fees would be conditioned on that particular employee being fully cooperative with the government and the firm.\footnote{Stein I, 435 F. Supp. 2d at 345; see also Mark Hamblett, Lawyers Spar over KPMG Legal Defense Fee Policy, N.Y. L.J., May 9, 2006, at 1.}

The form letter also stated that if any of its employees were charged with criminal wrongdoing, payment of legal fees would cease immediately.\footnote{Stein I, 435 F. Supp. 2d at 345-46; Hamblett, supra note 80, at 1.}

After viewing the letter, the USAO for the most part took no issue with the company’s announcement and proceeded in its investigation.\footnote{Stein I, 435 F. Supp. 2d at 346; see also Stein III, 495 F. Supp. 2d at 404 (referring to the government as “perfectly happy” to let KPMG advance the fees subject to those conditions).}

During the course of the investigation, when the AUSAs felt that the company personnel they were investigating were failing to adequately cooperate, the AUSAs would notify KPMG.\footnote{Stein I, 435 F. Supp. 2d at 347.}

Counsel for the company would then inform the attorneys for the individuals that legal fees would be
terminated absent indication from the government that the employees in question ceased in their refusal to participate in government interviews. In some cases, the individuals relented under the pressure from KPMG’s threats and participated in the interviews. For those who refused, KPMG terminated their employment and cut off payment of their fees.

Subsequent meetings between the government and counsel for KPMG revealed the severance packages that had been granted to certain executives, including the one provided for Mr. Stein. The Stein agreement became a particularly thorny issue for both sides for two reasons. First, the agreement had no real restrictions or conditions on the payment of legal fees, which was inconsistent with KPMG’s earlier statements to the government with regard to limitations it would impose on such payments to current employees under investigation. Second, the amount that KPMG spent on Mr. Stein’s defense—over $640,000 for the criminal investigation alone—clearly exceeded any amount KPMG had represented to the AUSAs that it would pay in legal fees for employees. Despite the cooperation KPMG had shown with respect to its remaining employees who were subject to investigation, the government’s discontent with such rich severance packages and the risk that it would be perceived as a failure to cooperate greatly concerned KPMG. In May 2005, the company severed the consulting agreement with Stein and cut off payment of his legal fees in an effort to display full cooperation with the government under the guidelines of the Thompson Memo.

\[84\] Id.
\[85\] Id.; see also Stein III, 495 F. Supp. 2d at 404 (stating that by allowing KPMG to advance legal fees to its employee-defendants in the pre-indictment investigation stage, but subject to conditions of full cooperation, the government obtained leverage over the defendants through KPMG by “holding[ing] over their head their job”).
\[86\] Stein I, 435 F. Supp. 2d at 347.
\[87\] Id.
\[88\] Id. at 347-48. The agreement KPMG negotiated with Mr. Smith did not raise an issue with the AUSAs because KPMG refused to effectuate it once it sensed the pressure from the government. See Stein III, 495 F. Supp. 2d at 409.
\[90\] Id. at 348, nn.74, 80.
\[91\] Id. at 347. The chief of the criminal division of the United States Attorney’s Office, relayed to KPMG counsel and executives that such severance packages are a “troubling issue under the ‘Thompson Memo,’” id. 344 n.51, which led to the firm viewing the severance agreements as a “ticking bomb,” id. at 347.
\[92\] Id. at 348; see also Browning, Tactic Questioned, supra note 16.
In late August 2005, the first nine indictments of individual KPMG employees, including that of Jeffrey Stein, were handed down by a federal grand jury.\textsuperscript{93} As per the terms of advancement that it had disclosed to the USAO, the company ceased payments of legal fees for the indicted defendants.\textsuperscript{94} Around that same period, KPMG and the government entered into a Deferred Prosecution Agreement ("DPA").\textsuperscript{95} Among the terms of the DPA were that KPMG would pay $456 million in penalties, would forego the indictment process and be charged with one count of conspiracy to defraud the government, and would admit to criminal wrongdoing.\textsuperscript{96} In exchange, the government would not prosecute the company, contingent on KPMG’s continued cooperation with the government’s investigation in accordance with the requirements of the DPA and the acceptance of certain restrictions on its tax practice.\textsuperscript{97} Shortly thereafter, in January 2006, the KPMG defendants challenged the actions of the AUSAs authorized by the Thompson Memo, claiming, among other things, that the government interfered with their Sixth Amendment right to

\textsuperscript{93} Jonathan Weil, Nine Are Charged in KPMG Case on Tax Shelters, WALL ST. J., Aug. 30, 2005, at C1. The indictments also named former senior tax chiefs Richard Smith and John Lanning, as well as Raymond Ruble, a former partner at the New York office of the law firm Sidley Austin Brown & Wood LLP, for his part in writing legal opinions supporting the tax shelters. Id. The indictments also disclosed allegations that "at least 14 KPMG partners used some of the shelters in question to shave their own tax bills." Id. In all, sixteen indictments were handed down by May 2006, a little over a month before the Stein I decision. See Browning, Prosecutor Denies Pressure, supra note 67 (referring to the “tax shelter trial of 16 former employees of the accounting firm KPMG”).

\textsuperscript{94} Stein I, 435 F. Supp. 2d at 350.

\textsuperscript{95} Weil, supra note 93.

\textsuperscript{96} Id. The DPA between KPMG and the government has been reported as a victory for the company because it “gives the firm a chance to avoid the kind of criminal case that proved fatal for . . . Arthur Andersen.” Jonathan D. Glater, KPMG’s Gain, Partners’ Loss, N.Y. TIMES, Aug. 29, 2005, at C1; see also supra note 74 and accompanying text.

\textsuperscript{97} Weil, supra note 93; see also Jonathan D. Glater, U.S. to Widen Inquiry of KPMG Tax Shelters, N.Y. TIMES, Sept. 7, 2005, at C1. The terms of this continued cooperation required KPMG to promptly provide “all documents, records, information and [any] other evidence” that the USAO, the IRS, or any other government agency designated by the USAO would need for its continued investigation. Stein I, 435 F. Supp. 2d at 349. In addition, KPMG agreed not to assert any claim of privilege “including, but not limited to the attorney-client privilege and the work product protection” in regards to any of the documents, records and other information requested by the government. Id. at 349-50. Finally, under the DPA, KPMG agreed that even after the dismissal of the Information, which contained the one charge against the company in place of a formal indictment, KPMG would “continue to fulfill the cooperation obligations set forth” in relation to any investigation, prosecution or proceeding (criminal or civil) that arose out of the conduct being investigated. Id. at 350.
the assistance of counsel by hindering KPMG’s advancement of attorneys’ fees.98 In addition, the defendants requested that the charges against them be dismissed.99 The District Court heard the defendants’ challenge and ruled in their favor, holding that the conduct of the AUSAs under the Thompson Memo did in fact infringe on the defendants’ Sixth Amendment right.100

III. OVERVIEW OF THE SIXTH AMENDMENT RIGHT TO COUNSEL AND ITS APPLICATION IN STEIN

The Sixth Amendment of the United States Constitution reads: “In all criminal prosecutions, the accused shall enjoy the right to . . . have the assistance of counsel for his defense.”101 While the assistance of counsel as a right in the United States traces its roots at least to the earliest days of the American Constitution,102 most of the modern interpretation and application of the right to counsel has only taken place in the last seventy-five years.103 It is as a result of this modern interpretation that the court in Stein expressed the need to protect this right of the KPMG defendants.

A. The Right to Counsel and the Development of the Modern Application

The right to counsel guaranteed by the Sixth Amendment represents a deviation from the standard practice at the time of the American Revolution.104 “Under English law, an accused had a right to have counsel in misdemeanor, but not

---

98 Stein I, 435 F. Supp. 2d at 350.
99 Id.
100 Browning, Tactic Questioned, supra note 16. The government’s activity “interfered with the ability of the KPMG Defendants to obtain resources they otherwise would have had . . . [which] almost certainly will affect what these defendants can afford to permit their counsel to do.” Stein I, 435 F. Supp. 2d at 362. This in turn infringed on “the rights of such employees to a fair trial and to the effective assistance of counsel and therefore violated the Fifth and Sixth Amendments to the Constitution.” Id. at 382.
101 U.S. CONST. amend. VI.
102 BeaneY, supra note 31, at 14-15 (stating that there is an ongoing debate over the extent of the influence of English common law on the right to counsel as it pertains to the pre- and post-Revolutionary period).
103 Cheatham, supra note 31, at 8-9 (listing Powell v. Alabama, 287 U.S. 45 (1932), and Johnson v. Zerbst, 304 U.S. 458 (1938), as “the two leading cases that established the legal right to counsel” under the U.S. Constitution).
felony, cases.” Following the American Revolution, and upon adoption of the Constitution, only one of the original thirteen states continued to follow this practice. The remaining twelve states “fully recognized the right to counsel in criminal prosecutions.” However, the extent to which the individual states observed the right to counsel in state prosecutions was determined exclusively by state law and therefore could vary greatly between states. The right to counsel guaranteed under the Sixth Amendment also did not immediately apply to the states, but rather applied only in federal prosecutions.

The modern interpretation of the Sixth Amendment right to counsel began in 1932, with the Supreme Court’s decision in Powell v. Alabama. The Powell Court determined that the state of Alabama’s failure to ensure proper counsel for defendants in a murder prosecution deprived the defendants of their Due Process rights under the Fourteenth Amendment. The Court also outlined the essential protections that the assistance of counsel provides defendants:

[N]otice and hearing are preliminary steps essential to the passing of an enforceable judgment, and . . . constitute basic elements of the constitutional requirement of due process of law . . . . Historically . . ., in our country at least, [a hearing] has always included the right to the aid of counsel when desired and provided by the party asserting the right. The right to be heard would be . . . of little avail if it did not comprehend the right to be heard by counsel.

105 Id. While the English Parliament “granted special treatment to those accused under the Treason Act of 1695, and required the court to appoint counsel upon the request of the accused,” defendants were not permitted to have counsel in ordinary felony cases in England until 1836. Id. (citing Faretta v. Cal., 422 U.S. 806, 821-26 (1975)).

106 Id.

107 Id. (citing Powell v. Alabama, 287 U.S. 45, 60-64 (1932), where Justice Sutherland noted that the new states embraced the right to counsel in their state constitutions although the extent of such acceptance did vary from one state to the next).

108 Id. at 840.

109 Id.


111 Powell, 287 U.S. at 69.

112 Id. However the Court’s decision should not be construed as incorporating Sixth Amendment protections to state criminal prosecutions through the Due Process clause of the Fourteenth Amendment. This was clearly stated by the Supreme Court only ten years after Powell in Betts v. Brady. See 316 U.S. 455, 461-62 (1942).

113 Powell, 287 U.S. at 69.
Notably, the Court detailed the immense complexities of the criminal process that make the protection of the right to counsel so vital for criminal defendants:

Even the intelligent and educated layman has small and sometimes no skill in the science of law. If charged with a crime, he is incapable . . . of determining for himself whether the indictment is good or bad. He is unfamiliar with the rules of evidence. Left without the aid of counsel, he may be put on trial without a proper charge, and convicted upon incompetent evidence . . . . He lacks both the skill and knowledge adequately to prepare his defense, even though he have a perfect one. He requires . . . counsel at every step in the proceedings against him. Without it, though he be not guilty, he faces the danger of conviction because he does not know how to establish his innocence. If that be true of men of intelligence, how much more true is it of the ignorant and illiterate, or those of feeble intellect.114

Since its decision in Powell, the Supreme Court has repeatedly asserted that the Sixth Amendment right to counsel is designed to guarantee a criminal defendant a fair trial in the adversarial criminal process.115 “The Constitution guarantees a fair trial through the Due Process Clauses, but it defines the basic elements of a fair trial largely through the several provisions of the Sixth Amendment, including the Counsel Clause.”116 In its cases following Powell, the Court has further fleshed out the nature of the rights encompassed by the right to counsel. Most notably, in Gideon v. Wainwright, the Court ruled that the protections of the right to counsel are a “fundamental [safeguard] of liberty immune from federal abridgment [and are] equally protected against state invasion by the Due Process Clause of the Fourteenth Amendment.”117

114 Id. at 68-69. Since Powell, courts have gone to great lengths to ensure that proper representation has been provided, even going so far as to force the defendant to accept counsel he wished to refuse when the defendant’s ability to represent himself was in question. See Faretta v. California, 422 U.S. 806 (1975) (holding that a State may not constitutionally force a lawyer upon a criminal defendant who “voluntarily and intelligently” chooses to proceed without counsel).


116 Strickland, 466 U.S. at 684-85.

117 Gideon, 372 U.S. at 341, 342. This decision also expressly overturned the Court’s holding in Betts v. Brady that the Sixth Amendment did not apply to states through the incorporation of Fourteenth Amendment Due Process. Id. at 345. See supra note 112 and accompanying text.
This decision made the appointment of counsel for defendants unable to obtain counsel on their own an affirmative obligation of both federal and state courts.\textsuperscript{118}

Since its monumental holding delivered in \emph{Gideon}, the Supreme Court has continued to expound on the extent to which the right to counsel applies. The Court’s assessments have addressed the right as it pertains to, among other things: the different stages of court proceedings at which the right to counsel attaches,\textsuperscript{119} the different types and severity of offenses,\textsuperscript{120} issues of self-representation,\textsuperscript{121} the requirement of effective counsel,\textsuperscript{122} and defendants’ right to spend their own money to obtain their counsel of choice.\textsuperscript{123}

\textbf{B. The Sixth Amendment Protections at Issue in \emph{Stein}}

While the court in \emph{Stein} applied this modern interpretation to a number of questions presented before it, for the purposes of this Note, only two of these applications—the requirement of effective counsel and the defendants’ right to spend their own money on counsel of their choice—are of significant importance.\textsuperscript{124}

\begin{itemize}
\item \textsuperscript{118} \emph{Gideon}, 372 U.S. at 340, 344, 345.
\item \textsuperscript{119} United States v. Wade, 388 U.S. 218 (1967) (holding that certain pre-trial proceedings such as suspect line-ups for witness identification are “critical confrontations of the accused by the prosecution,” and that the presence of counsel is “necessary to ensure a meaningful defense.”).
\item \textsuperscript{120} Argersinger v. Hamlin, 407 U.S. 25, 33 (1972). This case was critical in its interpretation of the right to counsel for two reasons. First, it held that coverage applies regardless of a crime’s classification as a felony or misdemeanor and regardless of the severity of the punishment for the crime (i.e., fine or imprisonment). \textit{Id} at 33. Second, the Court took special notice of the difficulties that can arise when defendants attempt to cooperate with government prosecutors:
\begin{quote}
Beyond the problem of trials and appeals is that of the guilty plea, a problem which looms large in misdemeanor as well as felony cases. Counsel is needed so that the accused may know precisely what he is doing, so that he is fully aware of the prospect of going to jail or prison, and so that he is treated fairly by the prosecution.
\end{quote}
\textit{Id.}
\item \textsuperscript{121} Faretta v. California, 422 U.S. at 807 (holding that even though defendants have a right to counsel, and despite the Sixth Amendment’s purpose to ensure a fair trial, the State may not force a defendant to accept the assistance of counsel when that defendant insists on conducting his own defense); Adams v. United States \textit{ex rel.} McCann, 317 U.S. 269 (1942) (holding that the Constitution does not force a lawyer on the individual).
\item \textsuperscript{122} McMann v. Richardson, 397 U.S. 759, 771 (1970).
\item \textsuperscript{123} See Morris v. Slappy, 461 U.S. 1, 10 (1983).
\item \textsuperscript{124} The court in \emph{Stein I} acknowledged that “[t]he Sixth Amendment attaches only upon indictment.” \emph{Stein I}, 435 F. Supp. 2d 330, 373 (S.D.N.Y 2006). However, as
1. The Right to Effective Assistance

The first application of the Sixth Amendment in *Stein I* was the court's suggestion that the KPMG defendants were deprived of the right to the effective assistance of counsel. In *McMann v. Richardson*, the Supreme Court stated unequivocally that the right to counsel means the entitlement to “the effective assistance of competent counsel.” While this entitlement was put forth by the Court as early as its decision in *Powell*, it is the Court’s decision in *McMann* which clarified that the Sixth Amendment contained an implicit right to effective counsel. “If the right to counsel . . . is to serve its purpose, defendants cannot be left to the mercies of incompetent counsel.”

There has been extensive debate both within and outside of the courts regarding what constitutes effective versus ineffective assistance of counsel. The Supreme Court has addressed claims of ineffective assistance pertaining to both the general incompetence of the attorney, as well as government interference with the defendant’s efforts to mount a defense. While the *Stein* court stressed that the defendants were deprived of their rights to counsel “irrespective of the

the court also found that the government’s pre-indictment actions were likely to have “an unconstitutional effect upon indictment,” this warranted a finding of attachment of Sixth Amendment rights. *Id.* at 366. Since the court’s decision regarding the attachment of the right has no bearing on the purpose of this Note, it does not warrant further discussion.

125 *Stein I*, 435 F. Supp. 2d at 382.
126 397 U.S. at 771.
127 “[S]uch designation of counsel as was attempted was either so indefinite or so close upon the trial as to amount to a denial of effective and substantial aid in that regard.” *Powell* v. Alabama, 287 U.S. at 45, 53 (1932). Having found that the counsel for the defendants had made no investigation because no opportunity to do so had been given, the Court held that the “defendants were not accorded the right of counsel in any substantial sense.” *Id.* at 58. “[T]he necessity of counsel was so vital and imperative that the failure of the trial court to make an effective appointment of counsel was likewise a denial of due process . . . .” *Id.* at 71.
128 JOHN M. BURKOFF & NANCY M. BURKOFF, *INEFFECTIVE ASSISTANCE OF COUNSEL*, § 1.1 (summarizing the history of the right to effective assistance of counsel).
132 *Strickland*, 466 U.S. at 686.
quality of representation they receive” due to the effects of the government’s interference, it also hinted at the potential risk of incompetent counsel.133

a. Incompetent Counsel

The influence of the Thompson Memo and the conduct of the AUSAs created a risk of rendering private counsel for the KPMG defendants incompetent. In Strickland v. Washington, the Supreme Court developed a two-pronged test to evaluate a convicted defendant’s claim of defective assistance.134 First, there must be a showing of deficient representation by proving that counsel made errors so grave that he or she was not functioning as counsel within the scope of the Sixth Amendment right.135 Second, the defendant must show that this deficiency prejudiced his or her defense, depriving that defendant of a fair trial.136 The quality of representation by the defendant’s counsel must fall “below an objective standard of reasonableness.”137 Among the factors to consider in determining whether either of these prongs has been met are the attorney’s failure to maintain a duty of loyalty to the client by avoiding conflicts of interest, failure to advocate the defendant’s cause, failure to consult with the defendant on important decisions, and failure to keep the defendant informed of important developments in the course of the prosecution.138 The inquiry looks at the totality of the circumstances to determine the reasonableness of counsel’s performance.139

The Supreme Court specifically addressed conflicts of interest in Cuyler v. Sullivan.140 In Cuyler, the Court ruled that

---

134 Strickland, 466 U.S. at 687.
135 Id.; see also infra notes 153-155 and accompanying text.
136 Strickland, 466 U.S. at 687.
137 Id. at 688; see also Burkoff & Burkoff, supra note 128, at § 1:2 n.4 (citing critics of Strickland who believe that the test is geared more towards judicial efficiency in adjudicating such claims, as opposed to removal of injustices caused by incompetent trial counsel).
138 Strickland, 466 U.S. at 688.
139 Id.
140 446 U.S. 335 (1980); see also Burkoff & Burkoff, supra note 128, ch. 3 (providing a ranging overview of different factors to be considered by courts hearing claims of ineffective counsel due to attorney conflicts of interest, including claims raised before and after trial, obligations of trial courts to identify and inquire into potential conflicts, and harmonization with ethical standards for attorneys).
a defendant is denied the right to effective assistance when
counsel for the defendant has a conflict of interest which
prejudices the defense.141 The Court also ruled that a defendant
is entitled to the same measure of effectiveness when he or she
employs retained counsel as when counsel is appointed by the
court: “The vital guarantee of the Sixth Amendment would
stand for little if the often uninformed decision to retain a
particular lawyer could reduce or forfeit the defendant’s
entitlement to constitutional protection.”142

In *Stein I*, though ineffectiveness based on a prejudicial
conflict of interest of retained counsel was not explicitly found,
the possibility was certainly raised.143 The court analyzed the
pressure placed by the AUSAs on KPMG to limit its assistance
in the form of advancing legal fees to its employees:

[The government [did not] question the obvious conflict of interest
manifest in [counsel for KPMG]’s offer to recommend as counsel to
targeted KPMG employees “law firms that were familiar with these
types of proceedings and who understood that cooperation with the
government was the best way for KPMG to proceed.”144

The conflict of interest in *Stein* stemmed from KPMG’s
conditioning and limiting of fee payment to counsel for its
targeted employees. This created a risk that attorneys for the
KPMG employees might not provide full assistance based on
the perceived need to avoid risking a criminal indictment
against their clients resulting in the termination of payment of
their fees and the need to cooperate with the government in
order not to risk an indictment of the company.145 However, as

---

141 *Cuyler*, 446 U.S. at 350; see also *Mickens v. Taylor*, 535 U.S. 162, 167-68
(2002) (distinguishing between conflict of interest cases warranting an automatic
reversal of any conviction when defense counsel objects to representing divergent
interests, and cases requiring defendants to establish prejudice to the defense where
the conflict is not objected to, and the trial court does not know of, nor reasonably
should know of the conflict, prompting an inquiry).

142 *Cuyler*, 446 U.S. at 344. The Court also held that, despite the private
relationship between a criminal defendant and his retained counsel, since a state
criminal proceeding is an action of the state, the mere obtaining of a conviction in a
trial where the defendant is inadequately represented constitutes the necessary state
action to give rise to a due process violation. *Id.* at 343; see also *Burkoff & Burkoff*,
supra note 128, at § 1:8.


144 *Id.* at 345 n.54; see also *Cuyler*, 446 U.S. at 347 (requiring trial courts to
initiate inquiries into conflicts of interest where the court “knows or reasonably should
know that a particular conflict exists”).

145 *Stein I*, 435 F. Supp. 2d at 344-46; see also *Berube*, supra note 110, at 1395
(discussing similar conflicts of interest arising in criminal cases where defendants’
assets, including attorneys’ fees, may be forfeited prior to trial).
any risk of ineffective counsel was created by the government’s actions and the influence of the Thompson Memo, the Stein court did not focus on incompetence of counsel, but rather analyzed the case in terms of government interference. In addition, for the reasons discussed below, the Stein I court did not require a showing of prejudice to the defense. Regardless, it is important to highlight how a conflict of interest risks rendering defense counsel incompetent as a potential byproduct of excessive government scrutiny over how white collar defendants retain and pay for their representation.

b. Claims of Ineffective Assistance Based on Government Interference

Government interference with the ability of counsel to make “independent decisions about how to conduct the defense” also constitutes a violation of a criminal defendant’s right to the effective assistance of counsel.146

More specifically, the right to the assistance of counsel has been understood to mean that there can be no restrictions upon the function of counsel in defending a criminal prosecution in accord with the traditions of the adversary fact-finding process that has been constitutionalized in the Sixth and Fourteenth Amendments.147

Examples of government interference that have been found to render counsel’s assistance ineffective include state statutes permitting judges the discretion to bar closing summations in non-jury trials,148 court orders denying defendants the right to speak with their counsel,149 state rules requiring a defendant to testify first or not at all,150 and state rules declaring defendants unfit to testify under oath at trial on their own behalf.151

148 Id. at 864 (discussing the different arguments that the defendant’s attorney might have made during closing arguments, potentially altering the ultimate judgment rendered in the case).
149 Geders v. United States, 425 U.S. 80, 88 (1975) (discussing the importance of allowing defense counsel to speak with the defendant during an overnight recess in a criminal trial due to “tactical decisions to be made and strategies to be reviewed”).
150 Brooks v. Tennessee, 406 U.S. 605, 612-13 (1972) (“By requiring the accused and his lawyer to make [the choice to testify first or not at all] without an opportunity to evaluate the actual worth of their evidence, the statute restricts the defense—particularly counsel—in the planning of the case.”).
151 Ferguson v. Georgia, 365 U.S. 570, 592-94, 596 (1961) (holding that allowing a defendant to make only an unsworn statement during trial and preventing defense counsel from conducting a direct examination of the defendant denied the defendant of his rights to counsel and due process).
In Stein I, the court found a violation of the defendants’ right to counsel based on the Thompson Memo and the AUSAs’ scrutiny of KPMG’s fee advancement policy. The court held that the government obstructed the employees’ access to a valid source of funding for their legal defense and therefore improperly intruded on the manner in which the individual employees wished to defend themselves.\textsuperscript{152} In arriving at its holding, the court took notice of the immense amount of time, document review, and complexity involved in the preparation of what the court recognized was at least one of the “largest tax fraud case[s] in United States history.”\textsuperscript{153} This also warranted consideration of the vast amount of money that would be needed in order for the defendants to mount the defense they desired.\textsuperscript{154} The court in Stein I and again in Stein III highlighted the potential risks ensuing from such an obstruction, pointing out that “[a]t least most of [the defendants] likely will be unable to afford to pay their attorneys to review all or even most of the documents the government has produced or . . . to interview even a fraction of the witnesses the government has interviewed.”\textsuperscript{155} The court in

\textsuperscript{152} Stein I, 435 F. Supp. 2d at 362. “The government here acted with the purpose of minimizing these defendants’ access to resources necessary to mount their defenses or, at least, in reckless disregard that this would be the likely result of its actions. In these circumstances, it is not unfair to hold it accountable.” Id. at 366-67.

\textsuperscript{153} Id. at 362; see also Lynneley Browning, Judge Delays KPMG Tax Trial Over Legal Fees Dispute, N.Y. TIMES, Nov. 15, 2006, at C3. In its decision in Stein III, the court again raised this issue, pointing out the significant increases in the amount of documentation produced since its decision in Stein I, obstacles arising in accessing the documents, numerous motions raised over the course of the investigation, and the expected length of the trial. Stein III, 495 F. Supp. 2d 390, 417-18 (S.D.N.Y. 2007).

\textsuperscript{154} Stein I, 435 F. Supp. 2d at 362 n.163, 371; Stein III, 495 F. Supp. 2d at 423-24.

\textsuperscript{155} Stein I, 435 F. Supp. 2d at 371; see also Stein III, 495 F. Supp. 2d at 423 (stating that the relatively small net worth of some of the defendants in the case, compared with the large amounts already owed to their attorneys, illustrates that “[n]one of them can afford to defend this case at any meaningful level”). These considerations seem to indicate concern on the part of the Stein court that the defendants might not be effectively represented within the minimum requirements of the Sixth Amendment, discussed supra Part III.B.1.a, due to the need for counsel to make “strategic choices . . . after thorough investigation of law and facts.” Strickland, 466 U.S. at 690. However, two determinations by the court in Stein I and III remove this point from consideration. First, the Stein I court found that the defendants were deprived of their rights to counsel “irrespective of the quality of representation they receive[d].” Stein I, 435 F. Supp. 2d at 369. Second, the Stein III court recognized that appointed counsel, which would statutorily receive dramatically lower fees and would likely be more restricted in its ability to investigate, could still potentially “provide the minimally effective defense” required by the Constitution. Stein III, 495 F. Supp. 2d at 421. Therefore, this particular Sixth Amendment concern does not warrant speculation in this Note.
Stein I also addressed the costs of tax experts that would likely be needed to rebut expert testimony presented by the government.\textsuperscript{156}

However, the KPMG defendants were not required to prove that their defense had been prejudiced, as is customary for a finding of ineffective assistance of counsel.\textsuperscript{157} The Stein I court ruled that the government’s conduct that interfered with the defendants’ right “to be represented as they choose” constituted a complete deprivation of their right to counsel without it hinging on the quality of the representation they received.\textsuperscript{158} Thus, before any criminal trial could even begin, the court had to determine whether corrective action could allow the defendants to defend themselves as they had wished, guaranteeing them their right to a fair trial.\textsuperscript{159}

Additionally, prejudice is not required where the governmental interference is severe enough that it creates an overarching structural defect in a defendant’s trial.\textsuperscript{160} Such a defect could prohibit even fully competent counsel from providing effective assistance,\textsuperscript{161} and would warrant a presumption of prejudice against the defendant.\textsuperscript{162} In Stein I, the court found that such a defect existed in violation of the defendants’ right to counsel as a direct result of the Thompson Memo and the actions of the AUSAs.\textsuperscript{163} The court expressly held that the government’s obstruction of the defendants’ access to funds created a very high risk of contaminating the entire proceeding and that due to the immense amount of documentation involved, the substantial time and cost expended, and the complexity of the case, it would be impossible to know whether the defendants could have altered the outcome absent the limitations imposed by the government.\textsuperscript{164} Thus, a presumption of prejudice was warranted without need to review specific

\textsuperscript{156} Stein I, 435 F. Supp. 2d at 371.
\textsuperscript{157} Id. at 369.
\textsuperscript{158} Id.
\textsuperscript{159} Id.
\textsuperscript{160} See United States v. Gonzalez-Lopez, 126 S. Ct. 2557, 2564-65 (2006) (citing Arizona v. Fulminante, 499 U.S. 279, 306-10 (1991)) (dividing constitutional errors into “trial errors,” which occur during presentation of the case to the jury and are subject to harmless-error review, and “structural defects,” which affect the framework of the entire trial).
\textsuperscript{161} Gonzalez-Lopez, 126 S. Ct. at 2563-64.
\textsuperscript{163} Stein I, 435 F. Supp. 2d at 371.
\textsuperscript{164} Id. at 371-72.
details of any ensuing trial.165 This entitled the defendant employees to relief to the extent that “the Thompson Memorandum and the activities of the USAO . . . interfered with the rights of such employees to a fair trial and to the effective assistance of counsel.”166

2. The Defendants’ Right to Spend Their Own Money on Counsel of Their Choice

The second application of the right to counsel in the Stein cases is the district court’s holding in Stein III that the defendants were improperly deprived of their right to counsel of their choice.167 The right of defendants to obtain the counsel of their choice is another application of the Sixth Amendment right to counsel that has been recognized since Powell.168 However, the extent to which a defendant may choose his or her own counsel has been circumscribed by the restriction that the defendant be able to afford that counsel.169 This limitation was made clear by the Supreme Court in Morris v. Slappy.170

In Morris, the Court was presented with an indigent defendant who was represented by an appointed attorney, which the trial court had replaced due to illness.171 The defendant argued for a continuance, over the objections of his

165 However, the Supreme Court has also held that the mere need to review large amounts of documentation, complexity of the case, and time constraints in preparing an adequate defense do not give rise to an automatic finding of ineffective representation by counsel. See United States v. Cronic, 466 U.S. 648, 652-53 (1984) (holding that representation of a criminal defendant by an appointed attorney who had only twenty-five days to prepare a defense in a fraud case involving review of thousands of documents, and where the government had over four and a half years to investigate and prepare, does not create an automatic presumption of ineffectiveness). This would still require the accused to show specific errors made by his counsel that “undermined the reliability of the finding of guilt,” Id. at 659 n.26, without which reversal would be required even when counsel’s actual representation was flawless. Id. at 653. However such factors are relevant considerations in determining whether counsel made errors prejudicial to the defendant in his particular case and the extent to which those errors and prejudice rendered counsel ineffective. Id. at 663; see also United States v. Cronic, 839 F.2d 1401, 1402-04 (10th Cir. 1988) (finding, on remand from the Supreme Court, that errors made by counsel during trial due to inadequate preparation time and counsel’s inexperience caused prejudicial error to the defendant warranting a finding of ineffectiveness).

166 Stein I, 435 F. Supp. 2d at 382.
168 Powell v. Alabama, 287 U.S. 45, 52 (1932) (“It is hardly necessary to say that the right to counsel being conceded, a defendant should be afforded a fair opportunity to secure counsel of his own choice.”).
170 Id.
171 Id. at 5.
second attorney, claiming that his new counsel did not have enough time to prepare an adequate defense.\textsuperscript{172} The trial court denied the defendant’s motion, and the defendant was subsequently convicted.\textsuperscript{173} On appeal, the Circuit Court found that defendant’s Sixth Amendment right was violated due to the absence of a “meaningful attorney-client relationship” with his appointed counsel.\textsuperscript{174} The Supreme Court rejected the ruling by the Circuit Court, holding that where a defendant is unable to afford counsel of his choice, the Sixth Amendment does not require the defendant have a meaningful relationship with his appointed counsel.\textsuperscript{175}

Even when a defendant has sufficient assets to retain counsel of his choosing, further limitations on the guarantee may apply. In \textit{Wheat v. United States},\textsuperscript{176} the Supreme Court evaluated a potential conflict of interest when the defendant’s desired counsel was disqualified by the trial court due to already representing other defendants charged in the same conspiracy.\textsuperscript{177} The attorney had already been involved in substantial contact with the prosecution, and the government was concerned that if one of the other defendants were to testify against the petitioner, the attorney would fail in his responsibility to provide effective counsel.\textsuperscript{178} The Court ruled that the trial court did not err in its disqualification of the defendant’s counsel of choice and the defendant’s conviction was upheld.\textsuperscript{179} The court made it clear that

\begin{quote}
the Sixth Amendment right to choose one’s own counsel is circumscribed in several important respects. . . . [A] defendant may not insist on representation by an attorney he cannot afford, or who for other reasons declines to represent the defendant. Nor may a defendant insist on the counsel of an attorney who has a previous or
\end{quote}

\textsuperscript{172} \textit{Id.} at 6.
\textsuperscript{173} \textit{Id.} at 6-7, 9.
\textsuperscript{174} \textit{Id.} at 10.
\textsuperscript{175} \textit{Id.} at 13 (“No court could possibly guarantee that a defendant will develop the kind of rapport with his attorney—privately retained or provided by the public—that the Court of Appeals thought part of the Sixth Amendment guarantee of counsel.”).
\textsuperscript{176} \textit{486 U.S.} 153 (1988).
\textsuperscript{177} \textit{Id.} at 157.
\textsuperscript{178} \textit{Id.} at 155.
\textsuperscript{179} \textit{Id.} The Court also weighed the risk of government abuse in manufacturing conflicts to prevent defendants from being represented by “able defense counsel,” but chose to rely on trial courts being aware of such a tactic. \textit{Id.} at 163.
ongoing relationship with an opposing party, even when the opposing party is the Government.180

Where a defendant is wrongfully denied his counsel of choice, such a deprivation is also a complete violation of the Sixth Amendment right to counsel, and a defendant might not be required to show prejudice to his defense.181 In United States v. Gonzalez-Lopez, the Supreme Court stated that “the right to select counsel of one’s choice . . . has never been derived from the Sixth Amendment’s purpose of ensuring a fair trial,” but rather is “regarded as the root meaning of the constitutional guarantee.”182 Thus, erroneous deprivations of counsel of choice may constitute structural errors which pervade the entire trial, removing any requirement to show prejudice.

To determine the effect of wrongful denial of choice of counsel [is not to look] for mistakes committed by the actual counsel, but for differences in the defense that would have been made by the rejected counsel . . . . We would have to speculate upon what matters the rejected counsel would have handled differently . . . . [a]nd then we would have to speculate upon what effect those different choices had or different intangibles might have had. The difficulties of conducting [assessments of prejudice for wrongful denial of counsel of choice and ineffective assistance of counsel] are not remotely comparable.183

The court’s holding in Stein III came after it noted that at least some of the defendants had retained multiple counsel prior to being indicted but were forced to terminate some of their counsel when KPMG cut off payments for their legal fees.184 However, the right to counsel of choice was implicated even prior to the district court’s affirmative holding in Stein III.

180 Wheat, 486 U.S. at 159. “The right to counsel of choice . . . is not absolute. When a defendant’s selection of counsel, under the particular . . . circumstances of a case, gravely imperils the prospect of a fair trial, a court may justifiably refuse to accede to the choice. Thus a trial court may in certain situations reject a defendant’s choice of counsel on the ground of a potential conflict of interest, because a serious conflict may indeed destroy the integrity of the trial process.” Id. at 166 (Marshall, J., dissenting); see also Lainfiesta v. Artuz, 253 F.3d 151, 154 (2d Cir. 2001) (stating that the right may be overcome if it is “outweighed by competing interests in the fair administration of justice or maintaining orderly trial procedures”).
182 Id. at 2563.
183 Id. at 2565. But see also Lainfiesta, 253 F.3d at 157 (holding, where a trial court limited cross-examination of witnesses to only one of defendant’s two attorneys, that such a denial may not warrant automatic reversal since it does not constitute an “[a]ctual or constructive denial of the assistance of counsel altogether” (internal quotation marks omitted)).
when the Stein I court dismissed the government’s claim that the Sixth Amendment did not entitle the KPMG defendants to spend “‘other people’s money’ on expensive defense counsel.”

The linchpin of the government’s argument against protection of the individual defendants’ use of other people’s money—in this case KPMG’s—was the Supreme Court’s decisions in Caplin & Drysdale, Chartered v. United States, and United States v. Monsanto. In Caplin and Monsanto, the Court addressed whether a federal statute allowing the government to seek a restraining order prohibiting the transfer of a defendant’s assets that were potentially forfeitable to the government as fruits of a violation of federal drug laws infringed on that defendant’s Sixth Amendment right to counsel of his choice. The Court held both in Caplin and Monsanto that no Sixth Amendment violation occurred since the statute only prohibited the use of forfeitable assets to obtain one’s counsel of choice.

Nevertheless, the Stein I court rejected the government’s reliance on these cases as relevant precedent for such an argument. First, the court pointed out that Caplin and Monsanto dealt with a defendant who sought to spend money that, being forfeitable under federal law, belonged to the government. Second, it interpreted Caplin as standing for the

---

186 Id.; see also Bloomberg News, supra note 60 (reporting on the federal prosecutors argument that “[t]he constitution only guarantees defendants the right to a lawyer, ‘not the best lawyer money can buy or a particular lawyer’”).
189 Caplin, 491 U.S. at 623-24; Monsanto, 491 U.S. at 614.
190 Caplin, 491 U.S. at 625. “A defendant has no Sixth Amendment right to spend another person’s money for services rendered by an attorney, even if those funds are the only way that that defendant will be able to retain the attorney of his choice . . . . [when t]he money, though in his possession, is not rightfully his . . . .” Id. at 626. Caplin and Monsanto were companion cases that were handed down by the Supreme Court on the same day. Consequently, the Court relied on its decision in Caplin to answer the same question in Monsanto. Monsanto, 491 U.S. at 614.
191 Stein I, 435 F. Supp. 2d at 367.
192 Id. at 367. In Caplin, the Supreme Court also compared defendant’s use of funds obtained through federal drug violations for payment of legal fees to that of a defendant wishing to use the proceeds of a bank robbery for the same purpose. 491 U.S. at 626. However, the comparison between the government’s interest in forfeited assets due to a drug transaction, and its interest in assets resulting from a bank robbery has been criticized by commentators. See Bruce J. Winick, Forfeiture of Attorneys’ Fees Under RICO and CCE and the Right to Counsel of Choice: The Constitutional Dilemma and How to Avoid It, 43 U. MIAMI L. REV. 765, 814 (1989) (claiming that the bank analogy fails because the assets are the rightful property of the bank, and must be returned). However, the purpose of the analogy in Caplin was to show that whether the
The proposition that the Sixth Amendment protects “a defendant’s right to spend his own money on a defense” and that the expectation of the KPMG employees that KPMG would cover their legal expenses arising out of any claims or charges based on their service with the firm was a property right that could not be interfered with by the government. The court went on to hold in *Stein III* that even though many of these defendants still retained their own counsel for their defense after KPMG had cut off legal fees, the right to counsel of choice “includes the right to a second lawyer or law firm if the defendant can afford it, either from his own resources or from those lawfully available to him from others.”

C. The Remedy Granted to the KPMG Defendants for Violation of Their Constitutional Rights and Its Impact on the Prosecution

After determining that the KPMG defendants’ Sixth Amendment rights had been violated, the court in *Stein I* and *Stein III* considered the appropriate remedy. In *Stein I*, the court did not grant the defendants’ motion to dismiss the indictments against the employees. Rather, it gave the defendants limited relief “tailored narrowly to the injury suffered” by suppressing certain statements made by the defendants that the government obtained as a result of its constitutional violations. However, the court did reserve the ability to make a decision regarding “whether to grant additional relief” including the option to dismiss the indictments at a later time.

---

Forfeitable assets are obtained by the defendant from a legally possessing third party (i.e., the bank), or obtained by the defendant without intruding on a legitimate third party claim (i.e., in drug cases), the fact that the assets are “tainted” by the illegal acquisition is what vests the government with its property interest through forfeiture. *Caplin*, 491 U.S. at 627.

193 *Stein I*, 435 F. Supp. 2d at 367.
194 *Id.*
195 *Stein III*, 495 F. Supp. 2d at 422.
196 Browning, *Tactic Questioned*, supra note 16.
197 *Stein I*, 435 F. Supp. 2d at 374.
199 *Stein I*, 435 F. Supp. 2d at 382.
200 See Lynnley Browning, *Judge Raises New Concerns About Tactics In Shelter Case*, N.Y. TIMES, July 14, 2006, at C3 [hereinafter Browning, *Judge Raises*] (reporting that one month after *Stein I* the judge suggested “he might consider postponing the trial or even dismissing the case”).
In Stein III, the court reexamined appropriate relief for the defendants after it was unable to force the government to reimburse the legal fees to the individual defendants as a result of the government’s sovereign immunity protection.\footnote{Stein I, 435 F. Supp. 2d at 376.} In addition, the defendant’s efforts to seek civil action against KPMG to have their legal fees paid failed for lack of jurisdiction.\footnote{The court in Stein I had originally held that it had the requisite jurisdiction over the defendants’ claims against KPMG for reimbursement and advancement of their legal fees. Stein I, 435 F. Supp. 2d at 379. However, this ruling was overturned by the Second Circuit on appeal. See Stein v. KPMG, LLP, 486 F.3d 753, 756 (2d Cir. 2007).} As a result, the court (albeit “with the greatest reluctance”) dismissed the indictments of thirteen employee defendants whose rights were violated due to KPMG’s discontinuing the payment of legal fees.\footnote{Stein III, 495 F. Supp. 2d at 423-27.} However, it denied dismissal to the remaining three defendants who were former employees of the firm and who had not shown that KPMG would have paid their defense costs “as a matter of either grace or obligation.”\footnote{Id. at 426-27.} In spite of this and as a result of its overreaching, the government lost the opportunity to prosecute many of the suspected individuals for their alleged criminal misconduct.

IV. THE DISTRICT COURT’S FACTUAL FINDINGS IN STEIN AND ITS WEAKNESSES

Before concluding that the government intruded on the Sixth Amendment right to counsel of the defendant employees, the Stein court arrived at a number of factual conclusions in order to ascertain whether the Sixth Amendment was even implicated. First, the court determined that KPMG would have advanced legal fees to its employees.\footnote{Stein I, 435 F. Supp. 2d at 353; Stein III, 495 F. Supp. 2d at 405.} This conclusion was based partly on a state statute giving KPMG the option (but not necessarily the obligation) to indemnify its employees through means which include the advancement of legal fees\footnote{Stein I, 435 F. Supp. 2d at 355 n.117 (citing the Delaware Revised Uniform Partnership Act, DEL. CODE ANN. tit. 6, § 15-110 (providing that, subject to the partnership agreement, “a partnership may, and shall have the power to, indemnify and hold harmless any partner or other person from and against any and all claims and demands whatsoever”)).} as well as on evidence showing that KPMG had a history of...
doing so. This finding was bolstered in *Stein III* by additional evidence that the court felt revealed KPMG’s true desire to cover litigation expenses for the indicted defendants. This included the voicemail message from KPMG to its partners offering to pay for counsel for any current or former members of the firm involved in the investigation, the severance agreement negotiated with Richard Smith that KPMG refused to sign after it sensed the pressure from the AUSAs, and notes taken by counsel for KPMG from the firm’s initial meeting with the USAO.

Next, the court concluded that the threat of an indictment based on the guidelines in the Thompson Memo and the pressure applied on KPMG by the AUSAs during the course of the investigation caused the company to “consider departing from its long-standing policy of paying legal fees and expenses.” However in arriving at these conclusions, the court was forced to make a number of circumstantial inferences that illuminate the difficulties faced in making constitutional determinations regarding the advancement of legal fees and that illustrate the weaknesses of the *Stein* court’s decisions.

### A. The Court’s Holdings Regarding Indemnification and Advancement

In assessing what legal obligation KPMG had to advance legal fees to its employees, the *Stein I* court first looked at whether the company was subject to state indemnification laws. Aware that all states have laws addressing company indemnification of employees and that these laws differ in terms of whether indemnification is

---

207 As the *Stein* court wrote:

KPMG had an unbroken track record of paying the legal expenses of its partners and employees incurred as a result of their jobs, without regard to cost. All of the . . . defendants therefore had, at a minimum, every reason to expect that KPMG would pay their legal expenses in connection with the government’s investigation and, if they were indicted, defending against any charges that arose out of their employment by KPMG.

*Stein I*, 435 F. Supp. 2d at 355-56; see also *Stein III*, 495 F. Supp. 2d at 402, 406.

208 *Stein III*, 495 F. Supp. 2d at 407.

209 Id. at 407-08.

210 *Stein I*, 435 F. Supp. 2d at 352 (also determining that KPMG had considered departing from its practice of advancing fees, even before any conversations with the AUSAs took place, based merely on the inherent threat of indictment posed in the Thompson Memo); *Stein III*, 495 F. Supp. 2d at 400.

211 *Stein I*, 435 F. Supp. 2d at 354.
permissive or required, the court had to determine what statutory obligations attached to KPMG. The court recognized that these statutes differ not just between states but also between different types of business entities within the same state. Since KPMG is a Delaware company, the court applied the indemnification laws of the state of Delaware. Next, since KPMG was formed as a limited liability partnership, the court distinguished between the indemnification statutes governing Delaware corporations and statutes governing Delaware partnerships. Finally, the court noted that the law governing individual defendants may change when an individual is an employee rather than a partner in the firm and works in a different state.

This analysis illustrates three issues courts must consider in evaluating a company’s statutory obligation to indemnify its employees. First, the court must identify the type of business entity employing the person subject to the litigation and whether it is covered under a state’s indemnification law.

---

212 Id. at 354-55 (citing 3A WILLIAM MEADE FLETCHER ET AL., FLETCHER CYCLOPEDIA OF THE LAW OF PRIVATE CORPORATIONS § 1344.10 (2002)).
213 Id. at 354. While indemnification laws vary between states, this Note will focus on Delaware’s indemnification law, both for simplicity as well as due to KPMG’s status as a Delaware Limited Liability Partnership.
214 Id. at 355 n.117.
215 Delaware law provides:

A corporation shall have power to indemnify any person who was or is a party or is threatened to be made a party to any . . . pending or completed action, suit or proceeding, whether civil, criminal, administrative or investigative . . . by reason of the fact that the person is or was a director, officer, employee or agent of the corporation . . . against expenses (including attorneys’ fees), judgments, fines and amounts paid in settlement actually and reasonably incurred by the person in connection with such action, suit or proceeding if the person acted in good faith and in a manner the person reasonably believed to be in . . . the best interests of the corporation, and, with respect to any criminal action or proceeding, had no reasonable cause to believe the person’s conduct was unlawful.

DEL. CODE ANN. tit. 8, § 145(a) (2006). Section 145(b) continues:

[N]o indemnification shall be made in respect of any claim . . . as to which such person shall have been adjudged to be liable to the corporation unless . . . the court . . . shall determine upon application that, despite the adjudication of liability but in view of all the circumstances of the case, such person is fairly and reasonably entitled to indemnity for such expenses which the Court . . . shall deem proper.

Id. § 145(b).
216 Id. tit. 6, § 15-110; Stein I, 435 F. Supp. 2d at 355 n.117.
217 See Stein I, 435 F. Supp. 2d at 356 n.119 (discussing KPMG defendants who were entitled to mandatory indemnification under California law).
cation law.\textsuperscript{218} Second, the court must then ascertain what classes of employees within the company (that is, officers, directors, partners, employees, etc.) are intended to be covered by the indemnification statute in question and whether a specific employee falls within that class. Finally, depending on those first two factors, the court must be sure that it is applying the indemnification statute of the proper state. This analysis becomes even more confusing when one realizes that, as in \textit{Stein}, the state statute may \textit{permit} indemnification but not \textit{require} it.\textsuperscript{219}

The \textit{Stein I} court noted that the Delaware indemnification law is subject to any “standards and restrictions” set out in KPMG’s partnership agreement.\textsuperscript{220} The court inferred that since KPMG’s agreement contained no such restrictions, rendering it “entirely free to indemnify its personnel,” KPMG would have necessarily done so.\textsuperscript{221} However, this argument also cuts the other way. KPMG’s statutory freedom to indemnify its personnel does not necessarily create an obligation to do so.\textsuperscript{222} Depending on the state whose indemnification law is applied, the statutory requirements, and the type of company in question, a showing of good faith by the individual being investigated may be required before he or she might be eligible for indemnification.\textsuperscript{223}

\textsuperscript{218} As a general example, the law in the state of Delaware governing companies set up as corporations does not govern companies set up as limited liability partnerships. See \textit{Del. Code Ann.}, tit. 8, § 145(a) (establishing permissive indemnification for Delaware corporations); \textit{id.}, tit. 6., § 15-110 (establishing permissive indemnification for Delaware partnerships).

\textsuperscript{219} \textit{See} Kurt A. Mayr, II, \textit{Note, Indemnification of Directors and Officers: The Double Whammy of Mandatory Indemnification Under Delaware Law in Walutch v. Conticommodity Services, Inc.}, 42 \textit{Vill. L. Rev.} 223, 223-24 (1997) (noting that, in response to concerns regarding director and officer personal liability, states enacted statutes to limit director and officer exposure through “indemnification statutes that empower corporations to indemnify their directors and officers . . . and, in some instances, requiring such indemnification” (emphasis added)).


\textsuperscript{221} \textit{Id.}

\textsuperscript{222} \textit{Sr. Tour Players 207 Mgmt. Co. v. Golftown 207 Holding Co.}, 853 A.2d 124, 127 n.5 (Del. Ch. 2004) (comparing the nearly identical language under Delaware limited liability company and limited partnership law, and stating that “[t]he statutory language is permissive and does not \textit{per se} create a right to indemnification”).

\textsuperscript{223} \textit{See Del. Code Ann.} tit. 8, § 145(a) (requiring that the person seeking indemnification act “in good faith and in a manner the person reasonably believed to be in or not opposed to the best interests of the corporation, and, with respect to any criminal action or proceeding, had no reasonable cause to believe the person’s conduct was unlawful”); Julie J. Bisceglia, \textit{Practical Aspects of Directors’ and Officers’ Liability Insurance—Allocating and Advancing Legal Fees and the Duty to Defend}, 32 \textit{UCLA L.
This exposes a weakness in the Stein court’s presumption that, absent the presence of the Thompson Memo, KPMG would have paid the defense costs for its partners and employees simply because it had the discretion to do so. With nothing in KPMG’s partnership agreement making any mention of indemnification, it is entirely possible that KPMG, or any company subject to government investigation, may choose to inquire as to whether an employee, partner, officer or director’s acts were in good faith and perceived as lawful before determining eligibility for indemnification. Additionally, as noted by the Stein I court, virtually all indemnification laws have one common characteristic: the right to indemnification is contingent on the defense to the legal proceedings being successful on the merits. Therefore, any obligation KPMG had to indemnify would not exist if the individual defendants were found guilty of the charges.

REV. 690, 696-97 (1985) (discussing how states have followed Delaware, permitting indemnification for expenses, fines, and judgment and settlement costs where the defendant acts in good faith and, in criminal actions, has no reasonable cause to believe his conduct is illegal); see also J. WILLIAM CALLISON & MAUREEN A. SULLIVAN, PARTNERSHIP LAW AND PRACTICE: GENERAL AND LIMITED PARTNERSHIPS § 10:6 (2007) (stating that a requirement of good faith may be set as a standard for indemnification); Sr. Tour Players, 853 A.2d at 128 (interpreting the operating agreement of a limited liability company which expressly denied indemnification for actions involving bad faith).


225 Indeed, it appears that even in KPMG’s case, an argument that good faith and perception of lawful conduct were absent could be made. See Weil, supra note 93, at C1 (reporting that one former partner who was indicted testified before the Senate Subcommittee regarding “his attempts to dissuade senior partners from approving one of the four shelters in question. . . . Emails showed he cautioned other KPMG partners that the strategy wasn’t legitimate”). This argument may be advanced even further by considering that KPMG’s decision to terminate the employment of some of its senior partners, see supra Part II.B, came even before its initial discussions with USAO. Stein I, 435 F. Supp. 2d at 339.

226 Stein I, 435 F. Supp. 2d at 355 (citing Homestore Inc. v. Tafeen, 888 A.2d 204, 211 (Del. 2005) as stating that while Delaware’s law “allows corporate officials to defend themselves in legal proceedings ‘secure in the knowledge that, if vindicated, the corporation will bear the expense of litigation[,] . . . indemnification cannot be established until after the defense to legal proceedings has been ‘successful on the merits or otherwise’” (footnotes omitted)); see also DEL. CODE ANN. tit. 8, § 145(c) (“To the extent that a . . . director or officer of a corporation has been successful on the merits . . . in defense of any action . . ., such person shall be indemnified against expenses (including attorneys’ fees) actually and reasonably incurred . . .”); CALLISON & SULLIVAN, supra note 223 (suggesting “success on the merits” as a “method for determining whether the standard for indemnification has been met” in the creation of a partnership agreement); Majkowski v. Am. Imaging Mgmt. Servs., LLC, 913 A.2d 572, 586 (Del. Ch. 2006) (“[A]n indemnification dispute cannot be resolved until after the merits of the underlying controversy are decided because the good faith standard requires a factual inquiry in the events that gave rise to the lawsuit.”).
Even assuming that indemnification does apply and that the defense can succeed on the merits, individual defendants in these types of criminal proceedings, without more, might still be faced with the extreme difficulty of covering the cost of the litigation.\textsuperscript{227} This exposes the defendants to “the personal out-of-pocket financial burden of paying the significant on-going expenses inevitably involved with investigations and legal proceedings.”\textsuperscript{228} At least in part out of these concerns, indemnification statutes generally allow for the advancement of legal fees before the conclusion of the case.\textsuperscript{229} One rationale for this is that advancement is needed as an inducement to attract the most capable individuals to positions of high responsibility in companies.\textsuperscript{230}

However, a major drawback to indemnification statutes that allow for the advancement of legal fees is that such advances are entirely permissive.\textsuperscript{231} A company may advance legal expenses in defense of litigation to its officers, directors, partners or employees, but may also choose not to provide for advancement at all, or to limit situations in which it would do so.\textsuperscript{232} One significant limitation is that defendants seeking advancement of legal fees prior to the final disposition of a proceeding may be required to deliver a written undertaking to repay any funds advanced if it is determined that they are not entitled to indemnification by the company.\textsuperscript{233}

\textsuperscript{227} Stein I, 435 F. Supp. 2d at 355. “The cost of a trial is out of the financial reach of many white-collar defendants. ‘It is hard to defend a white-collar case for less than $100,000, and most cost much, much more than that.’” Koppel, supra note 42 (quoting a Georgetown University McDonough School of Business professor).

\textsuperscript{228} Homestore, 888 A.2d at 211.


\textsuperscript{230} Homestore, 888 A.2d at 211; Falvey & Taylor, supra note 229, at 14 (arguing that advancement allows directors to act in the best interests of the corporation while knowing that they can resist meritless suits and not be forced to fund their own legal defense).

\textsuperscript{231} 18B AM. JUR. 2D Corporations § 1651; Bisceglia, supra note 223, at 709-10.

\textsuperscript{232} 18B AM. JUR. 2D Corporations § 1651.

\textsuperscript{233} 3A William Meade Fletcher et al., Fletcher Cyclopedia of the Law of Private Corporations §1344.10 (perm. ed., rev. vol. 2002) (discussing statutory indemnification and advancement in the context of corporations); DEL. CODE. ANN. Tit. 8, § 145(e) (requiring an officer or director to deliver an undertaking to repay legal fees where a corporation chooses to advance); see also Sr. Tour Players, 853 A.2d at 127 n.5, 129 (discussing the “broad authority” given to limited liability companies and limited partnerships under Delaware law to set their indemnification provisions and “to require a written undertaking as a condition to advancement”).
Thus, while indemnification and advancement may be correlative, they are “separate and distinct legal actions. [A defendant’s] right to advancement is not dependent on [his or her] right to indemnification.”

[A corporation may, through its certificate of incorporation or bylaws, or in a contract specifically addressing the issue, make mandatory the advancement of expenses to a director or former director for defending in a covered proceeding. Such a provision may be enforced as a contract. However, if the applicable corporate indemnification statute is permissive, then a provision in a corporation’s bylaws requiring the corporation to indemnify its directors, officers, employees, and agents to the extent permitted by law is not mandatory and does not require the corporation to advance litigation expenses before the termination of the proceeding in which the expenses were incurred.

This presents another issue that courts must address: determining what, if any, provisions regarding indemnification and advancement exist in a company’s bylaws or partnership agreement, or in other express contractual agreements, and then whether any such provision entitles a particular employee to advancement. If a company has no expressly stated provisions regarding indemnification and advancement in its corporate bylaws or partnership agreement, this becomes substantially more difficult.

As this was the case for KPMG in Stein, the court overcame this difficulty in two ways. First, the court found that

---

234 Homestore, 888 A.2d at 212; accord Sr. Tour Players, 853 A.2d at 128.
235 18B A.M. Jur. 2d Corporations § 1651; accord Homestore, 888 A.2d at 212 (stating that while the advancement authority granted by the indemnification statutes is permissive, “mandatory advancement provisions are set forth in a great many corporate charters, bylaws and indemnification agreements”).
236 Id. at 213 (“The scope of an advancement proceeding is usually summary in nature and limited to determining the issue of entitlement in accordance with the corporation’s own uniquely crafted advancement provisions.”).
237 Callison & Sullivan, supra note 223 (explaining that partnership agreements usually contain indemnification provisions where the limited partnership promises to indemnify for liabilities incurred as a result of partnership business, as long as the liabilities are not the result of “bad faith, willful misconduct, or gross negligence” and that these provisions should be drafted to authorize advancement of litigation costs prior to a determination of liability while requiring repayment if it is determined that the partner breached the standard of care); see also Sr. Tour Players, 853 A.2d at 130 n.23 (distinguishing between cases where an undertaking by an employee defendant was not required under Delaware corporate law since advancement was provided for under a corporation’s specific bylaws and was not conditioned on an undertaking, and cases where a corporation was entitled to deny advancement even where the corporate officer offered to submit to an undertaking since no mandatory advancement provision existed in the certificate of incorporation or bylaws).
the employees’ expectations of advancement based on KPMG’s prior practice of advancing legal fees constituted a property interest with which the government was not permitted to interfere.\textsuperscript{238} Second, the court suggested (although it admittedly declined to decide in this ruling) that all of the defendants were protected by a contract implied-in-fact (with the exception of Stein who had an express agreement with the company).\textsuperscript{239} These conclusions, however, are flawed.

The first flaw is that any expectation the employees’ had based on KPMG’s “long-standing policy,” as well as the suggestion that a contract implied in fact existed, is misplaced since one of the stipulations between the government and KPMG was that the company’s practice of advancing and paying legal fees prior to February 2004 was “voluntary.”\textsuperscript{240} Such a voluntary policy suggests that KPMG was free to exercise its discretion to deny advancement, either during the investigation or later during criminal proceedings, to any employee or former employee that it suspected did not act in good faith and was guilty of a criminal violation.\textsuperscript{241} Two factual circumstances in \textit{Stein} highlight this flaw: First, KPMG refused to extend payments for legal fees to certain former employees despite their involvement in the investigation. Second, KPMG also refused to extend fee payment to current employees even after the court found a Sixth Amendment violation in \textit{Stein I}.

With respect to the first circumstance, one might view KPMG as having already exercised such discretion under its voluntary practice when it declined to extend payment of legal fees to three former employees implicated in the alleged wrongdoing.\textsuperscript{242} Two of the former employees had left KPMG

\begin{footnotes}
\footnotetext[238]{\textit{Stein I}, 435 F. Supp. 2d at 366.}
\footnotetext[239]{\textit{Id.} at 356, n.119.}
\footnotetext[240]{\textit{Id.} at 352, 340; see also \textit{Stein v. KPMG, LLP}, 452 F. Supp. 2d 230 (S.D.N.Y. 2006), \textit{rev’d}, 486 F.3d 753, 762 n.3 (2d Cir. 2007) (stating in dictum that the stipulation by the defendants that KPMG chose to voluntarily advance legal fees in the past arguably estopped them from now arguing that KPMG had a contractual obligation—“implied or otherwise”—to pay their post-indictment legal fees, and that it is far from certain that KPMG would “lose on the merits” of any implied contract claim since the alleged “uniform practice” of paying legal fees for employees consists of “a single instance in which KPMG paid the legal fees of two partners indicted and convicted in a 1974 criminal case”).}
\footnotetext[241]{\textit{See Advanced Mining Sys., Inc. v. Fricke}, 623 A.2d 82, 84-85 (Del. Ch. 1992) (holding that a mandatory indemnification provision in a corporate by-law which is silent as to advancement does not “deprive the board of its function . . . to evaluate the corporation’s interest with respect to advancement of expenses”).}
\footnotetext[242]{\textit{Stein III}, 495 F. Supp. 2d 390, 425 (S.D.N.Y. 2007).}
\end{footnotes}
prior to the start of the investigation, but were suspected in the misconduct partially through conduct prior to their departure from the firm and partially through their formation of a separate company which played a “central role in the transactions at issue.”243 Yet the court in Stein III did not find the rights of those defendants to have been violated, despite evidence from the KPMG voicemail that the firm would cover expenses for “any present and former members of the firm.”244 Therefore, it may be argued that KPMG never intended to create an expectation that legal fees would be advanced unconditionally.

The other factual circumstance is that even after the district court’s finding in Stein I that the government had violated the defendant’s Sixth Amendment rights, KPMG still refused to extend payment of the legal expenses for their defendant employees either voluntarily245 or by conceding a contractual obligation to do so.246 This is significant because the Stein I court had viewed stipulations by KPMG to the AUSAs that it had no legal obligation to pay fees as being borne of the firm’s own self-interest in avoiding an indictment and not necessarily out of its actual belief.247 Yet even with the government’s obstructions removed, KPMG chose to exercise discretion even with regards to defendants who were current employees during the course of the suspected criminal wrongdoing. Such considerations cast further doubt on future courts’ efforts to reconcile such claims by attempting to infer what a company would have done, especially when a company chooses,

---

243 Id. at 426.
244 The Stein III court found that it could not be determined that KPMG had any legal obligation to defend, nor did the defendants show that the firm would have covered their legal costs “as a matter either of grace or of obligation, given that [they] left the firm so long ago.” Id. With respect to a third former defendant, the court found no intention of KPMG to cover his costs due to his departure under “strained circumstances” from the firm two years before the indictment, and that he had expressly released KPMG from any claims that he may have had against the firm. Id.
245 See Browning, Judge Raises New Concerns, supra note 200 (reporting a month after the Stein I decision was handed down that the firm had no intention of paying the legal fees for its defendant employees because while they had already paid $12 million in fees to that point, the employees had cost the firm $500 million by breaching their fiduciary duty).
246 United States v. Stein, 452 F. Supp. 2d 230, 260 (S.D.N.Y. 2006) (noting that KPMG contested claims of a contract implied in fact by asserting that past decisions regarding payment of employee legal expenses were made pursuant to “voluntary, unilateral decisions . . . on a case by case basis”).
247 Stein I, 435 F. Supp. 2d at 345 n.54 (“KPMG had an interest in avoiding advancement of fees if its legal obligation to do so might be questioned, as the government might view advancement of fees as protecting culpable personnel.”).
as KPMG did, to change course in its discretion as an investigation wears on.

A second major flaw in the court’s reasoning regarding the employees expectations is that KPMG arguably never intended to advance legal expenses to all employees being investigated by the DOJ. The only affirmative action taken by the company regarding advancement was to create an express contract for the two individuals the company unmistakably intended to cover: Jeffrey Stein and Richard Smith. Such action undercuts the argument that KPMG impliedly obligated itself to advance fees to all of its employees connected to the investigation. A better indication of the company’s true intent could be ascertained by looking at whether the company planned for such litigation expense issues by obtaining a liability insurance policy that provided protection in such instances. Such protection could make indemnification and advancement through company assets unnecessary.

Mandatory advancement clauses have been broadly interpreted to apply, even in situations where the advancement request was borne out of a lawsuit brought by the company providing the advance. Specific contract language is needed to entitle a person to mandatory advancement. A provision mandating indemnification “to the full extent permitted by Delaware law” will not “deprive the board of its function under Section 145(e) to evaluate a corporation’s interest with respect to advancement of expenses.” Where a bylaw mandates the advance of expenses, it creates a vested right, which cannot be unilaterally terminated, to advances once a triggering event for advances occurs. Without a bylaw or contract mandating the advance of expenses, a board determination to advance their personal litigation expenses is treated as a self-dealing transaction, governed by entire fairness. “A rubber-stamp resolution authorizing advances will not pass muster.”

Id. at 339. See generally Karl E. Stauss, Indemnification in Delaware: Balancing Policy Goals and Liabilities, 29 DEL. J. CORP. L. 143 (2004) (pointing out, at least with corporations, the inherent risks in blanket authorizations and clauses regarding mandatory advancement).

248 See David B. Bayless, Defending Your Client in the World of SEC Enforcement, Part 2: Cooperation and Litigation, SEC. LITIG. REP., Apr. 2006, at 1. The Bayless article tackles concerns arising from the Securities and Exchange Commission’s use of the Seaboard Report that are nearly identical to those emanating from the Thompson Memo, including pressure to cooperate with government investigations and a requirement that companies, in order to obtain settlements with the government, not exercise their right to indemnify individuals. Id. For corporations, this makes the purchase of a director and officer (“D&O”) insurance policy essential. Id. In addition, when purchasing such policies, companies can elect a “pay as you go” clause, which would permit advancement of defense costs on a current or quarterly basis. Id. But see John C. Tanner & David E. Howard, Blowing Whistles and Climbing Ladders: The Hidden Insurance Issues Behind Sarbanes-Oxley and Recent Corporate Governance Reform, ACC DOCKET, Apr. 2005, at 32, 35-37 (pointing out that many policies impose limits on cost coverage of government investigations, may contain terms that are subject to interpretation regarding coverage of certain costs, and may
B. The Court’s Findings Regarding the Threats and Pressure Imposed by the Government

In addition to finding that KPMG had a policy of advancing legal fees on which its employees were entitled to rely, the Stein decisions also focused on the effect that actions and threats by the government had on KPMG’s desire to advance fees.\textsuperscript{251} The court concluded first that the inherent threat of an indictment for failing to cooperate contained in the Thompson Memo caused KPMG to consider abandoning its policy of advancing legal fees, even before it first met with the AUSAs.\textsuperscript{252} One premise for this conclusion was the court’s belief that the Thompson Memo was “binding on all federal prosecutors.”\textsuperscript{253} Next, the court found that the threat of indictment was consistently reinforced by the USAO by focusing early and often on KPMG’s legal obligations regarding advancing legal fees and by allowing it to comply only with those obligations that were demonstrable.\textsuperscript{254} Third, the court found that the government’s conduct manifested a desire to minimize the involvement of defense counsel.\textsuperscript{255} Finally, the court determined that the firm’s decision to cease payments to any indicted employees and the conditions placed on the receipt of these payments were the product of direct pressure applied by the government pursuant to the Thompson Memo.\textsuperscript{256} Again, however, a number of weaknesses limit the effectiveness of the court’s decision.

First, KPMG’s fear of prosecution had little, if anything, to do with the Thompson Memo and more to do with the government’s general discretion to bring an indictment, which would have effectively crippled the company. As the Stein I

\textsuperscript{251} Stein I, 435 F. Supp. 2d at 352; Stein III, 495 F. Supp. 2d at 400.
\textsuperscript{252} Stein I, 435 F. Supp. 2d at 352.
\textsuperscript{253} Id. at 338.
\textsuperscript{254} Id. at 352; Stein III, 495 F. Supp. 2d at 402.
\textsuperscript{255} See Stein I, 435 F. Supp. 2d at 352.
\textsuperscript{256} Id.
court recognized, this discretion was grounds for concern for KPMG following the investigation, indictment, and prosecution of KPMG's former competitor, Arthur Andersen.\footnote{Id. at 341; see also supra notes 74, 97 and accompanying text.} Since prosecutorial discretion to seek an indictment against a company exists independently of the Thompson Memo guidelines,\footnote{See McNulty Statement, supra note 4, at 2 (“Federal prosecutors could lawfully exercise their discretion to charge a corporation in many instances where we have stayed our hand.”).} the decision to seek an indictment could hinge on the government's satisfaction with a company's cooperation irrespective of any procedural requirement\footnote{The Stein I court found that AUSAs, in determining whether to bring an indictment against a company under the Thompson Memo, are "obliged" to consider a company's cooperation and that the advancement of legal fees are a measurement of that cooperation. Stein I, 435 F. Supp. 2d. at 338. The court used this logic to differentiate between the Thompson Memo and the Holder Memo from 1999. Id. As authority for this holding, however, the court referred to a memorandum from Robert D. McCallum, Jr., Acting Deputy Attorney General. Memorandum from Robert D. McCallum, Jr., Deputy Attorney Gen. on Waiver of Corporate Attorney-Client and Work Product Protection to the Heads of Dep't Components, U.S. Attorneys (Oct. 21, 2005). However, this memo was dated October 21, 2005—over a year after the IRS made its criminal referral to the DOJ. Id. The meetings between the AUSAs and counsel for KPMG took place on February 25, 2004, twenty months before the McCallum memo. Stein I, 435 F. Supp. 2d at 341.} that it be considered.

Second, courts attempting to resolve such disputes regarding the advancement of legal fees risk intruding on the government's compelling interest and legitimate discretion in prosecuting individuals and companies liable for misconduct.\footnote{See Braswell v. United States, 487 U.S. 99, 115 n.9 (1988) (“White collar crime is 'the most serious and all-pervasive crime problem in America today.' Although this statement was made in 1980, there is no reason to think the problem has diminished in the meantime.” (quoting John Conyers, Jr., Corporate and White-Collar Crime: A View by the Chairman of the House Subcommittee on Crime, 17 AM. CRIM. L. REV. 287, 288 (1980))).} Prosecutors have broad discretion in determining when and what charges to bring against a defendant.\footnote{See Fisher, supra note 74, at 30 (“It is well settled that business entities may be indicted. . . . Prosecutors have discretion as to whom to prosecute. This discretion is limited by constitutional considerations . . . [but otherwise] the prosecutor is virtually without legal limit as to whom she prosecutes . . . .” (paragraph break omitted)).} Claims of abuse of prosecutorial discretion are rarely reviewed as such discretion is grounded in the separation of powers doctrine of constitutional theory.\footnote{Greg Hollon, Note, After the Federalization Binge: A Civil Liberties Hangover, 31 HARV. C.R.-C.L. L. REV. 499, 508 (1996); BENNETT L. GERSHMAN, PROSECUTORIAL MISCONDUCT § 4:3, at 4-6 (2d ed. 2001) (cautioning that while judicial deference to prosecutors arises from the separation of powers doctrine and respect for}
discretion, stating that “prosecutors appropriately are given great latitude in the arguments they make to juries,” and remarking on the importance of the government’s efforts to obtain cooperation during its investigations:

Any government’s interest in investigating and fairly prosecuting crime is compelling. . . . In order properly to accomplish that task, the government must have the ability to make just charging decisions and to prevent obstruction of its investigations. Hence, no one disputes the proposition that a willingness to cooperate with the government is an appropriate consideration in deciding whether to charge an entity. Nor does anyone suggest that an entity’s obstruction of a government investigation . . . should be ignored in a charging decision.264

In the post-Enron environment of company accountability, this government interest is also seen through the increased use of regulation designed to provide greater oversight of company conduct.265 Legislatures and administrative agencies have responded to a surge in white collar crime in recent years with statutes that increase transparency and reporting requirements and enhance criminal penalties for fraudulent conduct.266 However, as white collar corruption becomes increasingly complex, prosecution of such conduct remains an essential enforcement and deterrence mechanism.267 While

263 Stein I, 435 F. Supp. 2d at 359.
264 Id. at 363; see also Stein III, 495 F. Supp. 2d at 427.
265 See McNulty Statement, supra note 4 (arguing that following the outbreak of corporate scandals after the collapse of Enron, the DOJ’s ensuing “vigor and action” in prosecuting corporate crime along with congressional reform “have helped to instill a climate of accountability in corporate boardrooms, and to restore investors’ confidence in the integrity of our markets”).
266 See, e.g., Sarbanes Oxley Act, Pub. L. No. 107-204, 116 Stat. 745 (2002) (establishing penalties for criminal fraud in corporate record keeping in Title VIII, and increasing criminal penalties for mail and wire fraud, as well as under federal sentencing guidelines for white-collar offenses under Title IX); Disclosure of Proxy Voting Policies and Proxy Voting Records by Registered Management Investment Companies, Securities Act Release No. 8188, Exchange Act Release No. 47,304, Investment Company Act Release No. 25,922 (Jan. 31, 2003) (requiring mutual fund companies to provide disclosures on proxy voting policies relating to portfolio securities they hold, due in part to recent corporate scandals that have generated new investor interest in issues of corporate governance, as well as due to increased voting power that mutual funds enjoy as major shareholders and their effects on corporate accountability).
267 See Fisher, supra note 74, at 31 (commenting on accuracy and credibility in financial reporting, the effects of distrust on shareholders, creditors, managers and regulators, and the potential need to make an example of Arthur Andersen); see also Wray & Hur, supra note 2, at 1106 (“Satisfaction of the government’s interests of
abuse of prosecutorial discretion and the potential for vindictiveness by prosecutors are important concerns, forcing courts to evaluate such claims creates the risk that courts will overstep judicial authority under the separation of powers doctrine.

A third weakness limiting the effectiveness of the court’s holding is that decisions like those in Stein I and III may serve as catalysts for prosecutors to use alternative methods to obtain company cooperation, making legal fee advancement a moot point. The DOJ and USAO were chastised in Stein for placing an unfair burden on the individual KPMG defendants by denying them access to a source of funds they were lawfully entitled to for purposes of presenting a defense. However, the defendants’ access to these funds was subject to a second limitation: availability. Should a criminal indictment render a company insolvent, a separate burden would be placed on defendants wishing to gain access to these funds.

The negative financial impact that a formal indictment against a company has on that company’s assets and survivability may give prosecutors an interest in not seeking an indictment against that company. However, frustrating government investigations of illegal activity by burdening their acquisition of cooperation from a legitimate additional defendant—in this case the company—creates a risk that the government will simply bring an indictment, potentially rendering a company (as in the case of Arthur Andersen) retribution, deterrence, rehabilitation and restitution through the prosecution of culpable individuals will weigh against prosecution of the business entity.”

---

268 Fisher, supra note 74, at 32 (discussing how prosecutors can obtain indictments with “relative ease” and by using “questionable or even contrived evidence”); see also, Hollon, supra note 262, at 508 (stating that a prosecutor’s charging decisions are entitled to the presumption that they were made in good faith); Note, Prosecutorial Vindictiveness in the Criminal Appellate Process: Due Process Protection After United States v. Goodwin, 81 Mich. L. Rev. 194, 195-96 (1982).

269 Hollon, supra note 262, at 508.

270 Stein I, 435 F. Supp. 2d at 362.

271 See Tanner & Howard, supra note 249, at 46 (arguing that even where employees such as in-house attorneys have mandatory indemnification through state law or written agreements, “individual financial resources will be at risk if the company files for bankruptcy or becomes insolvent”); Bisceglia, supra note 223, at 710-11.

272 See Thompson Memo, supra note 1, at 3 (explicitly raising adverse risks to shareholders, pension holders, and innocent employees as factors to consider in deciding whether to charge a company); McNulty Statement, supra note 4.
This will place the burden of legal fees more heavily on defendants who would otherwise have sought company funds. Even assuming prosecutors did abuse their discretion in seeking an indictment, it is arguable whether a reversal or the dropping of any charges would restore the company's financial position.

This Note does not claim that the Stein court erred in its conclusions, or misrepresented the issues in arriving at its holdings. It merely seeks to present the wealth of issues, arguments, and claims that could be made in response to the ruling issued by the Stein court, as well as may be made to other courts in similar, if not identical, situations moving forward. In the face of these complications related to judicial reconciliation of legal fee disputes for defendants charged with corporate crime, an easier solution is needed.
V. PROPOSING A LEGISLATIVE SOLUTION

In arriving at its final conclusions, the court in the Stein cases engaged in a long and complicated analysis of the arguments presented by both the government and the defendants, while also balancing policy concerns. The court rejected the government’s claim that the right to counsel does not include the right to spend other people’s money. However, the government’s argument indirectly presents a more viable solution for balancing its interest in preventing the obstructive use of legal fee advancement and a defendant’s right to use a lawful source of funding to secure counsel and mount a defense. This solution is the use of preliminary injunctive relief in future government claims of obstructive fee advancement by employers to employees. Such relief would be similar to the injunctions granted under current legislative forfeiture provisions.

A. Use of Legislative Injunctive Restraints

The use of legislative preliminary injunctions, similar to those at issue in Caplin and Monsanto, is a superior alternative to the approach taken by the Thompson Memo because it reduces the risk of subsequent disputes among the government, employers, and their employees over the advancement of legal fees while withstanding constitutional scrutiny of Sixth Amendment claims. Such injunctions also alleviate the burden on future courts forced to reconcile similar claims and arguments, and they preserve a defendant’s Sixth Amendment rights prior to the need for judicial intervention to restore what the government has violated. This alternative would also minimize judicial intrusion on prosecutorial discretion, while at the same time limiting over-zealous prosecutors. This in turn reduces the risk of damage to legitimate criminal prosecutions caused by suppression of statements and other evidence and, in
cases like Stein, forced dismissal of unprosecuted criminal indictments should be tried on their merits.\textsuperscript{281}

The legislative injunction at issue in Caplin and Monsanto, which this Note suggests as a model, is the Continuing Criminal Enterprise Act (“CCE”).\textsuperscript{282} The CCE is one of two federal criminal statutes that were amended by the Comprehensive Forfeiture Act of 1984 (“CFA”)\textsuperscript{283} to give greater effect to government efforts to fight different types of crime.\textsuperscript{284} The CCE, as modified by the CFA, authorizes the forfeiture of “property constituting, or derived from, any proceeds . . . obtained . . . as the result of such [drug] violation[s].”\textsuperscript{285} Included in such property is the forfeiture of any “interest in, claims against, and property or contractual rights affording a source of control over, the continuing criminal enterprise.”\textsuperscript{286} The statute also declares that any “right, title, and interest” in the property obtained via a violation of the drug law “vests in the United States upon the commission of the act giving rise to forfeiture” even when those assets are subsequently transferred to a third party.\textsuperscript{287}

The forfeiture statute has a significant preemptive feature. This feature gives the government the ability to apply for a restraining order or injunction in order “to preserve the availability of [the] property” either upon the filing of an indictment or information charging a violation, or prior to the filing of an indictment or information if the court determines that there is a substantial probability that the U.S. will prevail on the issue of forfeiture and that failure to restrain the property will result in its being made unavailable.\textsuperscript{288} The court must also weigh the need to preserve the property against

\textsuperscript{281} Stein III, 495 F. Supp. 2d at 427 (“This indictment charges serious crimes. They should have been decided on the merits as to every defendant.”).
\textsuperscript{283} Id. § 881(e).
\textsuperscript{284} See Roderick D. Vereen, Comment, Attorneys Rights to Fees Under the Comprehensive Forfeiture Act of 1984: The “Bona Fide Purchaser,” 16 S.U. L. REV. 407, 408-09 (1989) (“[The CFA] revised the forfeiture provisions of both the RICO and the CCE statutes. . . which attack[] the enterprises of drug trafficking and racketeering. . . Congress believed that since profit and economic power were the motivating factors for this type of criminal activity enforcement of those statutes would strip these offenders and organizations of this economic power.”).
\textsuperscript{286} Id. § 853(a)(3).
\textsuperscript{287} Id. § 853(c). However, an exception applies when the third party who has obtained the assets proves that he or she is a “bona fide purchaser” of the property “reasonably without cause to believe” it is forfeitable. Id.
\textsuperscript{288} Id. § 853(e).
the hardship on the individual whose assets have been “preserved.”\textsuperscript{289} The duration of the restraint is limited to ninety days, unless either good cause is shown to the court to extend the order, or an indictment is filed against the individual.\textsuperscript{290}

A carefully drawn statutory enactment allowing the government, with certain limitations, to request that companies be enjoined from advancing legal fees to certain employees can achieve the goals of easing judicial reconciliation, preserving defendants’ Sixth Amendment right to counsel, and limiting intrusion on prosecutorial discretion, while also addressing other policy concerns. Such a provision would consist of two parts. The first part would allow prosecutors, after initiating a formal investigation against an employee suspected of wrongdoing in the capacity of his employment, to seek, subject to limitations, a preliminary injunction from a court barring the advancement of legal fees from the employer to that employee. Like the CCE, the government would not be required to bring an indictment against the individual employee or the company. However, in order for the court to grant such an injunction, the government would be required to show to the magistrate evidence sufficient to obtain an indictment against the company and a substantial likelihood that the fee advancement is being used for the purpose of obstructing the government’s investigation. In addition, all injunction requests would allow potential defendants as well as their employers to rebut the government’s claims through presentation of evidence sufficient to show either a lack of obstructive conduct or excessive hardship to the potential defendant.

The first part of this proposed injunction provision accomplishes two goals. First, it preserves the government’s interest in obtaining cooperation and preventing obstruction by giving prosecutors the authority to seek judicial intervention in the prevention of companies using fee advancement as a mode of improperly obstructing a criminal investigation.\textsuperscript{291} At the same time, the government’s interest would be sufficiently checked by setting a minimum on what must be proven to

\textsuperscript{289} Id. § 853(e)(1)(B)(ii).

\textsuperscript{290} Id.

\textsuperscript{291} See Wray & Hur, supra note 2, at 1146 (discussing the DOJ’s “more aggressive pursuit of obstructive conduct since the Thompson Memo’s issuance,” but also arguing that companies which impede investigators have for a long time been more likely to face criminal prosecution at the hands of the Justice Department).
establish the necessary urgency implicating that interest. Second, this provision would preserve the Sixth Amendment right to counsel for individual defendants by placing the burden on the government to show good cause for judicial interference with a company’s right to advance fees to its employees. It also removes the government’s unrestrained consideration of legal fee advancement in its evaluation of a company’s cooperation, alleviating the threat contained in the Thompson Memo of prosecutors seeking an indictment against the company. Should the court grant the injunction, the company would have no choice but to comply. Should the court deny the injunction, prosecutors would be otherwise barred, both before and after the denial of the motion, from considering fee advancement in deciding whether or not to seek an indictment.

The second part of the injunction provision would place additional limits on when such an injunction would be appropriate. First, prosecutors would not be permitted to request an injunction when it would be construed as forcing an employer to violate the terms of a statutory obligation requiring the advancement of legal fees. Second, prosecutors would not be able to seek an injunction that would intrude on any insurance policy obtained by the company from a third party insurer for the purpose of guaranteeing the availability of advanced legal fees. Third, prosecutors would be unable to seek an injunction when a company already has an existing uncontradicted and express agreement with its employees to advance legal fees, unless prosecutors can show that the agreement was put in place for purposes of obstructing the government’s investigation. In such a case the government would be required to prove the obstruction in accordance with part one of the provision. Courts would also be required to weigh the government’s interest in injunctive relief against any hardship to the defendant and could limit the time period of any restraint based on those perceived hardships. Finally, if none of the

292 See Stein I, 435 F. Supp. 2d at 363 (conceding the need of the government to prevent obstruction in its criminal investigations, but pointing out that “the Thompson Memorandum does not say that payment of legal fees may cut in favor of indictment only if it is used as a means to obstruct an investigation”). “If the government means to take the payment of legal fees into account in making charging decisions only where the payments are part of an obstruction scheme—and thereby narrowly tailor its means to its ends—it would be easy enough to say so.” Id. at 364.

293 See Wray & Hur, supra note 2, at 1103 (stating that the use of alternatives such as “pre-trial diversion” agreements or “deferred prosecution and non-prosecution agreements,” encourages greater cooperation by removing the “all-or-nothing choice between indicting (and destroying) a company and giving it a complete ‘pass’”).
above limitations apply and if the court finds that abrogation of fee advancement is necessary, nothing in the injunction provision would intrude on a company’s right or obligation to indemnify its employees at the conclusion of any litigation in accordance with any obligation imposed by statute or by a company’s bylaws, partnership agreement, or other express contractual obligations.

These limitations accomplish a number of goals. The first limitation protects the interests of states that wish to preserve the use of indemnification and fee advancement as a significant inducement to attract capable individuals to positions of high responsibility in a company. The second and third limitations prevent companies that may only be subject to permissive statutes dealing with fee advancement from being effectively punished by prosecutors for their attempts to plan for such a contingency, while encouraging other companies that have not yet done so to undertake one of these options in an effort to eliminate the issue from consideration going forward. All three limitations would constitute express legal obligations of companies to advance legal fees—obligations that the Thompson Memo had recognized by prohibiting prosecutors from considering such mandatory payments as a failure to cooperate. Thus, prosecutors would be in no worse position than they had voluntarily undertaken under the Thompson Memo, while also receiving an additional tool to combat a company’s obstruction under the guise of cooperation. The expensive nature of complex white-collar criminal litigation, such as that presented in the Stein cases, would also be taken into consideration by courts evaluating hardship to defendants.

294 See supra note 230 and accompanying text.

295 See Jonathan C. Dickey, Recent Decisions Concerning the Right of Advancement of Defense Costs, SEC. LITIG. REP., Feb. 2006, at 6 (explaining while companies can do only so much to mitigate the risk of being punished by prosecutors for supporting officers and directors through indemnification and advancement, certain steps such as amending company bylaws to make advancement mandatory; contractual exclusions eliminating the obligation to advance fees to employees who plead guilty to criminal misconduct; and purchasing non-rescindable D&O liability insurance are some suggestions to preserve the right to pay for the defense of officers and employees); see also Falvey & Taylor, supra note 229, at 16.

296 Thompson Memo, supra note 1; McNulty Statement, supra note 4 (removing from consideration a company’s statutory obligation to pay employee defense costs in the government’s evaluation of that company’s cooperation).
B. Arguments Against and Rebuttals in Favor of the Use of Injunctive Restraints

While this Note proposes using injunctive restraint as a tool against obstructive fee advancement, injunctive provisions similar to the one proposed have been increasingly criticized out of concerns pertaining to prosecutorial abuse as well as risks regarding availability and adequacy of counsel. The first criticism is that such injunctive provisions have undesirable impacts on the adversarial criminal justice system. The argument is that such provisions remove any incentive for private criminal defense attorneys to represent defendants whose assets are subject to pre-trial restraint because the defendant’s ability to cover his legal expenses is suddenly called into question. Under forfeiture provisions involving drug offenses, if the defendant is convicted, the restrained assets would become forfeited, potentially leaving nothing to cover attorney’s fees. Critics might argue that the proposal suggested by this Note creates a similar risk through the possibility that defendants who are found guilty would no longer be entitled to indemnification from their respective companies, a notable concern where defense costs can reach extraordinary levels.

A second argument against the use of an injunction provision is that such a restraint on a defendant’s assets has a direct implication on his or her Sixth Amendment right to counsel of choice and, more indirectly, on the right to the effective assistance of counsel. Third, it may be argued that

297 See Winick, supra note 192; Berube, supra note 110, at 1395.
299 Berube, supra note 110, at 1395; Winick, supra note 192, at 779, 785. The argument is that many private criminal defense attorneys require a fee that defendants cannot afford to pay if their assets are seized. Id. at 773. This argument is furthered in the sense that criminal defendants whose assets are subject to forfeiture and have been frozen prior to an indictment, can claim indigency, and have counsel appointed to them, while defendants whose assets have not been frozen are unable to claim indigency, are unable to retain counsel since “lawyers will refuse to represent him, fearing subsequent forfeiture of their fees.” Rackner, supra note 298, at 134.
300 This concern is increased by the requirement that the illegal assets transferred to a third party are also subject to forfeiture, including legal fees paid to defense attorneys. Vereen, supra note 284, at 409.
301 See supra note 154 and accompanying text.
302 Winick, supra note 192, at 784-85, 800-01 (arguing that, even where courts have accepted government contentions that giving defendants who have had their assets forfeited appointed counsel satisfies the requirements of the Sixth Amendment,
such an injunctive provision would not remove the risk of abuse by prosecutors seeking pre-trial freezing of assets in order to purposely eliminate a defendant’s chosen defense counsel. However, these arguments are weakened by a number of considerations.

1. Overcoming Negative Impacts on the Adversarial Criminal Justice System

First, the arguments regarding the impact of preemptive restraints and forfeiture on the adversary system rely on the assumption that private defense attorneys will be hesitant to represent defendants out of the fear that they either will not be paid or that the government will forfeit those fees that have already been paid. Second, the argument also relies on the absence of any exception for legitimate attorney’s fees and the overly broad use of seizures and injunctions by prosecutors in freezing the defendant’s personal assets, making no distinction between those which are tainted by the illegality and those which are legitimate. The effect is to leave the defendants with little or no other source to finance their defense.

“[a] criminal justice system relying on appointed rather than retained counsel . . . is not the adversary system contemplated by the sixth amendment, no matter how effective such appointed advocates are”).

303 Id. at 777-78 (arguing that such a result was neither anticipated nor intended by Congress); Peter W. Salsich, III, A Delicate Balance: Making Criminal Forfeiture a Viable Law Enforcement Tool And Satisfying Due Process After United States v. James Daniel Good Real Property, 39 ST. LOUIS U. L.J. 585, 586 (1995). This was also an express concern and conclusion of the Stein court when it ruled that one of the purposes of the AUSAs in exerting such pressure on KPMG pursuant to the Thompson Memo was their “desire to minimize the involvement of defense attorneys.” Stein I, 435 F. Supp. 2d at 353.

304 Winick, supra note 192, at 777-81; Berube, supra note 110, at 1395.

305 While an exception is carved out under the CCE for bona fide recipients of potentially forfeitable funds who are without cause to believe that the received funds are forfeitable, supra note 287, defense attorneys are almost always excluded from this category. Winick, supra note 192, at 785 (“[C]riminal defense lawyers are almost inevitably on notice that their clients’ payments may be from the proceeds of crime.”); Vereen, supra note 284, at 410-11; see also supra note 296.

306 See Jon E. Gordon, Prosecutors Who Seize Too Much and the Theories They Love: Money Laundering, Facilitation, and Forfeiture, 44 DUKE L.J. 744, 744-45 (1995) (“Since 1970, federal prosecutors have increasingly relied on civil and criminal forfeiture as tools for law enforcement. . . . [This] gives prosecutors undesirable power to seize property.”); see also Winick, supra note 192, at 770-71 (arguing that the ease with which prosecutors obtain pre-trial restraining orders combined with the threat of post-conviction forfeiture of assets “have deterred or prevented private criminal defense attorneys from taking such cases,” and give prosecutors “almost unfettered discretion to deprive defendants of the use of their assets to hire counsel of choice”).
However, the solution proposed in this Note assuages these concerns.

The injunction provision proposed here would actually give private defense counsel less reason to hesitate since it is not concerned with the seizure of illegal, and therefore forfeitable, assets such as in drug cases, but rather with preventing legal assets from being used for illegal purposes. As a result, the provision would not subject fees already paid to defense attorneys to any retroactive forfeiture by the government and would only apply to the further advancement of legal fees from the employer to the employee, leaving the employee's personal assets unaffected. Thus, the employee would be free to expend his own assets in retaining an attorney for his defense307 since all sources of personal financial income, such as an employee's salary and investment income (assuming no other illegitimacy) are preserved. As a result, private defense counsel would receive some compensation from the defendant’s personal assets, while still retaining the prospect of indemnification at the conclusion of the trial, alleviating at least some of the risk that attorneys will be deterred from representing such defendants.308 Additionally, not subjecting fees already paid to retroactive forfeiture and preservation of a defendant’s personal assets prevents prosecutors from sweeping too broadly in their application of such a provision.309

307 Winick, supra note 192, at 811 (“[The problem] is not whether exercise of the right to counsel of choice can be regulated, such as by restrictions on the choice of counsel who is otherwise engaged in order to prevent undue delay, or on the choice of an attorney disqualified by a factor such as a conflict of interest. Rather, it is whether the right may be completely destroyed by governmental action that renders the defendant unable to choose any private counsel.”). Since the provision leaves a defendant’s personal assets intact, such a concern would not be implicated under this proposal. As critics such as Winick note, “There is a vast difference between overriding a defendant’s choice of a particular lawyer and preventing him from employing any lawyer at all.” Id.

308 See Browning, Judges Press, supra note 33 (reporting that in at least one white collar criminal case against a former Enron accountant, retained defense counsel continued to represent the defendant at a “small fraction of his usual charges” because he was “convinced of her innocence”); Stein III, 495 F. Supp. 2d 390, 423-24 (S.D.N.Y. 2007) (finding that while some of the KPMG defendants had gone into debt in financing their defenses, most of the defendants were in “better financial circumstances,” and only one was threatened with a motion to withdraw by their chosen counsel).

309 One example of the broad application of forfeiture provisions in criminal cases is seen in charges of money laundering where, when dealing with assets subject to forfeiture, “dirty money” that is tainted by the alleged illegality is mixed with “clean money” that is unconnected. Gordon, supra note 306, at 744. When prosecutors have great difficulty sorting the dirty from the clean, they may attempt to try and seize it all. Id. at 744-45.
2. Overcoming Negative Impacts on Rights to Counsel of Choice and Effective Assistance of Counsel

Second, criticism based on the impact of injunctive restraint on a defendant’s rights to counsel of choice and the effective assistance of counsel is also addressed by the provision suggested in this Note. Part of this concern is based, again, on current forfeiture provisions which leave defendants with no other source of funding for their defense. Another source of this concern, however, is that private defense attorneys who agree to take on such cases are presented with a conflict of interest by having a pecuniary interest in the outcome of a criminal litigation. However, the narrow application of the provision recommended here subverts these concerns. Under this proposal, prosecutors would be unable to request a restraint of fee advancement in cases where such a restraint contravenes a statutory obligation imposed by the state or impinges on a company’s clearly demonstrated legal obligation to its employees through express provisions in their bylaws or other contractual agreement. This provision would therefore protect any property interest claim in fee advancement that an employee may have. To the extent that the government can seek such a restraint, it would be limited to the prosecutor’s ability to show that the fee advancement is being used for purposes of, or in connection with, efforts to obstruct their investigation, in which case the advancement would further an illegal interest not protected under the Constitution. Thus, even where a company is not protected by a prior existing legal obligation, either contractual or statutory, as long as their voluntary advancement of fees to investigated employees is done in good faith, the government will not be permitted to intrude.

310 See Winick, supra note 192, at 785; Rackner, supra note 298, at 135; Berube, supra note 110, at 1395.

311 Rackner, supra note 298, at 140-41 (arguing that such a conflict could violate the defendant’s Sixth Amendment right to conflict-free counsel); Winick, supra note 192, at 776-77 (discussing the added risk that defense attorneys will be representing a criminal defendant on a contingency basis, presenting an ethical violation of the Model Rules of Professional Conduct).

312 To the extent that a company uses a third party insurer to provide indemnification and advancement protection in the event of litigation, the problem could be avoided entirely.
3. Overcoming Risks of Prosecutorial Abuse

Third, the limitations in the proposal address the concerns of prosecutorial abuse. These arguments are based on a number of factors. The first is that prosecutors have increasingly relied on forfeiture provisions in their efforts to fight crimes such as money laundering, racketeering, drug trafficking, and mail or wire fraud. The second is that as this reliance has increased, so has prosecutors’ dependency on theories that allow them to sweep more broadly in order to seize and restrain a larger portion of a defendant’s assets while also enjoying a low burden of proof to establish the grounds necessary to implement the restraint. The provision proposed here eliminates these risks of abuse for a number of reasons.

First, prosecutorial reliance on this provision will be limited since, as was originally intended in the Thompson Memo, it is intended to assist prosecutors in deciding when to bring criminal charges against a company. As the memo itself stated, instances in which a company will be subject to indictment will be only in a minority of cases. Second, by placing the decision in the hands of a court, theoretically the only thing the prosecution can do to abridge the advancement of legal fees is request an injunction. Until the court grants

---

313 See Gordon, supra note 306, at 744-45 (warning against giving prosecutors “undesirable power to seize property”); Salsich, supra note 303, at 585-86.
314 Gordon, supra note 306, at 744.
315 An example of this is prosecutorial use of certain theories such as “taint” or “facilitation.” See id. at 744-45. The premise behind these theories is that prosecutors seeking the restraint of “dirty” money obtained as a result of money laundering or drug transactions will also seize “clean” money, not obtained through these activities by claiming either that the clean money was used to facilitate the illegal transaction, or that the dirty money was commingled with the clean money, causing it to be tainted and subject to forfeiture. Id. at 755.
316 Id. at 749. Under a number of forfeiture provisions prosecutors use, only a standard of probable cause is required to effectuate the restraint. See, e.g., 21 U.S.C. § 853 (e)(2) (1994) (allowing the government to obtain a pre-indictment temporary restraining without notice or opportunity for a hearing order upon a showing of probable cause that the property is forfeitable, and that notice would jeopardize the availability of the property for forfeiture); see also Monsanto, 491 U.S. at 615; Gordon, supra note 306, at 749-50.
317 Thompson Memo, supra note 1, at 1.
318 In this way, such a proposal also alleviates a potential conflict of interest created by the McNulty Memo, which requires prosecutors wishing to consider fee advancement as a failure to cooperate to obtain permission from the Deputy Attorney General. See supra note 277. By requiring a court to consider the merits of an injunction on the advancement of legal fees from employers to employees, this provision inserts the objectiveness of a neutral magistrate, eliminating the risk of abuse by prosecutors. See, e.g., Coolidge v. New Hampshire, 403 U.S. 443, 449-55 (1971) (invalidating a warrant authorizing the Fourth Amendment search and seizure
of a defendant’s automobile where it is not issued by a “neutral and detached magistrate,” and holding that a state’s Attorney General cannot serve as a neutral, detached magistrate for the purpose of determining probable cause and issuing warrants). “[T]he whole point of [this] basic rule . . . is that prosecutors and policemen simply cannot be asked to maintain the requisite neutrality with regard to their own investigations—the ‘competitive enterprise’ that must rightly engage their single-minded attention.” Id. at 450.

319 Gordon, supra note 306, at 760 (“Prosecutors have enjoyed mixed success in achieving forfeiture . . . under a civil forfeiture provision] under the facilitation theory. . . . [S]everal courts have applied the theory to justify forfeiture . . . [while] . . . on the other hand, most attempts to apply the theory to accounts containing proceeds of other offenses have failed, and some courts have rejected the theory altogether.”).
interests for all involved, but also preserves the ideal of the adversary system—assessing responsibility for culpable conduct and meting out proper punishment while ensuring a fair trial for defendants. While the court in Stein was forced to embark on an arduous fact-finding process in order to resolve competing interests on both sides of the investigation, a more balanced and direct approach, which preserves the interests on both sides and reduces the risk of disputes and abuse, presents a superior alternative for courts to resolve future controversies over these matters.

This Note analyzed the resolutions the Stein court had to make with respect to the actions and intentions of the government, KPMG, and the KPMG employees who were subject to the investigation; the company’s obligations to its employees; and the company’s reaction to the Thompson Memo. The court struggled to apply these facts in its determination of the employees’ rights to effective assistance and counsel of choice under the Sixth Amendment and of the scope of the protections afforded by those rights. The difficulties of such a piecemeal, fact-intensive analysis demonstrate the need for a clearer alternative.

The injunctive provision proposed here would simplify this process for future courts by removing from contention many of the disputed issues in Stein. The proposed provision would encourage both the government and private employers to preempt these contentions and refrain from abusive conduct. The narrowly drawn injunctive provision suggested should also assuage fears of abuse that arise with respect to similar provisions in other criminal contexts. The injection of the court as an intermediary to ensure an early and neutral consideration of the facts presented and the interests implicated would significantly limit, or at least ease the resolution of, Sixth Amendment violations to the right to the assistance of counsel pertaining to the advancement of legal fees from employers to employees.

Jonathan Bashi†

† J.D. Candidate, Brooklyn Law School, 2008; B.A., University of Massachusetts, 2000. Thanks to everyone at the Brooklyn Law Review for their endless efforts, particularly Bradley Benedict and Jason Zakai, as well as to Professor Jason Mazzone for his guidance and helpful insights during this process. Special thanks to my loving parents and siblings as well as to Elizabeth Vicens for all of their love, support, and constant encouragement.
Why Federal Preemption Is Necessary to Create Uniform Professional Boxer Safety Standards

I. INTRODUCTION

Boxing is a dangerous sport.1 This was tragically illustrated twice within three months in 2005 when both Martin Sanchez and Leavander Johnson died after suffering injuries during professional boxing matches.2 Sanchez died shortly after suffering a brain bleed during a fight with Rustam Nugaev; Johnson, who also received a fatal brain injury, died five days after his fight with Jesus Chavez.3 While boxing is undoubtedly dangerous,4 extensive regulations covering boxer safety have supposedly made the sport safer.5 Despite the presence of these regulations, the circumstances surrounding the recent deaths of Sanchez and Johnson, especially the death of Sanchez, have led many to believe that the current safety regulations are not being properly enforced.6 For example, one of these regulations requires each boxer to undergo a pre-fight physical examination.7 However, there is evidence that

3 Id.; see infra Part III (discussing the circumstances surrounding these fatalities).
4 Thomas Hauser, one of the most well respected and informed journalists covering the boxing industry, has done an excellent job of summing up the sport of boxing and the importance of safety regulations; he wrote, “When boxing is run right, it’s a scary sport. When it’s run wrong, the risks become unacceptable.” Thomas Hauser, Medical Issues and the AAPRP, SECONDSOUT, http://www.secondsout.com/usa/colhauser.cfm?cs=208&cs=15463 (last visited Feb. 19, 2008).
5 McCain & Nahigian, supra note 1, at 18-23.
6 Hauser, Fighter Safety, supra note 2.
7 The regulation states:

No person may arrange, promote, organize, produce, or fight in a professional boxing match without meeting each of the following requirements or an alternative requirement in effect under regulations of a boxing commission
Sanchez’s pre-fight physical was either not done or done in an extremely negligent manner.\textsuperscript{8} And, because Sanchez did not speak English, and the ringside doctors who were supposed to attend to him as required by the regulations did not speak Spanish, the medical staff was unable to communicate with him during the fight.\textsuperscript{9} Unfortunately, there are examples from multiple states which indicate that improper enforcement of current regulations is a nationwide problem that has had a negative impact on the health and safety of many fighters.\textsuperscript{10}

The recent boxing deaths should not be surprising to anyone familiar with the sport. In fact, the sport of boxing has a long history of legal problems in the United States.\textsuperscript{11} Boxing is different from other contact sports in that conduct that would otherwise be considered criminal is lawful in the boxing ring. Simply put, the goal of a professional boxer is to knock his or her opponent unconscious, but outside of the ring this is considered assault.\textsuperscript{12} Therefore, while the game of basketball is legal without having a law about putting a leather ball through a metal hoop, people can only box legally when it is sanctioned by law. Another reason there have been legal problems in the boxing industry is that, because boxers receive a large amount of physical punishment throughout their careers, with the most serious injuries coming from blows to the head,\textsuperscript{13} there have been calls over the years to regulate boxing in order to make it

\begin{flushright}
that provides equivalent protection of the health and safety of boxers: (1) A physical examination of each boxer by a physician certifying whether or not the boxer is physically fit to safely compete, copies of which must be provided to the boxing commission. . . .
\end{flushright}

\textsuperscript{8} Hauser, \emph{Fighter Safety}, supra note 2; see infra Part III.

\textsuperscript{9} Hauser, \emph{Fighter Safety}, supra note 2. This clearly violates the spirit of the regulation stating that there must be a "physician continuously present at ringside," 15 U.S.C. § 6304 (2006), since a physician cannot render competent medical advice if he cannot communicate with the patient.

\textsuperscript{10} See discussion infra Part III.

\textsuperscript{11} See McCain & Nahigian, supra note 1, at 9-18.

\textsuperscript{12} See Model Penal Code § 211.1 (“(1) Simple Assault. A person is guilty of assault if he: (a) attempts to cause or purposely, knowingly or recklessly causes bodily injury to another; or (b) negligently causes bodily injury to another with a deadly weapon; or (c) attempts by physical menace to put another in fear of imminent serious bodily injury.”).

\textsuperscript{13} The U.S. military conducted a study of injuries commonly sustained by boxers, which “showed that sixty-eight percent of the boxers suffered head injuries, including intracranial injuries, concussions, intracranial hemorrhages, and skull fractures.” Kevin M. Walsh, \emph{Boxing: Regulating a Health Hazard}, 11 J. CONTEMP. HEALTH L. & POLY 63, 65 (citing Robert W. Enzenauer et al., \emph{Boxing-Related Injuries in the US Army, 1980 Through 1985}, 261 JAMA 1463, 1464 (1989)).
safer for the participants, or to ban the sport altogether. The calls for increased boxer safety, which seem to cycle in and out of the public’s consciousness, have once again been renewed based on a number of recent boxer injuries and the deaths of Sanchez and Johnson. This is especially true because many within the boxing industry believe that improved safety standards, or even stricter enforcement of the existing standards, could have prevented these injuries and deaths.

Many of the issues relating to boxing safety result from the economics of the sport and a lack of federal oversight that has led to a patchwork of state laws. There are two classes of fighter that fill the ranks of professional boxing: the premier fighters, whose names are widely recognizable, and the club fighters, who are unknown to the general public but make up the majority of the sport’s participants. While the premier fighters are highly compensated, the club fighters often receive little money. While the boxing industry depends on the club fighters, they are likely to be exploited both physically and financially. In part, this exploitation continues because of the

---

14 While the idea of banning boxing entirely has not received much attention in the United States, some countries have either banned the sport or are seriously considering the possibility. See Mark Barden, Boxing on the Ropes, BBC SPORT, Dec. 18, 2000, http://news.bbc.co.uk/sport2/hi/other_sports/1076244.stm.

15 For an illustration of the cycles of safety regulation, especially on the federal level, see McCain & Nahigian, supra note 1.

16 See Hauser, Fighter Safety, supra note 2. The circumstances surrounding the deaths of Leavander Johnson and Martin Sanchez will be discussed further in Part III, infra.

17 See Hauser, Fighter Safety, supra note 2.

18 McCain & Nahigian, supra note 1, at 8.

19 Two of the sport’s most popular fighters, Oscar De La Hoya and Floyd Mayweather, Jr., fought in Las Vegas on May 5, 2007. De La Hoya was expected to make at least $25 million and Mayweather was expected to make at least $10 million. Dan Rafael, Finally! The De La Hoya-Mayweather Bout Is Official, ESPN BOXING, Nov. 13, 2006, http://sports.espn.go.com/sports/boxing/news/story?id=2661028. However, because the fight made more money than expected, grossing over $150 million, De La Hoya actually received roughly $50 million while Mayweather earned around $20 million. Dan Rafael, De La Hoya ‘Ecstatic’ that Fight Was Richest Ever, ESPN BOXING, May 9, 2007, http://sports.espn.go.com/sports/boxing/news/story?id=2865549. In contrast, it is common for a club fighter to receive less than $200 for a fight. Mike Mosedale, In This Corner, CITY PAGES, Jan. 12, 2000, http://citypages.com/databank/21/997/article8360.asp.

20 One illustration of this dependence and exploitation is in the area of matchmaking (which is the process where it is determined who a fighter’s opponent will be). In order to build up a promising young fighter’s win-lose record, his handlers will schedule fights with opponents of considerably less skill. This allows the young fighter to showcase his skills without any real risk of losing. An example of this practice can be seen by looking at the record of Shad Howard. Shad Howard has a record of thirteen wins and ten losses. As of February 29, 2008 Howard had lost four of his last five fights with three of the losses coming against young contenders who have a
large amount of money that a boxing event can generate. Boxing is especially important to the city of Las Vegas, which has no major professional sports teams.21 For example, tickets for the Oscar De La Hoya/Floyd Mayweather Jr. fight at the MGM Grand Casino in Las Vegas were priced from $150 to $2000.22 Incredibly, almost all of the tickets were sold before the tickets officially went on sale.23 It was estimated that the live gate alone would exceed $19 million,24 and the event almost reached that goal by bringing in over $18 million.25 In fact, so much money was made that the two fighters are in negotiations for a rematch even though the first fight was extremely one-sided.26 With such large amounts of money coming into the state, it must be hard for state regulators to remain objective when deciding whether or not to allow a fight to take place.

Boxing has traditionally been regulated by the states through state athletic commissions.27 There are variations from state to state, but New York provides an example of how state athletic commissions are generally set up.28 New York State law requires the establishment of an athletic commission.29 The

23 The advance tickets were sold to sponsors, casinos, the fighters’ camps, and HBO. Id.
24 Id.
27 See infra Parts II, III.
body consists of three commissioners who are appointed by the governor and a support staff that includes a physician and a medical advisory board. The commission is responsible for setting safety standards and procedures involved in the regulation of boxing and wrestling activities. In addition, the commission administers all required licenses.

While the state commissions theoretically protect the safety of the participants, there is concern that in reality many state commissions have not adequately developed or enforced safety procedures. As a result, reformers have called for the federal government to regulate boxing in order to curtail some of the safety problems that have plagued the sport. These calls led Congress to use its Commerce Clause power to pass the Professional Boxing Safety Act (“PBSA”) of 1996. Through the PBSA, Congress tried to raise safety standards by requiring certain minimum safety measures for all professional boxing matches. The PBSA’s most important provision was aimed at increasing the standards of the less-regulated states by requiring all the state commissions to honor one another’s medical suspensions. In 2000, the PBSA was amended by the Muhammad Ali Boxing Reform Act (“MABRA”). The MABRA amended the PBSA but was primarily concerned with protecting professional boxers economically, and it left most of

30 Id.
31 Id.
32 Id.
33 See McCain & Nahigian, supra note 1, at 13-20.

Another catalyst for reform was the perceived corruption in boxing’s four major sanctioning organizations: the World Boxing Association, the World Boxing Council, the World Boxing Organization (“WBO”), and the International Boxing Federation (“IBF”). Each of these organizations has its own champions and ranking systems. Boxing Sanctioning Bodies, http://boxinggymms.com/sanctioningbodies.htm (last visited Feb. 7, 2008). There is widespread belief that the organizations are corrupt and do not accurately determine who the best boxers are. This position gained credence in 2001 when the WBO moved Darrin Morris two spots up in the rankings even though he had only fought once in the last three years, and had been deceased for over a year. Tim Graham, New WBO Division: Dead Weight, ESPN BOXING, http://espn.go.com/boxing/columns/ghanam/1097210.html (last visited Feb. 7, 2008).

36 Id. § 6306.
37 Id. §§ 6301-6313. Since the MABRA left most of the PBSA safety provisions unchanged, this Note will refer to the PBSA when discussing boxer safety regulation.
the safety provisions unchanged. The major problem with the PBSA and MABRA is that they rely heavily on the state athletic commissions. Thus, while these statutes do provide some federal oversight of the boxing industry, they do not significantly remedy the problem of state commissions failing to develop and enforce safety standards because of their reliance on these same commissions. As a result, even though the PBSA is generally seen as a positive step, there has been a large amount of debate about how effective the federal legislation has been in regulating the sport of boxing.

Because he does not believe that the previous acts have been effective in promoting boxer safety, Senator John McCain, who was instrumental in passing the PBSA and the MABRA, proposed the Professional Boxing Amendments Act (“PBAA”). The PBAA attempts to address the problem of inadequate oversight by providing for the formation of the United States Boxing Commission (“USBC”), an additional bureaucratic level providing federal oversight. The problem with this proposal is that, even with the establishment of the USBC, the state commissions would still be responsible for most of the oversight of the sport.

This Note will argue that the PBSA has failed to create adequate boxer safeguards due to the poor enforcement of its provisions by the state athletic commissions and the fact that a recent state court decision has effectively nullified the most important aspect of the Act. It will further argue that the only way to provide proper oversight of the sport is for the federal government to preempt the field of boxing regulation. This can be achieved by creating a federal boxing commission, similar to Senator McCain’s proposed USBC, that is capable of replacing the state commissions. Finally, this article will explain how Congress can constitutionally create such a commission.

Part II of this Note will provide a brief overview of the PBSA and the MABRA, and will discuss the role of the Association of Boxing Commissions (“ABC”) in regulating the

38 See id. §§ 6301-6313.
39 Id. § 6306 (placing primary responsibility for the establishment of numerous procedures on the Commissions).
40 See infra Part III.
42 Id.
43 Id. § 4(a).
44 See infra Part III.
sport. Part III will provide examples of how the PBSA has failed to achieve its goals by neglecting to ensure uniform safety standards. Part III will also concentrate on the case of Joe Mesi, a boxer who was suspended in one state but was then allowed to fight in other jurisdictions even though he was never cleared by the medical board that suspended him.45 The role that the state athletic commissioners played in enforcing the PBSA in the Mesi case will be discussed. Part IV will explain why the proposed PBAA will not produce the intended results of increased boxer safety. Finally, Part V will illustrate how federal oversight, independent of state athletic commissions, can improve boxer safety and will argue that Congress can constitutionally achieve federal oversight. This argument will be made by citing previous areas where Congress has preempted the states from acting by passing exhaustive regulations, and will explain how Congress can pass similar laws for the purpose of controlling boxing regulation.

II. OVERVIEW OF THE PBSA, MABRA, AND ABC

While it is not known when boxing first developed as a sport, there is evidence that it was popular in Rome prior to 500 A.D.46 Almost from the beginning of the sport, attempts were made to regulate it.47 In England “Broughton’s Rules” were developed to make the sport safer for the participants in 1743,48 and ongoing concern for boxer safety resulted in the establishment of the Marquis of Queensberry rules in 1867.49 In the United States, boxing was first legalized in California and Louisiana,50 but New York was the first state to regulate boxing with the passage of the Horton Act.51 As soon as boxing was legalized it was profitable for those who ran it, while the fighters who participated usually were doing so only to escape
Eventually the state athletic commission scheme was adopted by a majority of the states. In 1955, in *United States v. International Boxing Club of N.Y., Inc.*, the Supreme Court ruled that Congress could regulate professional boxing under the Commerce Clause, at least to the extent that the government could sue in civil court for violations of the Sherman Act by the boxing industry. However, at the time, the Court ruling did not lead Congress to adopt any federal standards concerning the boxing industry. Since there were no universal standards, safety conditions varied from state to state, depending on the regulations adopted by the state athletic commissions. This led to “forum shopping” by professional boxers; fighters who could not pass the medical tests of those states with strict medical requirements or who were medically suspended by a state would travel to states with less stringent regulations in order to fight. The wide variety of safety regulations and the ease with which fighters could evade them prompted congressional action to help protect professional boxers.

---

52 Id. at 10-11. For example, Jake La Motta, the former middleweight champion, described the difficulties of growing up poor during his early childhood. It is clear that his upbringing shaped the rest of his life and was something that he drew on during his boxing career. La Motta wrote:

> What I remember about the tenement as much as anything else is the smell. It’s impossible to describe the smell of a tenement to someone who’s never lived in one. You can’t just put your head in the door and sniff. You have to live there, day and night, summer and winter, so the smell gets a chance to sink into your soul. There’s all the dirt that the super never really manages to get clean even on the days when he does an hour’s work, and this dirt has a smell, gray and dry and, after you’ve smelt it long enough, suffocating.


54 See McCain & Nahigian, supra note 1, at 19-20.


56 The standards of the State Athletic Committees still vary widely. For example, only a small number of states require neurological exams as a condition of being licensed. See Association of Boxing Commissions, State Medical Requirements, http://www.canadianboxing.com/abcboxing_commission_medical_requirements.htm (last visited Feb. 14, 2008).

57 McCain & Nahigian, supra note 1, at 19. Senator McCain wrote that one of the aims of the PBSA was to “prohibit[] medically-suspended fighters from participating in boxing matches in other states and assure[] that states are aware that a fighter may be suspended in another state.” Id.
A. The PBSA

The PBSA was passed in 1996, pursuant to the authority recognized in International Boxing Club of N.Y., in an attempt to regulate the safety standards in professional boxing.58 Section 6302 of the Act states that two of the purposes of the Act are, “(1) to improve and expand the system of safety precautions that protects the welfare of professional boxers; and (2) to assist State boxing commissions to provide proper oversight for the professional boxing industry in the United States.”59 In order to achieve these large-scale goals, the Act concentrates on specific practices: for example, because Congress realized that one of the obstacles facing the boxing industry was accurate record-keeping, the Act provides that the state athletic commissions must establish procedures to evaluate the records, both medical and win-loss, of each boxer fighting in the state, and requires the state to prevent the fighter from fighting within the state if appropriate.60 Congress also addressed the issue of “forum shopping” in the Act by ordering state athletic commissions to establish “[p]rocedures to ensure that . . . no boxer is permitted to box while under suspension from any boxing commission due to—(A) a recent knockout or series of consecutive losses; (B) an injury, requirement for a medical procedure, or physician denial of certification . . . .”61 By requiring states to recognize one another’s medical suspensions, Congress hoped to end the practice in which boxers avoid medical suspensions in more restrictive states by traveling to other, less restrictive ones.62

59 Id. § 6302.
60 Id. § 6306.
61 Id. § 6306(a)(2)(A)-(B). The statute allows a suspended boxer to fight in another state if:

(1) for any reason other than those listed in subsection (a) of this section if such commission notifies in writing and consults with the designated official of the suspending State's boxing commission prior to the grant of approval for such individual to participate in that professional boxing match; or (2) if the boxer appeals to the Association of Boxing Commissions, and the Association of Boxing Commissions determines that the suspension of such boxer was without sufficient grounds, for an improper purpose, or not related to the health and safety of the boxer or the purposes of this chapter.

62 See McCain & Nahigian, supra note 1, at 19.
B. The MABRA

The MABRA, passed in 2000, amended the PBSA. It is aimed at protecting boxers financially and, aside from some minor textual changes, leaves the safety standards developed by the PBSA untouched. For example, under the MABRA, states still have to honor the medical suspensions of other states. The MABRA begins with a list of findings by Congress, including:

State officials are the proper regulators of professional boxing events, and must protect the welfare of professional boxers and serve the public interest by closely supervising boxing activity in their jurisdiction. State boxing commissions do not currently receive adequate information to determine whether boxers competing in their jurisdiction are being subjected to contract terms and business practices which may violate State regulations, or are onerous and confiscatory.

The finding that state officials are the proper regulators of the sport is the main reason why attempts to regulate boxers' safety have been unsuccessful up to this point. A pattern has emerged with respect to boxing regulation: the federal government finds that the states are not adequately protecting boxers, Congress passes legislation designed to provide protection, the state athletic commissions fail to follow the legislation, the boxers are once again left unprotected, and Congress responds by passing new legislation. However, adequate protection would be provided if the federal government simply took boxing regulation out of the states' hands. An examination of the effectiveness of the PBSA and the MABRA reveals that almost all of the major failures result from improper oversight from the state athletic commissions. One attempt to achieve proper oversight was the creation of the Association of Boxing Commissions.

C. The ABC

The creation of the ABC was an initial attempt to provide uniformity among the state athletic commissions, but

---

65 Id.
66 Id. § 6302 (emphasis added).
67 See infra Part III.C.
so far it has proven to be largely ineffective.\textsuperscript{68} The ABC is a non-profit organization whose membership includes the athletic commissioners of states that have formed commissions to regulate boxing in accordance with the PBSA of 1996.\textsuperscript{69} It is recognized by both the PBSA and the MABRA.\textsuperscript{70} According to its constitution, one of the main purposes of the ABC is \`{t}o promote continual improvement of, and for, the sport of professional boxing.\textsuperscript{71} The ABC also provides training for referees and other ringside personnel.\textsuperscript{72} Additionally, the PBSA allows a state that does not have an athletic commission to hold boxing events under the supervision of a neighboring commissioner or the ABC.\textsuperscript{73} While the ABC mainly has a supervisory role, Congress also envisioned that it would make suggestions for substantive reforms.\textsuperscript{74}

When enacting the MABRA, Congress provided general guidelines for some of the reforms and left the specific details in the hands of the ABC.\textsuperscript{75} This was probably based on the belief that the ABC, whose sole responsibility is the boxing industry, is better informed than Congress about the needs of the boxing industry. For example, when addressing the issue of

\textsuperscript{68} See infra Part III.A.


\textsuperscript{70} For example, § 6303 states:

(a) No person may arrange, promote, organize, produce, or fight in a professional boxing match held in a State that does not have a boxing commission unless the match is supervised by a boxing commission from another State and subject to the most recent version of the recommended regulatory guidelines certified and published by the Association of Boxing Commissions as well as any additional relevant professional boxing regulations and requirements of such other State.

(b) For the purpose of this Act, if no State commission is available to supervise a boxing match according to subsection (a), then—(1) the match may not be held unless it is supervised by an association of boxing commissions to which at least a majority of the States belong; and (2) any reporting or other requirement relating to a supervising commission allowed under this section shall be deemed to refer to the entity described in paragraph (1).


\textsuperscript{71} ABC Constitution, supra note 69, art. I, Sec. 1.3(A).

\textsuperscript{72} Id. art. I, sec. 1.3(E).


\textsuperscript{74} Id. § 6307c ("It is the sense of the Congress that sanctioning bodies and State boxing commissions should follow these ABC guidelines [for boxer rating standards].").

\textsuperscript{75} Id.
fighter rankings, the MABRA stated, “Within 2 years after the date of the enactment of the [MABRA], the [ABC] shall develop and shall approve by a vote of no less than a majority of its member state boxing commissioners, guidelines for objective and consistent written criteria for the ratings of professional boxers.” Even though Congress did not expressly state that the findings of the ABC would be adopted, it did express its willingness to allow ABC recommendations to influence decisions traditionally made by the individual state commissions. This is clear because the section concludes, “It is the sense of the Congress that sanctioning bodies and state boxing commissions should follow these ABC guidelines.”

The ABC has subsequently used its ability to influence policy to try to augment and improve safety standards in the sport of boxing. The ABC Constitution includes the following goals as part of its mission:

- To promote the uniformity of health and safety standards and other requirements pertaining to the conduct of professional boxing events.
- To promote standard reporting of boxing contests between members, including results, injury reports, suspensions and other medical information.
- To encourage adherence to, and enforcement of, applicable federal laws by each member of the ABC.

In trying to achieve these goals, the ABC has made a variety of recommendations. Unfortunately, the ABC, for a variety of reasons set forth below, has not had much success in promoting boxer safety.

Most of the recommendations made by the ABC came after extensive studies were undertaken by the group. The minutes of the ABC’s 2005 annual meeting include presentations by medical professionals regarding such issues as hand injuries, weigh-in procedures, and mouthpiece selection.

---

76 For the major controversy surrounding the rankings, see supra note 34 and accompanying text.
78 Id.
79 See ABC Constitution, supra note 69, art. I, sec. 1.3.
80 Id.
81 For example, the ABC has made recommendations for minimum safety standards. See Boxing Severity Index, http://www.canadianboxing.com/abcboxing_minimum_medical_qualifications_championship_bouts_BSI.htm (last visited Feb. 14, 2008).
82 It was determined that most boxers have significant hand injuries by the time they become professionals. There is a higher risk of hand injury for boxers with small hands. The ABC believes that this problem can be combated by estab-
Although not all of these presentations led to changes in procedure (for example, the ABC decided to keep weigh-ins at the day before the fight), the presentations illustrate the ABC's willingness to improve regulations that are already in place. However, despite the ABC's laudable goal of promoting boxer safety, it has no real authority to make any substantive changes. This is because the power of the ABC is constricted by its constitution. For example, section 2.3 of the ABC Constitution states, “After a thorough investigation by the ABC’s Disciplinary Committee of any complaint or allegation of wrongdoing, any member may be counseled, reprimanded, suspended or suffer loss of membership in the ABC upon a majority vote of the Board of Directors.”

Another limitation on the reform powers of the ABC is that it is composed mostly of state athletic commissioners who in many states are picked more on the basis of political affiliations than on knowledge of the boxing industry. Another reason the ABC has not had much success in promoting boxer safety is that there have recently been allegations that the ABC is not as insulated from the sanctioning organizations as it is supposed to be. Further, while the ABC has conducted several studies about issues concerning boxer safety, it has no enforcement powers to

83 Fights occur between opponents of the same weight. However, most weigh-ins occur over twenty-four hours before the fight takes place, which can result in a significant weight disparity. Id. For example, Jorge Arce, a popular fighter, routinely gains over fifteen pounds between the time that he weighs-in and the time that he fights. This is an especially large amount of weight to gain in a day considering that he fights at a 115-pound weight limit. Jon Wertheim, Packing a Big Punch, SPORTS ILLUSTRATED.COM, Dec. 8, 2006, http://sportsillustrated.cnn.com/2006/writers/the_bonus/12/06/arce/index.html. This weight disparity situation has become a serious safety issue. Boxer Joey Gamache has sued his 2000 opponent Arturo Gatti claiming that he suffered physical injury in the fight because Gatti weighed too much. See Associated Press, Boxer Injured in 2000 Fight Sues Gatti, http://www.msnbc.msn.com/id/11643791 (last visited Feb. 14, 2008).

84 See Canadianboxing.com, supra note 82.


86 See ABC Constitution, supra note 69, art. I, sec. 1.3(G).

87 Id. art. II, sec. 2.3.

88 See id.

89 See infra Part III.

ensure change.91 Finally, if the ABC did have enforcement powers, it is possible that it would cease to be as diligent in addressing the safety concerns facing boxing, since it would then have to calculate the cost of enforcement when making its findings.

The PBSA, MABRA, and ABC were important steps in promoting boxer safety. However, they have in large part not achieved their goals. Their impact has been limited because there are no universal safety standards, state courts are able to undermine their effectiveness, and many state athletic commissioners fail to enforce the existing regulations.

III. FAILURES OF THE PBSA AND MABRA

The PBSA and MABRA were passed for the purpose of increasing professional boxer safety, in part by maintaining universal safety standards across multiple jurisdictions. Despite this federal legislation, there continues to be significant differences in safety regulations based on where the fight is being held. This section first illustrates that there are no uniform safety standards in the sport of boxing. It then explains that this failure is due to the fact that the state athletic commissions have not been active enough in their role as enforcers of the PBSA, and that this problem will get worse as a result of a recent Nevada ruling.

A. Failure to Achieve Uniform Safety Standards

Despite the passage of the PBSA and the MABRA, as well as the creation of the ABC, medical standards continue to vary greatly from state to state. These varying standards manifest themselves before the boxer even steps in the ring because the tests that a boxer must pass in order to be licensed to box differ greatly depending on which state is issuing the license.92 For example, the Pennsylvania State Athletic Commission only requires a boxer to undergo an annual physical examination and to provide negative HIV/AIDS and Hepatitis C test results (which cannot be older than sixty days)

91 See ABC Constitution, supra note 69, art. I, sec. 1.3.
92 See supra note 56.
in order to be licensed. In contrast, the New York State Athletic Commission, in addition to requiring annual physical examinations and HIV/AIDS and Hepatitis tests, requires a boxer to pass a wide variety of tests, which include an annual EKG, eye exam, and neurological tests along with an MRI exam at least once every three years. Admittedly, this situation, in which state standards differ vastly, was foreseeable at the time of the passage of the PBSA (which was not meant to instantly create uniform safety standards across the states). However, there is now an example of the PBSA failing to prevent a fighter from fighting in one jurisdiction after being medically suspended in another, which is one of the most important purposes for which it was designed.

B. The Case of Joe Mesi

The failure of the PBSA to protect boxer safety can best be illustrated by the circumstances surrounding the suspension and reinstatement of Joe Mesi. Mesi is a popular heavyweight boxer with a professional record of thirty-six wins and no losses. While he has yet to fight a top ranked heavyweight, he is considered a legitimate prospect. However, over the last

---


94 An electrocardiogram measures the electric signals that travel through the human heart. It is used to detect irregularities in the heart rhythm or heart structure. See MayoClinic.com, Electrocardiogram: Tracing the Electrical Path Through the Heart (June 30, 2006), http://www.mayoclinic.com/health/electrocardiogram/HB00014.

95 There are a variety of eye tests that make sure that sight is not impaired. See MayoClinic.com, Eye Exams: What to Expect (Oct. 20, 2006), http://www.mayoclinic.com/health/eye-exams/MC00021.

96 Magnetic resonance imaging uses a magnetic field and radio waves to create cross-section images of the head and body. These images can be used to diagnose a wide variety of neurological disorders. See MayoClinic.com, MRI: Viewing Your Brain (Nov. 30, 2006), http://www.mayoclinic.com/health/mri/SM00035.


98 “Nothing in this chapter shall prohibit a State from adopting or enforcing supplemental or more stringent laws or regulations not inconsistent with this chapter, or criminal, civil, or administrative fines for violations of such laws or regulations.” 15 U.S.C. § 6313 (2006).

99 The state commissions must create “[p]rocedures to ensure that, except as provided in subsection (b), no boxer is permitted to box while under suspension from any boxing commission due to . . . a recent knockout or series of consecutive losses.” 15 Id. § 6306(a)(2).


101 Federal Judge Rules Mesi’s Medical Suspension Expired, supra note 45.
few years he has received more attention for his fight outside of the ring with the Nevada Athletic Commission.\textsuperscript{102} As a result of his legal campaign to resume his professional career, Mesi is currently fighting even though he was never cleared from a prior medical suspension.\textsuperscript{103} As indicated above, one of the key protections for boxers that the PBSA provides is that each state athletic commission must establish “[p]rocedures to ensure that . . . no boxer is permitted to box while under suspension from any boxing commission due to—(A) a recent knockout or series of consecutive losses.”\textsuperscript{104} The statute goes on to state that each state athletic commission must not allow a boxer who is suspended in another state to fight if the suspension is because of “(B) an injury, requirement for a medical procedure, or physician denial of certification; (C) failure of a drug test; [or] (D) the use of false aliases, or falsifying, or attempting to falsify, official identification cards or documents.”\textsuperscript{105} Subsection 6306(b) allows a fighter to fight if he can show that the suspension in the other state is no longer warranted.\textsuperscript{106} Subsection (b) does not apply to medical suspensions from other states. In a decision that nullifies the key safety provision of the PBSA, which prevents “forum shopping,” Mesi is currently being allowed to fight in other jurisdictions without being cleared by the Nevada medical board that suspended him.\textsuperscript{107}

\textsuperscript{102} Id.
\textsuperscript{103} Id.
\textsuperscript{105} Id. § 6306(a)(2)(B)-(D) (emphasis added).
\textsuperscript{106} Subsection 6306(b) provides:

(b) Suspension in another State. A boxing commission may allow a boxer who is under suspension in any State to participate in a professional boxing match—(1) for any reason other than those listed in subsection (a) if such commission notifies in writing and consults with the designated official of the suspending State’s boxing commission prior to the grant of approval for such individual to participate in that professional boxing match; or (2) if the boxer appeals to the Association of Boxing Commissions, and the Association of Boxing Commissions determines that the suspension of such boxer was without sufficient grounds, for an improper purpose, or not related to the health and safety of the boxer for the purposes of this act.

\textsuperscript{107} Federal Judge Rules Mesi’s Medical Suspension Expired, supra note 45.
Mesi suffered two subdural hematomas\(^{108}\) during a fight with Vassiliy Jirov that took place on March 13, 2004 in Las Vegas, Nevada.\(^{109}\) The Nevada Administrative Code states, “The Commission will not issue or renew a license to engage in unarmed combat to an applicant . . . who has suffered cerebral hemorrhage.”\(^{110}\) When the Nevada Athletic Commissioners learned of Mesi’s hematomas, they followed Nevada procedure and suspended him indefinitely due to his injuries.\(^{111}\) Mesi challenged the Nevada Athletic Commission’s ruling in a Nevada (Clark County) court, and in December of 2005, his suspension was overturned.\(^{112}\) The court found that the state of Nevada could not suspend a fighter for a period longer than he was licensed for.\(^{113}\) The court also ruled that a boxer has a property right in his license, which means that it cannot be rescinded by the state without due process.\(^{114}\) Notably, the judge did not address the issue of whether or not Mesi was healthy enough to fight; he merely ruled that the Nevada suspension could not outlive the license.\(^{115}\) This result was important, because under the PBSA once the Nevada medical suspension was lifted other jurisdictions were free to license Mesi.\(^{116}\) This ruling also effectively ties the hands of the Nevada State Athletic Commission because the Commission can no longer suspend fighters for as long as it would like without making severe concessions such as the extension of the licensing period. This is because under the Clark County ruling, in order to increase the length of a suspension, the length of the licensing period must be correspondingly

\(^{108}\) A hematoma in the head “occurs when a blood vessel ruptures . . . between your skull and your brain. The collection of blood (hematoma) compresses your brain tissue. . . . Treating an intracranial hematoma often requires surgery to remove the blood.” MayoClinic.com, Intracranial Hematoma (June 28, 2007), http://www.mayoclinic.com/health/intracranial-hematoma/DS00330.


\(^{112}\) Federal Judge Rules Mesi’s Medical Suspension Expired, supra note 45.

\(^{113}\) Id.

\(^{114}\) Id.

\(^{115}\) Id.

\(^{116}\) Allowing Mesi to fight in jurisdictions outside Nevada after the court’s ruling was permissible because he was no longer under a medical suspension in any jurisdiction. 15 U.S.C. § 6306 (2006).
increased (fighters in Nevada, and in most other states, are currently licensed through the calendar year). Extending the licensing periods creates serious issues of boxer safety because it prevents the Commission from being able to check fighters regularly.

However, the most important aspect of the Clark County ruling is its impact on federal legislation. Keith Kizer, the deputy attorney general who handled the Mesi case, stated that the ruling “takes a lot of teeth out of the Ali Act . . . . It takes away Congress’s ability to prevent a fighter who has not proven he’s fit to fight from going elsewhere. But the adverse impact is on the federal law. It’s not on the state law or on the commission.” This is a valid assessment of the situation. Mesi cannot get a license to fight in Nevada because Nevada has stricter regulations than most other states. However, since Nevada is only withholding a license, rather than issuing a medical suspension, Mesi is free to fight in states that have lower licensing standards. The Clark County ruling is especially troubling because Nevada is one of the only states that require fighters to pass an MRI exam, and thus once a fighter with a head injury serves his Nevada suspension, it is unlikely that any other state will uncover the injury. While it is true that Mesi was prevented from fighting for over a year in Nevada, he still fought in several other states without ever having been cleared by Nevada. Mesi’s situation is exactly the type of forum-shopping scenario that Congress intended to eliminate.

Since his medical suspension was lifted, Mesi has fought seven times (none of these fights occurred in Nevada). One

---


118 For example, if the licensing period were extended to five years a suspension could last for five years. However, a boxer could be healthy when he applies for his license, suffer an undetected injury early in the licensing period, and continue to fight for five years without having to renew his license.


120 See State Medical Requirements, supra note 56.

121 This has been cited as an example of federal safety regulations having the potential for success. See Michael F. Jurek, Janitor or Savior: The Role of Congress in Professional Boxing Reform, 67 OHIO ST. L.J. 1187, 1204-06 (2006).

hopes that Mesi will finish his career without suffering any adverse health consequences as a result of his prior injuries. However, whether or not Mesi is injured in the ring, the precedent of disregarding safety procedures is likely to cause harm to future boxers.

C. The Role of State Athletic Commissioners

Even if the Mesi case had not removed some of the statutory protections provided by the PBSA, there would still be a need for additional federal legislation, because the biggest reason that the safety provisions of the PBSA have failed is that the state athletic commissions are not enforcing them. For example, the PBSA, in a section that was not amended by the MABRA, states:

No person may arrange, promote, organize, produce, or fight in a professional boxing match without meeting each of the following requirements or an alternative requirement in effect under regulations of a boxing commission that provides equivalent protection of the health and safety of boxers: (1) A physical examination of each boxer by a physician certifying whether or not the boxer is physically fit to safely compete, copies of which must be provided to the boxing commission. (2) Except as otherwise expressly provided under regulation of a boxing commission promulgated subsequent to the enactment of this Act [enacted Oct. 9, 1996], an ambulance or medical personnel with appropriate resuscitation equipment continuously present on site. (3) A physician continuously present at ringside.

While in theory this provision of the Act provides adequate protection for boxers by requiring review of their medical histories and providing medical supervision during a fight, there are many examples of the provision failing because falsified medical records went undetected due to improper state commission examination or only superficial medical supervision was provided during fights.

An example of the PBSA’s failure to provide adequate medical oversight is revealed by the circumstances surrounding appears that all of his fights have been sanctioned by the Association of Boxing Commissioners. Id. Mesi has also fought once in Canada, see id., which raises interesting questions about international safety standards, which are beyond the scope of this Note.

See Hauser, Lamon Brewster, supra note 119.

124 15 U.S.C. § 6304 (2006). The provision also requires, “(4) Health insurance for each boxer to provide medical coverage for any injuries sustained in the match.” Id.

125 Hauser, Fighter Safety, supra note 2.
Tommy Morrison’s retirement and subsequent return to boxing. Morrison reached the height of his boxing career in 1993 when he earned a decision victory over George Foreman to win the World Boxing Organization heavyweight title. In 1996, the Nevada Athletic Commission suspended Morrison after a pre-fight blood test revealed that he was HIV positive. Morrison fought one more time in 1996—in Japan, which at the time did not have any rules which prevented an HIV positive fighter from fighting—and then retired. However eleven years later Morrison began his boxing comeback in West Virginia with a knockout victory over John Castle.

The initial licensing of Morrison in West Virginia occurred in an unusual manner. West Virginia does not normally require fighters to take an HIV test in order to be licensed; however, because of concern over Morrison’s HIV status he was required to submit one. Since the West Virginia Commission does not normally require a blood test, it accepted the results of a test that was conducted in Arizona; the test indicated that Tommy Morrison was not HIV positive. While Morrison won his West Virginia fight, problems arose when he tried to get licensed in Texas for his next bout. After initial reports that he had been granted a license, Texas state officials informed the press that Morrison would not be granted a license due to incomplete lab results. Shortly after the announcement, Morrison withdrew his application for a Texas license. Since then, he won a professional boxing match in Mexico, and he has fought in a mixed


127 Id.

128 Id.


130 See Rafael, supra note 126.

131 Id.


133 Id.


martial arts event. Mixed martial arts events are also generally regulated by state athletic commissions, but Morrison’s fight was unsanctioned by the Arizona State Commission. In the week leading up to Morrison’s mixed martial arts debut, his former agent came forward with allegations that Morrison was HIV positive and had only passed the Arizona blood test through the use of fraud. Whatever the truth about Morrison’s HIV status is, it is clear that there are serious questions which should have been resolved before Morrison was allowed to return to the ring.

Unfortunately, examples of poor medical oversight are prevalent in many states. In Ohio, Lamon Brewster suffered a detached retina during a fight with Sergei Liakhovich on April 1, 2006. Despite reports that Brewster had undergone laser eye surgery weeks before the fight, he passed his eye exam and there is no mention of the laser surgery on the medical report. However, a previous eye examination had indicated that Brewster was having vision problems and had previously undergone eye surgery. While Ohio only requires fighters to undergo an eye exam once a year in order to be eligible to fight, failure to comply with this rule was not an issue for this fight because Brewster had had an eye exam eight days before the fight took place. This means that either

---


138 See Hauser, Lamon Brewster, supra note 119.

The retina is the light-sensitive tissue that lies smoothly against the inside back wall of your eye and sends messages to your brain through your optic nerve. . . . Retinal detachment occurs when the retina separates from the choroid. . . . Retinal detachment is a medical emergency, and time is critical. Unless the detached retina is promptly surgically reattached, this condition can cause permanent loss of vision in the affected eye.


139 See Hauser, Lamon Brewster, supra note 119.

140 Id.


143 See Hauser, Lamon Brewster, supra note 119.
the Commission knew about the surgery but allowed the fight to take place anyway to generate revenue for the state, or the Commission was fooled by a fraudulent medical report. No matter which scenario occurred, the Ohio State Athletic Commission failed to fulfill its safety obligations by allowing an injured fighter to fight.\textsuperscript{144}

In some cases lax enforcement of safety regulations has arguably lead to tragedy. For example, on July 1, 2005, Martin Sanchez died after suffering a brain bleed during his fight with Rustam Nugaev in Nevada.\textsuperscript{145} Sanchez was a twenty-six-year-old fighter with an 18 and 8 win-loss record who was fighting outside of Mexico for the first time.\textsuperscript{146} Examples of poor oversight can be seen before, during, and after his fight and injury. To start, Sanchez was granted a license by the Nevada State Athletic Commission despite the fact that his application to fight in Nevada lists his height as 5 feet 9 inches, while his pre-fight physical examination lists his height as 6 feet 1 inch.\textsuperscript{147} Even a superficial review of the application and physical would have alerted a diligent official to the possibility that there were serious questions about the validity of the application. The lack of adequate oversight continued during the fight. After Sanchez was hit with a blow to the back of the head in the fourth round, he did not throw punches with the same authority that he had displayed prior to the hit, and he appeared to be moving much more slowly.\textsuperscript{148} Although medical personnel were present at the fight, nobody examined Sanchez between rounds.\textsuperscript{149} Moreover, even if someone had examined

\textsuperscript{144} It is often the case that the state commission needs to protect the fighter from himself. While it may seem strange that a fighter would risk his health by fighting with a pre-existing injury, it happens all the time for a variety of reasons. For example, “The decision to proceed with Brewster-Liakhovich is believed to have been made in part because of Lamon’s desire to fulfill his contractual obligations to Don King so he could move to a new promoter.” See id.

\textsuperscript{145} Hauser, \textit{Fighter Safety}, supra note 2.

\textsuperscript{146} Id. There is growing concern that Mexican fighters are getting around the safety standards required by law by providing medical records from Mexico that cannot be verified. Id. This phenomenon involves not only club fighters; in 1997 Marco Antonio Barrera, one of the highest profile boxers in the sport, underwent brain surgery in Mexico in an attempt to hide his surgery from the state athletic commissions. See Thomas Hauser, \textit{Boxing’s Medical Mess}, \textsc{SecondsOut}, May 27, 2004, http://www.secondsout.com/usa/colhauser.cfm?ccs=208&cs=13484. Although it is unclear if this is where he learned to cover his tracks, before becoming a professional boxer Barrera was a law student in Mexico. See HBO.com, Bio: Marco Antonio Barrera, http://www.hbo.com/boxing/fighters/barrera_marcobio.html (last visited Feb. 19, 2008).

\textsuperscript{147} See Hauser, \textit{Fighters Are Dying}, supra note 141.

\textsuperscript{148} Hauser, \textit{Fighter Safety}, supra note 2.

\textsuperscript{149} Id.
him, communication barriers may have hindered any medical treatment as neither of the doctors present at the fight spoke Spanish and Sanchez did not speak English.\footnote{Id.} The fight was allowed to continue until Sanchez was knocked down and counted out in the ninth round.\footnote{Id.} After the fight Sanchez was left in his dressing room without medical supervision even though he appeared to have trouble walking under his own power.\footnote{Id.} While he was in his dressing room he had a seizure.\footnote{Id.} He was then taken to the hospital where he underwent brain surgery.\footnote{Id.} Unfortunately his injury was too severe and he died the next morning.\footnote{Id.}

After the fight the Nevada State Athletic Commission asked then deputy attorney general of Nevada Keith Kizer\footnote{Kizer later became a Nevada State Athletic Commissioner. Kevin Iole, 
\textit{Fighter Safety Is Kizer’s Priority}, \textit{LAS VEGAS REV.-J.}, Apr. 1, 2006, \textit{available at} \url{http://www.reviewjournal.com/lvrj_home/2006/Apr-01-Sat-2006/sports/6656180.html}.} to review the facts surrounding Sanchez’s death.\footnote{Hauser, \textit{Fighter Safety, supra note 2}.} His conclusion was, “I don’t have any recommendation. I don’t see any way to improve.”\footnote{Id.} This statement seems to ignore the fact that the Commission could have made sure that the licensing application was truthful and that medical personnel could communicate with the fighters. One Nevada State Athletic Commissioner, long time safety advocate Dr. Flip Homansky, disagreed with Keith Kizer’s conclusion.\footnote{See Hauser, \textit{Fighters Are Dying, supra note 141}.} Dr. Homansky went public with the fact that there was a discrepancy between the height listed on Sanchez’s application and the height recorded during his pre-fight physical, and noted that this discrepancy should have been discovered before the fight.\footnote{Id.} Rather than being commended for doing his job, Dr. Homansky was removed from his position.\footnote{Id.} This suggests that the Nevada State Athletic Commission wants to generate as much money as possible from the fights within the state without spending the money and time required to properly protect the fighters.
Another tragedy occurred in Las Vegas the night of September 17, 2005, when Leavander Johnson defended his IBF lightweight\textsuperscript{162} title against Jesus Chavez.\textsuperscript{163} Unlike the fight involving Sanchez, Johnson was checked by medical personnel between rounds, the fight was stopped before he was counted out, and he was taken to the hospital immediately after showing neurological systems.\textsuperscript{164} However, despite these precautions Johnson had to undergo surgery to relieve swelling in his brain that was a result of the fight and died five days later.\textsuperscript{165} The way the fight was handled prompted Jim Lampley, HBO's blow-by-blow commentator, to state, “Some ring tragedies are avoidable; others aren't. If you can't live with Leavander Johnson's death, you ban boxing.”\textsuperscript{166} However, despite the belief by some that everything was done correctly, there are rumors that Johnson was knocked unconscious during his training for the fight; if this is true it should have resulted in a medical suspension,\textsuperscript{167} since in Nevada when a fighter is knocked out, either in training or in a fight, he is automatically suspended.\textsuperscript{168} Also, regardless of whether or not Johnson had been knocked out during training, the referee should have been aware that despite winning his last fight against Stefano Zoff he had received a great number of punches in that match.\textsuperscript{169} This is where having experienced boxing people involved in the fights can be invaluable. Good referees know the styles, reputations, and histories of the fighters

\textsuperscript{162} There are seventeen weight classes in professional boxing. While the different sanctioning organizations have different names for the weight classes, the weight requirements are roughly the same. The lowest weight class requires participants to be below 105 pounds, while the highest weight class consists of boxers who weigh at least 200 pounds. See Ring of Dreams, Amateur and Professional Boxing Weight Classes, http://www.ringofdreams.com/articles/boxingweightclasses.shtml (last visited Feb. 19, 2008).

\textsuperscript{163} Hauser, Fighter Safety, supra note 2.

\textsuperscript{164} Id.

\textsuperscript{165} Id.

\textsuperscript{166} Id.

\textsuperscript{167} Hauser, Fighters Are Dying, supra note 141.

\textsuperscript{168} Id. The issue of boxers being knocked out in training has become the subject of much debate. This is because, while they are not supposed to fight after being knocked out, there is no oversight in the gyms. This creates a situation where the only people who know if a fighter has been knocked out are the people on the fighter's training team, and they have a financial interest in making sure that the fight takes place. Thomas Hauser has pointed out that if Johnson was knocked out during training, then his handlers lied on his application to fight, which may make them criminally liable for his death. Id.

\textsuperscript{169} Id.
involved in the fights, allowing them to anticipate potential problems so that they can be addressed before they get out of hand. An informed referee would have known that Johnson received a lot of punishment in his last fight, would have seen that he was not responding well to the punches he was taking, and might have stopped the fight sooner. These examples of failures to properly protect boxers before, during, and after their fights highlight the need for new regulations.

D. Appointing State Athletic Commissioners

Referees are not the only group involved with the boxing industry that benefits from experience. Outside of the failure to enforce medical standards, many of the problems with safety oversight occur because a large number of state athletic commissioners have no experience in the boxing business. In Nevada, after Governor Kenny Guinn removed Dr. Homansky from the Commission for publicly raising the issue of the discrepancies in Martin Sanchez’s medical records, he appointed as a replacement T.J. Day, a man with no experience in the boxing industry. Day is a businessman and large contributor to the Nevada Republican Party. When asked about his lack of experience in the boxing industry (a large concern given the recent deaths in Nevada), he stated, “Boxing has a major financial impact on the state, and I’m a major financial man in the state . . . . If I need some help from doctors and the safety people, I can find those people, but it’s a lot harder to find a qualified businessman.” He also

170 For example, before one event, New York State Commissioner Ron Scott Stevens talked to the referees while referring to the specific individual fighters that were competing that night. He then said:

These guys are coming in as opponents. Give them a fair chance to win. But if they’re getting hurt, do what you have to do sooner rather than later. The fans are entitled to see good honest boxing and that includes guys getting hit.

But I don’t want anyone taking unnecessary punishment.


171 See id.

172 See id.; supra notes 145-161 and accompanying text.

173 Hauser, Fighters Are Dying, supra note 141.

174 Id.

175 Id.

acknowledged that his close relationship with Governor Guinn was a motivating factor behind his appointment.\textsuperscript{177}

This phenomenon, in which political connections rather than experience in the boxing industry determines who participates in reform efforts, can also be seen outside of the Commission itself.\textsuperscript{178} After the deaths of Martin Sanchez and Leavander Johnson in Nevada, a five-person safety committee was formed to determine if any procedural changes should be made to Nevada’s laws regulating professional boxing.\textsuperscript{179} The committee was chaired by Sig Rogich, a Republican consultant and fundraiser.\textsuperscript{180} Before the committee’s suggestions were released, it was noted that “Boxing regulators have criticized the safety panel because every member has links to the boxing commissioners in Nevada or has political ties to the state’s Republican governor, Kenny Guinn, who appoints the commissioners.”\textsuperscript{181} Once released, the committee suggestions were not taken seriously by those in the boxing industry because they failed to address the substantive problems of the sport.\textsuperscript{182} Also, some suggestions seemed like they were aimed at protecting the financial interests of the state rather than the health of the participants. For example, one suggestion, which seems like it would hurt the fighters rather than help them, was that in order to prevent boxers from dehydrating themselves to make weight for a fight, boxers who fail to make weight should be fined ten percent of their fight purse.\textsuperscript{183} Rather than trying to protect fighters from dehydrating themselves before a fight, this suggestion seems geared towards “punish[ing] a fighter who endangers a fight.”\textsuperscript{184}

\textsuperscript{177} An article stated, “Day said his close relationship with Guinn was the motivating factor behind his appointment. According to campaign records, he contributed $9,750 to Guinn’s 2002 campaign . . . .” \textit{Id.}

\textsuperscript{178} \textit{Id.}

\textsuperscript{179} \textit{Id.}

\textsuperscript{180} \textit{Id.}

\textsuperscript{181} \textit{Id.}


\textsuperscript{183} \textit{Id.}

\textsuperscript{184} \textit{Id.} In an attempt to gain an advantage many boxers lose a drastic amount of weight in a short period of time in order to fight an opponent who they believe is naturally smaller. For example, while Ricky Hatton fights at 140 pounds, his weight between fights is somewhere between 170 and 180 pounds. See HBO.com, Juan Urango vs. Ricky Hatton: Compubox Post Fight Analysis, http://www.hbo.com/boxing/events/2007/0120_hatton_urango/columns/compubox_post.html (last visited Feb. 13, 2008). While the Nevada Commission claims to be concerned with the fighters’ health, it has
Nevada is not the only state that has a history of using political patronage to determine the makeup of the state athletic commission. In New York, from the mid 1990s until 2003, the State Athletic Commission was filled with political appointees who were corrupt or disinterested. Bernard Kerik served as the chairman of the Commission and was notorious for not even attending fights within the state. Kerik is but one example of a commissioner apparently appointed based on political connections rather than experience in boxing. However, this changed in 2003 when Kerik resigned and was replaced by Ron Scott Stevens. Stevens has a long history in boxing and has greatly improved the way boxing events are conducted in New York by taking an active role in overseeing how they are run. Stevens has also increased the number of boxing events held in New York since taking over. For example, there were seventeen events statewide in 2003, compared to fifteen events in just the first six months of 2005. Increasing the number of events held in a state is important for improving boxing safety because more events allow officials to gain more experience. Stevens also showed great concern for boxer safety shown that it is more interested in making sure that the fight takes place. For example, when Jose Luis Castillo failed to make weight for his fight with Diego Corrales, the first thing that Castillo did, although already clearly dehydrated, was to go into a sauna to try and lose more weight. It was only after this attempt failed that Castillo was suspended from fighting for the rest of the year. See Matt Wells, The Return of Jose Luis Castillo, SECONDSOUT, http://www.secondsout.com/USA/news.cfm?css=229&cs=21141 (last visited Feb. 13, 2008).


186 Bernard Kerik was the New York City police commissioner from 2001 until the end of 2002. In that position he gained national acclaim for his response to the September 11th attacks. He was later nominated for the position of Director of Homeland Security by President George W. Bush, but withdrew his name after information surfaced that he had hired an illegal immigrant as a nanny. Mr. Kerik moved into the private sector where he worked in consulting, but controversy continued to follow him. In 2006 he pled guilty to two misdemeanor ethical violations, which resulted from his business practices. See William K. Rashbaum, Testimony by Giuliani Indicates He Was Briefed on Kerik in ’00, N.Y. TIMES, Mar. 30, 2007, at A1. Most importantly, there was never any inquiry into his qualifications to be a commissioner. Thomas Hauser, Joe Dwyer and the New York State Athletic Commission, http://www.secondsout.com/usa/collhauser.cfm?css=208&cs=3951 (last visited March 24, 2008).

187 Hauser, Ron Scott Stevens, supra note 170.

188 Before an event he states that that the fights are live and, “It’s not a play. It happens once and won’t come back again tomorrow night, so everything has to be done right the first time.” Id.

189 Id.
when, in two different cases, he placed popular boxers, Evander Holyfield and Al Cole, on medical suspensions. Despite causing the state to lose the money that would have come into New York from future fights involving Holyfield and Cole (from licensing fees and tourism), Stevens acted correctly by placing their safety above the financial interests of the state.

As illustrated by the examples above, most commissioners are selected based on political affiliations as opposed to qualifications. As long as state commissioners are chosen in this way, any federal legislation that must be enforced at the state level will fail because the commissioners will be more concerned with pleasing the governor that put them in their position than with enforcing the safety precautions established by Congress. Although there are examples of competent commissioners, such as Ron Scott Stevens, unfortunately they are the exceptions.

IV. PROPOSED CHANGES TO MABRA

Because of the problems with safety enforcement and because he does not believe that the current regulations are adequate, Senator McCain has proposed amending the PBSA through the Professional Boxing Amendments Act of 2005 (“PBAA”). Although the Senate passed the legislation, the House did not vote on it, and it did not become law. The major problem with the PBAA is that it creates a new level of federal bureaucracy, but effectively leaves the enforcement of its provisions in the hands of the state commissions by putting them in charge of the day-to-day oversight of the boxing industry. The PBAA effectively requires lawmakers to expend a great amount of time and money on a program that is not likely to produce results because of its continued reliance on the state athletic commissions.

190 Despite Commissioner Stevens’s efforts, Evander Holyfield’s suspension was successfully appealed, and he has since fought outside of New York. BoxRec.com, Evander Holyfield, http://www.boxrec.com/list_bouts.php?human_id=000499&cat=boxer (last visited Mar. 24, 2008). Al Cole traveled to Nevada where he worked as the sparring partner of Samuel Peter, who has a reputation as one of the hardest punchers in the heavyweight division. Hauser, Fighter Safety, supra note 2.

191 Hauser, Ron Scott Stevens, supra note 170.


A. Reforms

The main reform proposed in the PBAA is the creation of a new bureaucratic agency, the United States Boxing Commission (“USBC”).\textsuperscript{195} At first glance, the proposed creation of the USBC appears to address many of the problems that exist in boxing today.\textsuperscript{196} The PBAA states that no person can promote or arrange a professional boxing match “within the United States unless the match—(1) is approved by the [USBC]; and (2) is held in a State, or on tribal land of a tribal organization, that regulates professional boxing matches in accordance with standards and criteria established by the [USBC].”\textsuperscript{197} This would finally result in uniform safety standards across the United States. The USBC would “consist of 3 members appointed by the President, by and with the advice and consent of the Senate.”\textsuperscript{198} In order to combat the problem of commissioners with no experience, the members of the USBC must “ha[ve] extensive experience in professional boxing activities or in a field directly related to professional sports . . . .”\textsuperscript{199} In addition, the members of the Commission cannot

\begin{itemize}
\item[(i)] be engaged as a professional boxer, boxing promoter, agent, fight manager, matchmaker, referee, judge, or in any other capacity in the conduct of the business of professional boxing;
\item[(ii)] have any pecuniary interest in the earnings of any boxer or the proceeds or outcome of any boxing match;
\item[(iii)] serve as a member of a boxing commission.\textsuperscript{200}
\end{itemize}

These provisions would ensure that the problems of conflicting loyalties that plague state athletic commissions would not occur at the federal level. However, under the PBAA as currently written, it is unlikely that many substantive issues will come to the attention of the USBC.

B. Problems with the Suggested Reforms

The problem with the USBC is that it would still rely on the state athletic commissions to oversee the day-to-day

\begin{itemize}
\item[195] See id. § 21.
\item[197] Id. § 5(a) (amending PBSA sec. 4(a)).
\item[198] Id. § 21(a) (adding PBSA sec. 202).
\item[199] Id. (adding PBSA sec. 202(b)(2)(A)(i)).
\item[200] Id. (adding PBSA sec. 202(b)(2)(C)).
\end{itemize}
running of the sport.201 The PBAA is set up so that many of the boxing matches will have presumed approval from the Commission: “[T]he [USBC] shall be presumed to have approved any match . . . .”202 The matches that will not receive presumed approval include, “a match with respect to which the Commission has been informed of an alleged violation of this Act and with respect to which it has notified the supervising boxing commission that it does not approve . . . .”203 The PBAA is also concerned with matches that are advertised as championships; it therefore reserves the presumed approval from “a match advertised to the public as a championship match” or “a match scheduled for 10 rounds or more . . . .”204 Finally, the PBAA addresses fighter safety and withholds presumed approval from “a match in which 1 of the boxers has—(i) suffered 10 consecutive defeats in professional boxing matches; or (ii) has been knocked out 5 consecutive times in professional boxing matches.”205 The end result is that for almost every fight, approval by the USBC will be presumed. The fact that for the majority of fights the approval of the USBC will be presumed means that the majority of fights will continue to be regulated by the state athletic commissioners alone, which have already proven themselves to be ineffective.206 It is useless to expend the time and money to create another level of bureaucracy if it is not clear that it will address the problems that it was created to address.

V. **HOW OVERSIGHT OF STATE ATHLETIC COMMISSIONERS CAN ACHIEVE BOXER SAFETY**

In most states the athletic commissions do not do an adequate job of enforcing the safety standards required by the PBSA and the MABRA.207 The state commissioners fail to enforce the law because they are more concerned with pleasing

---


203 Id. (amending PBSA sec. 4(b)(1)(A)).

204 Id. (amending PBSA sec. 4(b)(1)(B)-(C)).

205 Id. (amending PBSA sec. 4(b)(1)(D)).

206 See supra Part III.

207 Id.
the governors that control their appointments than protecting the interests of boxers with no political power.\textsuperscript{208} Therefore, if federal legislation is going to be effective in regulating the boxing industry, the legislation must not rely on the state athletic commissions for enforcement.

Federal enforcement of boxing regulations would likely produce better results than state enforcement because under a scheme of federal enforcement, there would be a greater level of disconnect between the regulating agency and the economic consequences which result from a fight.\textsuperscript{209} Under the current system, a state must police itself in circumstances where strictly enforcing regulations can mean millions of dollars in lost revenue for the state.\textsuperscript{210} Given these circumstances, it was, and is, not realistic to expect the states to follow the letter of the law, and the poor record of the state commissions (such as Nevada's) has shown that indeed the law has not been followed.\textsuperscript{211} In addition to financial independence, federal enforcement would also create a universal standard of enforcement. A universal federal standard would greatly improve the current system in which individual states have no incentive, and are in fact discouraged, from developing high safety standards.\textsuperscript{212} This occurs because a state with high safety standards will watch as money goes to other states with less stringent standards without any corresponding liability.

In order for Congress to regulate an industry there must be something that brings that industry within Congress’ power. Article I, section 8 of the U.S. Constitution states, “The Congress shall have Power To . . . regulate Commerce with foreign Nations, and among the several States, and with the Indian Tribes . . . .”\textsuperscript{213} In United States v. International Boxing Club of N.Y., Inc., the Supreme Court, in order to determine if the Government’s civil suit for antitrust violations under the Sherman Act could continue, found that boxing affects inter-state commerce.\textsuperscript{214} The Court stated that even though the

\textsuperscript{208} See supra Part III.D.
\textsuperscript{209} The economic impact of a fight on a state can be huge. See Rafael, supra note 22. However, because of the relative sizes of the revenue streams, this will not affect the federal government as much.
\textsuperscript{210} Id.
\textsuperscript{211} See supra Part III.
\textsuperscript{212} Rafael, supra note 22.
\textsuperscript{213} U.S. CONST. art. I, § 8.
boxing match itself takes place in only one state, this “fact alone does not bar application of the Sherman Act to a business based on the promotion of such matches, if the business is itself is engaged in interstate commerce or if the business imposes illegal restraints on interstate commerce.”215 The Court then went on to find that since much of the International Boxing Club’s revenue came from interstate operations, it could be regulated by Congress.216 The Court finally concluded that exemptions to the Sherman Act were properly determined by Congress and not the Court (thereby allowing Congress to enact an antitrust exemption for professional baseball without having to create similar exemptions for other professional sports).217

While International Boxing Club of N.Y. established that parts of the boxing industry could be regulated on the ground that they were in commerce,218 it is still not apparent that the federal government can directly regulate the state athletic commissions. The simplest way for the federal government to ensure boxer safety would be to exercise direct control over the commissions. However any regulation that was formulated such that federal authorities exercised direct control over state athletic commissions would likely be struck down as unconstitutional.219 An example of an effective, but likely unconstitutional, regulation would provide that a federal entity, such as the USBC, watch over the state athletic commissions to make sure that they were enforcing the PBSA. If the state commissions continued to fail to enforce the PBSA, the federal government could compel them to comply with the statute by using its political and economic power. This regulation would be relatively inexpensive and easy to administer because it would only entail regulating commissions, as opposed to regulating each individual fight that takes place within the state. While the most practical approach to improving boxer safety would be federal oversight of the state athletic commissions, the best way to guarantee that the solution would be constitutional is for Congress to engage in federal preemption, without first trying to work with the

215 Id. at 241.
216 Id.
217 Id. at 243.
218 Id. at 241.
219 See infra text accompanying notes 221-230.
states, in light of the relationship between the federal and state governments.

Any attempt by the federal government to regulate boxing would have to comply with the Tenth Amendment of the U.S. Constitution. The Tenth Amendment states, “The powers not delegated to the United States by the Constitution, nor prohibited to it by the States, are reserved to the States respectively, or to the people.”220 One of the most important cases interpreting the scope of the Tenth Amendment is Printz v. United States.221 In Printz, the Supreme Court, in a 5-4 decision, held that state officers cannot be “dragooned” into carrying out federal functions,222 meaning that the federal government cannot use state officials for the purpose of enforcing federal regulations.

Printz involved a challenge to interim provisions of the Brady Handgun Violence Prevention Act.223 The interim provisions required a firearms dealer who wished to transfer ownership of a handgun to receive information from the buyer and transfer it to the chief law enforcement officer (“CLEO”) of the buyer’s residence.224 The CLEO then had to “make a reasonable effort to ascertain within 5 business days whether receipt or possession would be in violation of the law, including research in whatever State and local recordkeeping systems are available and in a national system designated by the Attorney General.”225 After that point, the CLEO was not required to take any further action.226 Also, the CLEO was not required to do a background check “if the purchaser possesses a state handgun permit issued after a background check or if state law provides for an instant background check.”227 Based on these requirements, the Court stated that “it is apparent that the Brady Act purports to direct state law enforcement officers to participate, albeit only temporarily, in the administration of a federally enacted regulatory scheme.”228 A

220 U.S. CONST. amend. X.
222 See id. at 928.
223 Id. at 902.
224 Id. at 902-03.
225 Id. at 903 (quoting 18 U.S.C. § 922(s)(2)).
226 Id.
227 Id. (citations omitted).
228 Id. at 904.
majority of the Court found that this was constitutionally impermissible.229

Applied to the area of boxing regulation, the ruling in Printz prevents the federal government from engaging in direct oversight of the state athletic commissions for the purpose of ensuring that the federal laws are being complied with. However, the majority in Printz conceded that actions which achieved similar results were permissible when the Court looked at federal statutes that utilized state actors and stated, “Some of these are connected to federal funding measures, and can perhaps be more accurately described as conditions upon the grant of federal funding than as mandates to the States . . . .”230 As a result, it is possible that federal oversight could be constitutional if it was a requirement of receiving federal funds. In South Dakota v. Dole, the Court held that the federal government could use the threat of withholding money for highways to South Dakota to encourage the state to set the drinking age at twenty-one years old.231 However, it is unlikely that Congress will be able to tie federal money into an area that would provide a sufficient nexus to allow for federal oversight of the state athletic commissions.232 Regardless of whether or not Congress could find a way to tie regulation into funding, it could directly preempt the field of boxing regulation because that does not entail using state actors to enforce federal law.233

The Printz ruling does not prevent the federal government from preempting the field of boxing regulation. In fact, while the Printz decision sought to protect states’ rights, it may have the opposite effect. Instead of using state actors to carry out federal law indirectly, the federal government has to enforce the law directly. The decision refers to the separation of powers: “The power of the Federal Government would be augmented immeasurably if it were able to impress into its service—and at no cost to itself—the police officers of the 50 States.”234 The dissent points out that “the majority’s rule seems more likely to damage than to preserve the safeguards

229 Printz, 521 U.S. at 904.
230 Id. at 917-18.
232 While a nexus can possibly be found to Medicaid or Medicare, for example, cutting off of these funds would be an extreme threat, especially given the low level of interest in establishing minimum safety standards for professional boxing.
233 See infra text accompanying notes 234-238.
234 Printz, 521 U.S. at 922.
against tyranny provided by the existence of vital state governments.”\textsuperscript{235} This is because “[b]y limiting the ability of the Federal Government to enlist state officials in the implementation of its programs, the Court creates incentives for the National Government to aggrandize itself.”\textsuperscript{236} This is certainly a valid point in the area of boxing regulation where, since federal control cannot be used over state agencies, the result must be the dissolution of the state athletic commissions and the creation of a federal one.

It is clear from prior case law interpreting the U.S. Constitution that Congress can directly regulate boxing; it just cannot regulate boxing by overseeing the state commissions. It has already been established that the boxing industry is in interstate commerce,\textsuperscript{237} and that Congress may constitutionally regulate items in interstate commerce.\textsuperscript{238} Therefore, Congress has the authority to directly regulate the boxing industry.

Federal preemption of boxing regulation should be achieved by amending the PBAA. An entity like the USBC should be created, but provisions should be added for the selection of a federal boxing commissioner for each state. These federal commissioners could be set up in the same way that the state commissions are currently operated. While this system would require much more federal attention than the proposed PBAA, it would be preferable for two reasons. First, it would be effective. The PBAA increases federal bureaucracy while still heavily relying on the state athletic commissions, thus calling into question whether it will produce tangible results. Federal preemption, while requiring much more federal bureaucracy, would eliminate any reliance on the state commissions, making success almost a certainty. Second, the costs of creating additional federal bureaucracy could be offset by taxing the boxing industry. While this step would represent a significant departure from current practice, federal control of the boxing industry is necessary to provide a safer environment for the participants.

\textsuperscript{235} \textit{Id.} at 959 (Stevens, J., dissenting).
\textsuperscript{236} \textit{Id.}
\textsuperscript{238} U.S. CONST. art. I, § 8.
CONCLUSION

From the time of its inception in the United States, the sport of boxing has not had an adequate system to ensure the safety of the fighters. While attempts to regulate the boxing industry have been made with the formation of the state athletic commissions and the passage of the PBSA, they have not been successful. Since current regulations have fallen short of their goals, the best way to achieve comprehensive enforcement is for Congress to take control of all boxing regulation. While this is a major departure from the current system, it is necessary to ensure that the fights are conducted under the safest possible conditions. Even though this would require somewhat complex legislation from Congress, the recent deaths of Martin Sanchez and Leavander Johnson illustrate the high cost of not properly regulating the sport. These deaths, along with the erosion of the PBSA through the Joe Mesi case, have created a climate where Congress should realize that increased federal regulation is necessary for the sport. If not, more preventable deaths will occur.

J. Bradley Clair†

† J.D. Candidate, Brooklyn Law School, 2008; B.A., Tulane University, 2003. Thanks to everyone at the Brooklyn Law Review. Thanks also to Michael Kim for several conversations about MMA and boxing regulations. Most importantly, thanks to John and Dr. Andrea Clair and Davis Iszard for all their love and support.
They Won’t Come Knocking No More

HUDSON V. MICHIGAN AND THE DEMISE OF THE KNOCK-AND-ANNOUNCE RULE

The requirement is no mere procedural nicety or formality attendant upon the service of a warrant. Decisions in both federal and state courts have recognized, as did English courts, that the requirement is of the essence of substantive protections which safeguard individual liberty.

*Ker v. California*¹

I. INTRODUCTION

A tenet central to the Founders’ conception of the Fourth Amendment was the protection of the right to privacy in one’s home.² Throughout its history the Supreme Court has recognized this purpose by imposing limitations on the ability of law enforcement officials to enter one’s home.³ One such limitation is the knock-and-announce rule. This rule requires police officers to knock and announce their presence and then wait a reasonable amount of time before making a forcible entry when executing a search.⁴ A common law principle established well before the Bill of Rights, this rule was incorporated into the reasonableness analysis of Fourth Amendment search-and-seizure law in order to safeguard individuals’ rights against unannounced police entry into their homes.⁵

---

¹ 374 U.S. 23, 49 (1963) (Brennan, J., concurring).
² *Id.* at 51-52. The Fourth Amendment provides:

> The right of the people to be secure in their persons, houses, papers and effects, against unreasonable searches and seizures, shall not be violated and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the person or things to be seized.

> U.S. CONST. amend. IV.


⁵ *See id.* at 934 *infra* Part II.A.
Typically, when the government and police fail to uphold Fourth Amendment protections, the evidence gathered during the unlawful search must be suppressed under the exclusionary rule.6 In Hudson v. Michigan, however, the Supreme Court announced in a 5-4 decision that the exclusionary rule does not apply to evidence obtained when police officers fail to knock and announce, and accordingly, the evidence seized can be used at trial to determine a defendant’s guilt.7

The Supreme Court’s decision in Hudson appears to be another step in the weakening of the knock-and-announce rule. In 1995, Wilson v. Arkansas made the rule a constitutional command.8 Since Wilson, however, the Court has carved out exceptions that have significantly watered down the rule’s impact.9 By not applying the exclusionary rule to knock-and-announce violations, the Hudson majority effectively took away any legal incentive for police officers to conform their actions to what the Court held was constitutionally required in Wilson.10 The Court’s decision not to apply the exclusionary rule was largely based on the substantial social costs generated by expanding its use.11 In reaching its conclusion, the majority reasoned that the increased professionalism of today’s police officers effectively deters law enforcement officials from violating the knock-and-announce rule.12

This Note contests the Court’s conclusion in Hudson that the increased professionalism of the police force effectively deters an officer’s unlawful behavior. First, it is doubtful that such a trend toward increased professionalism actually exists. On the contrary, the growing involvement of paramilitary units in local police activities have made militaristic police tactics increasingly commonplace.13 Second, when this trend is

---

8 514 U.S. at 936.
10 See Hudson, 547 U.S. at 605 (Breyer, J., dissenting) (discussing how police incentive to comply with knock-and-announce violations has significantly diminished as a result of the majority’s holding); see also Radley Balko, CATO INSTITUTE, OVERKILL: THE RISE OF PARAMILITARY POLICE RAIDS IN AMERICA 34 (2006), http://www.cato.org/pubs/wtpapers/balko_whitepaper_2006.pdf.
11 Hudson, 547 U.S. at 594-99; see infra text accompanying notes 86-89.
12 Id. at 598-99.
considered alongside the fact that *Hudson* provides officers with little incentive to comply with the knock-and-announce guidelines, it is likely that combative police behavior and an increase in wrong-door raids will result.\(^{14}\) This Note contends that in order to uphold the protections of the knock-and-announce requirement and ensure police discipline, the exclusionary rule must be applied. In support of this argument, this Note looks to the experience of the *Miranda* rule in order to demonstrate the problems that can result when the Court weakens an established requirement to the point that there is no longer any real incentive to comply with its procedures.\(^{15}\)

Part II of this Note briefly surveys the history of the knock-and-announce rule as it has evolved from its common law origins and highlights the exclusionary rule as it applies to knock-and-announce violations. Part III summarizes the Court's decision in *Hudson v. Michigan*, primarily focusing on the majority's argument that the exclusionary rule should not apply to evidence seized in violation of the knock-and-announce requirement because the substantial social costs generated by the rule's application outweighs the deterrence benefits on police misconduct. Part IV discusses the Court's cost-benefit analysis and suggests that the mechanisms the Court puts forth insufficiently deter police misconduct and instead foster aggressive police tactics and detrimental behavior. Part V compares how police training has responded to the Court's *Miranda* jurisprudence and contends that unless the exclusionary rule is applied to knock-and-announce violations, police officials may encourage officers to bypass knock-and-announce procedures.

---

\(^{14}\) See BALKO, supra note 10, at 3-5.

\(^{15}\) See Charles D. Weisselberg, *Saving Miranda*, 84 CORNELL L. REV. 109 (1998) [hereinafter Weisselberg, *Saving Miranda*] (arguing that the original vision of *Miranda* has been transformed, and as a result police departments have begun to promote a policy of questioning "outside *Miranda*"). The policy of training officers to question outside *Miranda* became apparent in the late 1990s as officers in California were being taught that as a result of Supreme Court decisions, "it is perfectly acceptable to violate *Miranda* because *Miranda* . . . has no application except to bar certain statements from the prosecution's case-in-chief. Proponents of this new vision tell police that they need not cease interrogating a suspect who has asserted his or her Fifth Amendment rights." *Id.* at 132.
announce guidelines in much the same way they have encouraged officers to question suspects “outside Miranda.”

II. LEGAL BACKGROUND

A. The Knock-and-Announce Rule

Under the Fourth Amendment, police officers must knock and announce their presence when executing a search warrant at an individual’s home. This knock-and-announce rule provides residents an opportunity to willingly open their door before police officers forcibly enter. The rule is designed to protect people’s privacy interests in their home against unreasonable searches and seizures, which is a right central to the Fourth Amendment. In particular, the Supreme Court has set forth three main purposes of the rule: “1) reducing the potential for violence to both the police officers and the occupants of the home into which entry is sought; 2) curbing the needless destruction of property; and 3) protecting the individual’s right to privacy in his or her house.”

The knock-and-announce rule is a common law principle that dates to thirteenth-century England. In 1604, English courts formally established the knock-and-announce requirement in Semayne’s Case. Based on the principle that “a man’s house is his castle,” the court recognized the privacy interest an individual has in his or her home and determined that before forcibly entering one’s home, officers should announce their presence and allow the occupant time to open his or her door.

---

18 U.S. CONST. amend. IV; see Wilson, 514 U.S. at 934.
19 United States v. Dice, 200 F.3d 978, 982 (6th Cir. 2000) (holding that evidence obtained in violation of the knock-and-announce rule should be suppressed at trial, regardless of its use to help fight the War on Drugs).
20 Miller v. United States, 357 U.S. 301, 307 (1958) (holding that under 18 U.S.C § 3109 the police officers violated the rights of the defendant when they knocked and said “police” in a low voice, but did not announce their purpose and their reason for arresting him). Records show that in the 13th Yearbook of Edward IV, it was illegal for a sheriff to forcibly enter a man’s home and arrest him for a suit which involved either debt or trespass. See id.
22 See Miller, 357 U.S. at 307-08 (1958); see also Sabbath v. United States, 391 U.S. 585, 589 (1968). Based on the principle, “a man’s house is his castle,” the English court believed that, “the house of every one is to him his castle and fortress, as well for his defense against injury and violence, as for his repose.” Martin Estrada, A Toothless Tiger in the Constitutional Jungle: “The Knock-and-Announce Rule” and The Sacred Castle Door, 16 U. FLA. J.L. & PUB. POL’Y 77, 80 (2005) (quoting 77 Eng. Rep.
This common law principle, now codified at 18 U.S.C. § 3109, became part of American statutory law when Congress passed the Espionage Act in 1917.23

In 1995, the Supreme Court unanimously held in Wilson v. Arkansas that the knock-and-announce rule should be considered in the Fourth Amendment’s reasonableness analysis of a search or seizure.24 In reaching its conclusion, the Court surveyed the common law’s knock-and-announce requirement and determined that the Framers of the Constitution intended the method used to enter an individual’s residence should be a factor when evaluating the reasonableness of a search or seizure.25 This decision transformed the knock-and-announce rule from a common law or statutory mandate into a constitutional command.26

Additionally, the Wilson Court recognized that the knock-and-announce requirement is not an inflexible or rigid rule. The Court suggested that there are certain circumstances in which an unannounced entry would be reasonable under the Fourth Amendment.27 Specifically, the Court determined that the “knock-and-announce requirement could give way ‘under circumstances presenting a threat of physical violence’ or ‘where police officers have reason to believe that the evidence

194, 195 (K.B. 1603)). Moreover, in recognizing the sacredness of the home, in an English Parliamentary debate, William Pitt, Earl of Chatham, made a powerful statement regarding one’s right to privacy in their home:

[The poorest man may in his cottage bid defiance to all the forces of the Crown. It may be frail; its roof may shake; the wind may blow through it; the storm may enter; the rain many enter; but the King of England cannot enter-all his force dares not cross the threshold of the ruined tenement.]

Miller, 357 U.S. at 307.

23 Hudson v. Michigan, 547 U.S. 586, 589 (2006); see Miller, 357 U.S. at 308. Section 3109 of the U.S. Code governs the breaking of doors or windows for entry or exit, stating in relevant part:

The officer may break open any outer or inner door or window of a house, or any part of a house, or anything therein, to execute a search warrant, if, after notice of his authority and purpose, he is refused admittance or when necessary to liberate himself or a person aiding him in the execution of the warrant.


24 Wilson, 514 U.S. at 936.
25 Id. at 934.
26 See id. at 936.
27 Id. at 934-36. As the Wilson Court stated, “[W]e simply hold that although a search or seizure of a dwelling might be constitutionally defective if police officers enter without prior announcement, law enforcement interests may also establish the reasonableness of an unannounced entry.” Id. at 936.
would likely be destroyed if advance notice was given.”

In *Wilson*, however, the Court refrained from making any bright line rule and asserted that the lower courts would be responsible for assessing when an unannounced entry is reasonable.

Two years later, in *Richards v. Wisconsin*, the Supreme Court carved out exceptions to the requirement that a police officer must knock and announce. In *Richards*, the Court invalidated Wisconsin’s per se rule that a no-knock entry is always justifiable in felony drug cases, yet concluded that certain circumstances warrant a no-knock entry. The Court established that where police have a reasonable suspicion that knocking-and-announcing would be “dangerous or futile” or would “inhibit the effective investigation of the crime by, for example, allowing destruction of the evidence,” then a no-knock entry under the circumstances is lawful. Accordingly, this reasonableness inquiry is to be done on a case-by-case basis, balancing the concerns of law enforcement officials with an individual’s right to privacy in his or her home.

Six years later the Court attempted to clarify some of the ambiguities resulting from *Wilson* and *Richards*. In *United States v. Banks*, the Court determined that in a case where police have knocked-and-announced their presence and have a reasonable belief that destruction of the evidence is possible, fifteen to twenty seconds is a reasonable amount of time to wait before forcibly entering an individual’s home.

Recognizing that the purpose of the knock-and-announce rule

---

29 Wilson, 514 U.S. at 936.
30 Richards, 520 U.S. 385 (holding that a police officer’s failure to knock and announce would be justifiable if the officer had reasonable suspicion that knocking and announcing would be dangerous or futile or would prohibit an effective investigation).
31 Id. at 394. *E.g.*, United States v. Dice, 200 F.3d 978, 983 (6th Cir. 2000) (stating that circumstances in which the knock-and-announce rule is not required include (1) the occupant knows that the police are present and their purpose for being there; (2) the police have a justifiable belief that an individual may be in imminent danger; and (3) the officers have a justifiable belief that the residents in the home are aware of police presence and as a result are trying to escape or discard the evidence).
32 Richards, 520 U.S. at 394. As the Supreme Court recognized, the showing of reasonable suspicion “is not high,” but if challenged, the police must be able to justify their no-knock entry. Id. at 394-95.
33 Id.
35 Id. at 40. In forming its conclusion, the Court reasoned that “15 to 20 seconds does not seem an unrealistic guess about the time someone would need to get in a position to rid his quarters of cocaine.” Id.
is to allow a resident to willingly open his or her door, the Court again refrained from making a blanket categorical rule as to the appropriate manner and timing of a police officer's entry. Rather, the Court stated that the reasonableness of an officer's entry should depend on the totality of the circumstances in a given case.

The knock-and-announce rule is embedded in our nation's history. In recognizing its importance in protecting an individual's right to privacy, the Supreme Court made it a constitutional command and part of the reasonableness inquiry of the Fourth Amendment. In *Richardson* and *Banks*, the Court began to carve out exceptions to this requirement. *Hudson* represents the Court's latest, and most troubling, dilution of a once tenacious constitutional rule.

**B. The Exclusionary Rule**

The exclusionary rule is a judicially created remedy that requires the suppression of evidence obtained in violation of the Fourth Amendment. Any physical or tangible goods, as well as verbal statements, that are gathered pursuant to an unreasonable search or seizure may not be used to prove a defendant's guilt at trial. However, there are exceptions to the exclusionary rule that allow parties to introduce the illegally

---

36 Id. at 41. The Court reasoned that, “in a case with no reason to suspect an immediate risk of frustration or futility in waiting at all, the reasonable wait time may well be longer . . . since [police] ought to be more certain the occupant has had time to answer the door.” Id.

37 Id. at 42.


39 See infra Part V.

40 Weeks v. United States, 232 U.S. 383, 392 (1914) (holding that it was a violation of the defendant’s Fourth Amendment rights when the government illegally seized letters and papers from defendant which were later used as evidence against him at trial). The Court felt that the letters should have been returned to the defendant and that their use constituted a prejudicial error. Id. at 398.

41 Wong Sun v. United States, 371 U.S. 471, 484-86 (1963) (holding that verbal evidence that was obtained during an unlawful entry and arrest is considered “fruits of the agents unlawful action” and therefore must be suppressed under the exclusionary rule). The Court has determined:

The exclusionary rule reaches not only primary evidence obtained as a direct result of an illegal search or seizure, but also evidence later discovered and found to be derivative of an illegality or “fruit of the poisonous tree.” It “extends as well to the indirect as the direct products” of unconstitutional conduct.

obtained evidence at trial. The Supreme Court has concluded that the use of evidence seized under these exceptions does not violate an individual’s Fourth Amendment rights.42

The exclusionary rule was first applied to federal prosecutions in *Weeks v. United States*.43 In *Weeks*, the Supreme Court determined that evidence obtained without a warrant is a violation of an individual’s constitutional rights and therefore shall not be admitted at trial.44 The Court reasoned that to hold otherwise would render the Fourth Amendment ineffective, as it would have “no value . . . [and] might as well be stricken from the Constitution.”45

Subsequently, the Supreme Court expanded the use of the exclusionary rule in *Mapp v. Ohio* and held it applicable to state courts through the Due Process Clause of the Fourteenth Amendment.46 The majority’s opinion in *Mapp* appears to be supported by two justifications, both of which guided the Court’s application of the exclusionary rule in federal trials.47 First, the Court reasoned that the exclusionary rule would deter police misconduct, and second, the Court believed that the rule’s application would uphold judicial integrity.48 As the *Mapp* majority emphasized, “the purpose of the exclusionary rule ‘is to deter—to compel respect for the constitutional

42 See *The Exclusionary Rule*, 35 GEO. L.J. ANN. REV. CRIM. PROC. 186, 186-201 (2006) (discussing the exceptions to the exclusionary rule). For the purposes of this Note, the exceptions to the exclusionary rule do not need to be discussed in detail. However, it is worth noting that several exceptions to the exclusionary rule do exist. These include: the good faith exception established in *United States v. Leon*, 468 U.S. 897 (1984); the attenuation exception, see *Wong Sun*, 371 U.S. 471; the independent source exception, see *Silverthorne Lumber Co. v. United States*, 251 U.S. 385 (1920); the inevitable discovery exception, see *Nix v. Williams*, 467 U.S. 431 (1984); and other collateral uses, see, e.g., *Stone v. Powell*, 428 U.S. 465 (1976) (where the court made an exception in a habeas corpus proceeding). For more information and cases in which these exceptions can be found, see *The Exclusionary Rule*, supra.
43 *Weeks*, 232 U.S. at 392.
44 Id.
45 Id. at 393.
46 *Mapp v. Ohio*, 367 U.S. 643, 655-56 (1961) (holding that evidence obtained based on the police’s illegal warrantless entry can not be admissible in state courts). *Mapp* signified the overruling of *Wolf v. Colorado*, where twelve years prior the Supreme Court refused to extend the exclusionary rule to states. *Id.* at 654-55; see also *Wolf v. Colorado*, 338 U.S. 25, 25-26 (1949) (holding that Due Process of the law does not automatically mean that the first eight Amendments of the Constitution are incorporated into Fourteenth Amendment; here the Court did not believe that the federal exclusionary rule should be applied to the states).
47 See *Mapp*, 367 U.S. 643, 657-59; see also *Stone*, 428 U.S. at 484-85.
48 See *Mapp*, 367 U.S. at 658-59; see also *Stone*, 428 U.S. at 484-85. Additionally, these two principles appeared to first be laid out pre-*Mapp*, in *Elkins v. United States*, when the Court applied the exclusionary rule in federal trials. *Id.* at 484.
guaranty in the only effectively available way—by removing the incentive to disregard it.” Moreover, the Court noted the importance of preserving judicial integrity, asserting that “nothing can destroy a government more quickly than its failure to observe its own laws... If the government becomes a law breaker it breeds contempt for law; it invites every man to become a law unto himself; it invites anarchy.” Accordingly, these guiding principles supported the Court’s reasoning that when evidence is obtained in violation of the Fourth Amendment, it should be suppressed at both the federal and state level.

In the 1974 case United States v. Calandra, the Supreme Court clarified that the main purpose of the exclusionary rule is to deter future unlawful police misconduct. As the Calandra majority stated, “[T]he rule is a judicially created remedy designed to safeguard Fourth Amendment rights generally through its deterrent effect, rather than a personal constitutional right of the party aggrieved.” As a deterrence mechanism, the exclusionary rule is not intended for individual deterrence and rehabilitation but, in contrast, focuses on institutional deterrence aimed at correcting the future conduct of the police force as a whole. Furthermore, the Court determined that the exclusionary rule is not applicable in all cases where evidence is illegally obtained. Rather, a balancing test must be conducted to see if the

49 Mapp, 367 U.S. at 656 (quoting Elkins v. United States, 364, U.S. 206, 217 (1960)).
50 Id. at 659 (quoting Olmstead v. United States, 277 U.S. 438, 485 (1928) (Brandeis, J., dissenting)).
51 Id. at 655, 660.
52 United States v. Calandra, 414 U.S. 338, 347 (1974) (holding that evidence that was seized outside the scope of the warrant at the defendant’s place of business does not qualify him to refuse to testify in grand jury proceedings on the grounds that the evidence was obtained unlawfully and should be suppressed in accordance with the exclusionary rule). The Court’s decision in Calandra appears to be grounded in the precedent of Elkins v. United States. Id. at 347; see also Elkins, 364 U.S. at 217.
53 Calandra, 414 U.S. at 348 (emphasis added) (explaining how the purpose of the exclusionary rule is not to remedy the injury to privacy that resulted from an illegal search, but rather to prevent future police misconduct). Additionally, in supporting the main objective of police deterrence, the Court suggested, “that the application of the rule has been restricted to those areas where its remedial objectives are thought most efficaciously served.” Id.
54 United States v. Leon, 468 U.S. 897, 917 (1984) (holding that although the evidence was obtained illegally, the officers acted in reasonable reliance on a search warrant and therefore the evidence need not be suppressed at trial). Additionally, Leon established the Good Faith exception to the exclusionary rule. Id. at 922-23.
55 Calandra, 414 U.S. at 348.
deterrence benefit of applying the exclusionary rule outweighs the costs generated by its use.\(^{56}\) The Court indicated that costs of the exclusionary rule include the possibility that the guilty go free or, at the very least, that defendants will use the error to plea bargain for a reduced sentence.\(^{57}\) On the other hand, as effective deterrence mechanisms focus on creating incentives to prevent police misconduct, the Court determined that the exclusionary rule’s deterrence benefits should be applied when “the police have engaged in willful, or at the very least negligent, conduct which has deprived the defendant of some right.”\(^{58}\) Therefore, by suppressing evidence gained as a result of officer’s unlawful behavior, the Court encourages a “greater degree of care toward the rights of an accused.”\(^{59}\)

Thus, with respect to preventing future police misconduct, the exclusionary rule’s deterrence mechanism is crucial in safeguarding an individual’s constitutional rights. This Note suggests that it is this focus on institutional deterrence that makes the exclusionary rule the only viable and effective remedy for knock-and-announce violations.\(^{60}\) Because other deterrence mechanisms have proven ineffective, it is only through the procedural incentives of the exclusionary rule that police behavior will be rectified.\(^{61}\)

III. **HUDSON V. MICHIGAN**

In a 5-4 decision, the U.S. Supreme Court held that evidence obtained in violation of the knock-and-announce rule is admissible at trial and can be used to prove a defendant’s guilt.\(^{62}\) In reaching this conclusion, the Court failed to extend the exclusionary rule, a doctrine that orders that evidence seized in violation of the Fourth Amendment be suppressed at trial.\(^{63}\) In *Hudson*, Michigan police officers obtained a warrant permitting the search of defendant Brooker T. Hudson’s home

\(^{56}\) See id.; see also Leon, 468 U.S. at 906-07 (Leon contains an in-depth analysis of the cost-benefit analysis. The *Calandra* case establishes that a balancing test is necessary to determine when the exclusionary rule is an appropriate remedy and it requires that the costs and benefits must be weighed.).

\(^{57}\) Id. at 907.

\(^{58}\) *Id.* at 919 (quoting United States v. Peltier 422 U.S. 531, 539 (1975)).

\(^{59}\) *Id.*

\(^{60}\) See infra Part V.

\(^{61}\) See infra Part V.


\(^{63}\) *Id.*
for drugs and firearms. Upon executing the warrant on August 27, 1998, the police arrived at Hudson’s door and announced their presence but did not knock. The officers then waited only three to five seconds before opening Hudson’s unlocked door and entering his home. Inside police recovered drugs, including cocaine rocks found in Hudson’s pocket, and a loaded gun that police discovered between the cushion and armrest of the chair in which Hudson was seated. Consequently, police charged Hudson with unlawful drug and firearm possession in violation of Michigan state law.

Hudson subsequently moved to suppress the evidence. He argued that the police’s premature entry did not comply with the knock-and-announce rule and, therefore, violated his Fourth Amendment rights. The trial court granted Hudson’s motion to suppress the evidence, but the Michigan Court of Appeals reversed the decision upon interlocutory review. In making this determination, the Michigan Court of Appeals followed Michigan Supreme Court precedent and concluded that suppression was not the appropriate remedy when police gathered evidence in violation of the knock-and-announce rule but pursuant to a valid search warrant. Subsequently, the Michigan Supreme Court denied leave to appeal, and Hudson was convicted of drug possession and sentenced to eighteen months probation. Following his conviction, Hudson renewed his Fourth Amendment claim on appeal, but the Michigan Court of Appeals affirmed his conviction. Thereafter, the Michigan Supreme Court once again denied review, and on January 27, 2005, the U.S. Supreme Court granted certiorari. The question before the Court was whether the exclusionary rule applied to evidence that police obtained in violation of the

---

64 Id. at 588.
65 Brief Amicus Curiae of the Criminal Justice Legal Foundation in Support of Respondent at 2, Hudson, 547 U.S. 586 (No. 04-1360) [hereinafter CJLF Brief].
66 Hudson, 547 U.S. at 588.
67 Id.
68 Id.
69 Id.
70 Id.
71 Id.
72 Hudson, 547 U.S. at 589. See People v. Stevens, 597 N.W.2d 53, 64 (1999) (holding that the exclusionary rule does not apply to evidence obtained in violation of the knock-and-announce requirement).
73 CJLF Brief, supra note 65, at 2; see also Hudson, 547 U.S. at 589.
74 Hudson, 547 U.S. at 589.
75 CJLF Brief, supra note 65, at 2.
knock-and-announce rule, and therefore was required to be suppressed at trial.\textsuperscript{76} Justice Scalia, in writing the opinion for the Court, declared that evidence seized in violation of the knock-and-announce rule need not be suppressed.\textsuperscript{77}

Although the majority opinion contains four separate discussions, this Note will focus on the Court's argument that the substantial costs generated by applying the exclusionary rule outweigh its deterrence benefits.\textsuperscript{78} The Court's opinion began by acknowledging that the knock-and-announce rule stems from common law principles and is now a command of the Fourth Amendment.\textsuperscript{79} In relying on case law that has shaped both the exclusionary rule and the knock-and-announce rule, the Court explained why this type of violation does not warrant suppression.

First, Justice Scalia addressed the causation argument that suggests a court can admit evidence at trial that would not have been discovered “but-for” police misconduct if the causal connection between the misconduct and seizing of the evidence is notably attenuated.\textsuperscript{80} Following this principle, Scalia suggested that the exclusion of evidence should not occur automatically, even though a constitutional violation was the but-for cause of obtaining the evidence.\textsuperscript{81} Accordingly, Scalia believed that even if there is a direct causal connection between the evidence obtained and the illegal entry, the evidence should only be suppressed when suppression would directly remedy the constitutional interests violated. In \textit{Hudson}, Scalia determined that the officers' illegal entry by failing to knock and announce was not a but-for cause of obtaining the evidence because police would have ultimately seized the evidence pursuant to the execution of a search warrant.\textsuperscript{82} As Scalia noted, “Whether that preliminary misstep had occurred or not, the police would have executed the

\begin{itemize}
\item \textsuperscript{76} \textit{Hudson}, 547 U.S. at 590.
\item \textsuperscript{77} \textit{Id.} at 594.
\item \textsuperscript{78} See \textit{id.} at 594-600.
\item \textsuperscript{79} See \textit{id.} at 588; see also Wilson v. Arkansas, 514 U.S. 927, 936 (1995) (holding that whether an officer followed the knock-and-announce rule forms part of the reasonableness inquiry as to whether a search is valid under the Fourth Amendment).
\item \textsuperscript{80} \textit{Hudson}, 547 U.S. at 592-94; see also Wong Sun v. United States, 371 U.S. 471, 484-85 (1963) (holding that verbal evidence seized in violation of an unlawful entry and arrest is considered “fruits” of the agents' unlawful action and therefore must be suppressed under the exclusionary rule).
\item \textsuperscript{81} \textit{Hudson}, 547 U.S. at 592-93.
\item \textsuperscript{82} \textit{Id.}
\end{itemize}
warrant they had obtained, and would have discovered the guns and drugs inside the house.” Moreover, Scalia asserted that the exclusionary rule is not designed to safeguard evidence from the government and stated, “Since the interests that were violated in this case have nothing to do with the seizure of the evidence, the exclusionary rule is inapplicable.”

Second, Scalia noted that the exclusionary rule is only applied when the deterrence benefits outweigh the social costs generated by its application. Here, Scalia found that the social costs of applying the exclusionary rule were substantial and significantly outweighed the deterrence effects on police misconduct. Scalia alleged that the social costs of applying the exclusionary rule were twofold. First, it would allow criminals go free because evidence of their guilt was suppressed. Second, it would produce a flood of cases where individuals would claim that either officers violated the knock-and-announce rule or that the various exceptions that permit officers not to knock were inapplicable. In contrast, Scalia suggested that the deterrence benefits of preventing police officer misconduct would be nominal and that other remedies were available to ensure police acted responsibly. In particular, Scalia noted both the availability of civil rights suits against police officers under 42 U.S.C. § 1983 as well as the “internal discipline” and “increasing professionalism” of police forces.

83 Id. at 592 (emphasis added).
84 Id.
85 Id. at 594 (emphasis added).
86 Id.; see also United States v. Leon, 468 U.S. 897, 907-08 (1984) (stating that the exclusionary rule should be applied when the effect of deterring police misconduct outweighs the social costs generated by applying the rule).
87 Hudson, 547 U.S. at 595.
88 Id.
89 Id. at 595-96.
90 Id. at 596-99.
91 Id. Section 1983 of the U.S. Code provides citizens a cause of action:

Every person who, under color of any statute, ordinance, regulation, custom, or usage, of any State or Territory or the District of Columbia, subjects, or causes to be subjected, any citizen of the United States or other person within the jurisdiction thereof to the deprivation of any rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured in an action at law, suit in equity, or other proper proceeding for redress, except that in any action brought against a judicial officer for an act or omission taken in such officer’s judicial capacity, injunctive relief shall not be granted unless a declaratory decree was violated or declaratory relief was unavailable. For the purposes of this section, any Act of Congress applicable
In relying on the availability of civil remedies against police officers, the majority concluded that the potential threat of being confronted with a civil rights suit for officer misconduct would significantly deter law enforcement officials from violating knock-and-announce guidelines. The majority rejected the notion that courts do not award significant damages in civil-right suits against law enforcement officials. Rather, the Court reasoned that Congress has made it easier to bring such suits by approving attorney fees in civil rights proceedings and that lawyers are now more disposed to taking on these types of cases. Additionally, the Court found that the internal discipline of today’s police forces will ensure that officers abide by the guidelines set out in the knock-and-announce requirement and respect individuals’ constitutional rights when executing search warrants at their homes. As Scalia noted, there are “[n]umerous sources . . . now available to teach officers and their supervisors what is required of them under this Court’s cases, how to respect constitutional guarantees in various situations, and how to craft an effective regime for internal discipline.”

Lastly, Scalia discussed how the Court’s holding in *Hudson* is consistent with pre-existing law governing the rules

---


92 *Hudson*, 547 U.S. at 596-97.

93 Id. at 597-98.

94 Id.; see also 42 U.S.C. § 1988(b). Section 1988(b) authorizes attorney fees in the following circumstances:

In any action or proceeding to enforce a provision of sections 1981, 1981a, 1982, 1983, 1985, and 1986 of this title, title IX of Public Law 92-318 [20 U.S.C. § 1681 et seq.], the Religious Freedom Restoration Act of 1993 [42 U.S.C. § 2000bb et seq.], the Religious Land Use and Institutionalized Persons Act of 2000 [42 U.S.C. § 2000cc et seq.], title VI of the Civil Rights Act of 1964 [42 U.S.C. § 2000d et seq.], or section 13981 of this title, the court, in its discretion, may allow the prevailing party, other than the United States, a reasonable attorney’s fee as part of the costs, except that in any action brought against a judicial officer for an act or omission taken in such officer’s judicial capacity such officer shall not be held liable for any costs, including attorney’s fees, unless such action was clearly in excess of such officer’s jurisdiction.

42 U.S.C. § 1988(b) (2000). Moreover, the Court asserted that because cases are settled, it may appear that the lower courts are not going forward with civil rights suits against officers. *Hudson*, 547 U.S. at 598.

95 *Hudson*, 547 U.S. at 598-99.

96 Id. at 599.
of suppression. Scalia focused on the cases Segura v. United States, New York v. Harris, and United States v. Ramirez to support his conclusion. Like Hudson, Segura v. United States also involved an illegal entry. In Segura, the officers did not have a search warrant, did not knock, and did not receive permission to enter the defendant’s apartment. Due to delays in obtaining a warrant, police had to wait nineteen hours inside Segura’s apartment until they received a valid warrant and could lawfully conduct a search. The Supreme Court denied Segura’s motion to suppress the evidence gathered, and held that the grounds on which the warrant was granted did not relate in any way to the police’s illegal entry. Accordingly, the Court found that there was an independent source for the warrant under which the evidence was obtained, and therefore suppression of the evidence was not required. Scalia concluded that if the evidence obtained as a result of the

---

97 See id. at 599-602.
101 Segura, 468 U.S. at 798.
102 Id. at 800.
103 Id. at 801.
104 Id. at 814.
105 Id. at 816. In Segura, the Court stated that “the exclusionary rule has no application [where] the Government learned of the evidence from an independent source.” Id. at 805 (1984) (quoting Silverthorne Lumber Co. v. United States, 251 U.S. 385, 392 (1920) (alteration in original)). Accordingly, Silverthorne describes the usage of the independent source doctrine:

The essence of a provision forbidding the acquisition of evidence in a certain way is not merely evidence so required shall not be used before the Court but that it shall not be used at all. Of course this does not mean that the facts thus obtained become sacred and inaccessible. If knowledge of them is gained from an independent source they may be proved like any others, but the knowledge gained by the Government’s own wrong cannot be used by it in the way proposed.

Silverthorne Lumber Co., 251 U.S. at 392. Additionally, in Segura, the search warrant was based upon the testimony of Rivudalla-Vidal who told agents that he had bought cocaine from Segura that day and had plans to meet with the defendant later that evening to receive an additional supply. See Segura, 468 U.S. at 800. The Court deemed the illegality of the initial entry irrelevant by applying the “but-for” causation test. The Court relied on precedent stating that “evidence will not be excluded as ‘fruit’ unless the illegality is at least the ‘but for’ cause of the discovery of the evidence.” Id. at 815. Accordingly, the Court determined that because “[t]he illegal entry into the petitioner’s apartment did not contribute in any way to discovery of the evidence seized under warrant; it is clear, therefore, that not even the threshold ‘but for’ requirement was met in this case.” Id.
illegal entry in Segura was not suppressed, then it would be “bizarre to treat more harshly the evidence in this case.”

In New York v. Harris, the police arrested Harris in his home without a warrant. This illegal arrest, however, did not require the suppression of an incriminating statement Harris had made at the station house. In Harris, the Supreme Court determined that the statement made at the station house was not a result of the defendant’s being in “unlawful custody.” Accordingly, although he had made the statement after a warrantless and non-consensual entry into the defendant’s home, the exclusionary rule did not bar it. In comparison, Scalia stated, “While acquisition of the gun and drugs [in Hudson] was the product of a search pursuant to warrant, it was not the fruit of the fact that the entry was not preceded by knock and announce.”

Lastly, in United States v. Ramirez, the Court found that although unnecessary destruction to property may trigger a Fourth Amendment violation, the circumstances of Ramirez did not produce such a result. The Court did note, however, that if the police’s destruction of property was unreasonable, a causation test would be applied to determine if there was a causal relationship between the property damage and the evidence obtained, which would justify suppression. In citing Ramirez, Scalia asked, “[W]hat clearer expression could there be of the proposition that an impermissible manner of entry does not necessarily trigger the exclusionary rule?”

Justice Kennedy’s concurrence supported two major claims of the majority’s opinion: (1) the knock-and-announce rule does not meet the causation requirement that limits the application of the exclusionary rule; and (2) civil suits and the

---

107 Id. at 601.
108 Id. The incriminating statement in question was a written inculpatory statement signed by Harris after he was arrested and taken to the station house. New York v. Harris, 495 U.S. 14, 16 (1990).
109 Harris, 495 U.S. at 19.
110 Id. at 21.
111 Hudson, 547 U.S. at 601.
112 Id. at 602. In Ramirez, police entered the defendant’s home pursuant to a “no-knock warrant.” Upon being informed that the defendant had guns and drugs in his garage, the police believed the best method of entry would be to break a window in the garage and exhibit their guns to try and deter the defendant from attempting to use his weapons. United States v. Ramirez, 529 U.S. 65, 68-69 (2006).
113 Hudson, 547 U.S. at 602.
114 Id.
internal discipline of police officers serve as effective deterrence remedies for knock-and-announce violations.115 Furthermore, Kennedy asserted that the knock-and-announce rule is still very much the law and if the Court’s decision led to a “widespread pattern of violations . . . and particularly if those violations were committed against persons who lacked the means or voice to mount effective protest,”116 that the Court’s decision in Hudson would have to be revisited.117 Justice Kennedy, however, did not agree with the majority’s reliance on Segura and Harris.118

Justice Breyer wrote a powerful dissent asserting that not only did the majority’s decision take away the incentive to follow the knock-and-announce requirement, but it did so without any support from past precedent.119 First, Breyer believed that the majority should have applied the exclusionary rule, and he stated that the Court’s decision “weakens, perhaps destroys, much of the practical value of the Constitution’s knock-and-announce protection.”120 Second, Breyer criticized the majority’s reliance on civil suits and internal discipline as effective deterrence methods for knock-and-announce violations.121 He noted that the majority failed to cite any cases where a plaintiff has received substantial monetary damages for knock-and-announce violations.122 Moreover, he reasoned

115 Id. at 602-03 (Kennedy, J., concurring).
116 Id. at 604.
117 See id.
118 Id. In his concurrence, Justice Kennedy simply states, “[W]hile I am not convinced that Segura v. United States and New York v. Harris have as much relevance here as Justice Scalia appears to conclude, the Court’s holding is fully supported by Parts I through III of its opinion. I accordingly join those Parts and concur in judgment.” Id.
119 Hudson, 547 U.S. at 604-05 (Breyer, J., dissenting). Additionally, in supporting his opinion that the majority departed from past precedent, Breyer included in the appendix the decisions of all the Fourth Amendment cases that the Supreme Court has decided since the exclusionary rule was established. See id. at 630-32. Breyer apparently used these cases to emphasize that, notwithstanding the established exceptions to the exclusionary rule, “in every case involving evidence seized during an illegal search of a home (federally since Weeks, nationally since Mapp) the Court, with the exceptions mentioned, has either explicitly or implicitly upheld (or required) the suppression of the evidence.” Id. at 613. For discussions of the exceptions mentioned, see id. at 611-14.
120 Id. at 605.
121 Id. at 608-14.
122 Id. at 610-11. Breyer notes:

To argue, as the majority does, that new remedies, such as 42 U.S.C. § 1983 actions or better trained police, make suppression unnecessary is to argue that Wolf, not Mapp, is now the law. To argue that there may be few civil suits because violations may produce nothing “more than nominal injury” is
that the qualified immunity defense available to police officers makes it difficult for individuals to bring suits against them, and thus, civil claims are not an adequate replacement for the exclusionary rule.\footnote{Id. at 611 (emphasis added) (alteration in original) (citations omitted).} Third, Breyer rejected the Court’s but-for analysis, finding that the evidence obtained in \textit{Hudson} resulted from the police’s illegal entry into Hudson’s home.\footnote{Id. at 610.} As Breyer argued, “it is not true that, had the illegal entry not occurred, ‘police would have discovered the guns and drugs inside the house.’ Without that unlawful entry they would have not been inside the house; so there would have been no discovery.”\footnote{Id. at 618 (quoting Justice Scalia, \textit{id.} at 592 (majority opinion)).} Fourth, Breyer emphasized the fact that the knock-and-announce rule is designed to safeguard an individual’s privacy in his or her home from government intrusion.\footnote{Id. at 620-21.} He believed that the majority’s opinion consequently undermined this right, which is central to the Fourth Amendment.\footnote{Id. Supporting his conclusions, Breyer cited a Court decision from this year (2006) which emphasized the longstanding importance that the Court has placed on an individual’s privacy interests in regards to their home. As Breyer asserted, “[J]ust this

\textit{Id.} at 611 (emphasis added) (alteration in original) (citations omitted).

\footnote{Harlow v. Fitzgerald, 457 U.S. 800, 818 (1982).}

The rationale of the qualified immunity defense has two prongs: “[I]t allows officials to carry out their duties confidently, without fear of incurring unexpected liability, and it allows courts to dispose of insubstantial claims prior to trial, sparing officials from unnecessary litigation.” Pounds v. Griepenstroh, 970 F.2d 338, 340 (7th Cir. 1992) (alteration in original). Thus, in determining if an officer is entitled to the qualified immunity defense, the court must make a two part inquiry. First, do the plaintiff’s allegations establish a violation of a constitutional right, and second, was the constitutional right clearly established at the time of the injury? Kathryn R. Urbonya, \textit{Selected Fourth Amendment Issues in Section 1983 Litigation, in 2 14TH ANNUAL SECTION 1983 CIVIL RIGHTS LITIGATION 67, 101 (Hon. George C. Pratt & Martin A. Schwartz eds., 1988).} This two prong test is an objective standard that assumes liability if a reasonable officer would have or should have known, given the same circumstances and the law at the time, that his or her actions were violating a clearly established law. \textit{See id.} If a court finds that the qualified immunity defense applies, officers are saved from the burdens of litigation and do not have to stand trial. Saucier v. Katz, 533 U.S. 194, 200-01 (2001).
rejected the majority’s application of Segura, Harris, and Ramirez, and found they did not lend support to the proposition that suppression is unwarranted when a knock-and-announce violation occurs.128

IV. WHAT DETERRENCE? AN EXAMINATION OF THE INCREASED PROFESSIONALISM AND DISCIPLINE OF TODAY’S POLICE FORCES

In 1963, Justice Brennan warned of the risks of weakening the knock-and-announce rule:

[P]ractical hazards of law enforcement militate strongly against any relaxation of the requirement of awareness. First, cases of mistaken identity are surely not novel in the investigation of crime. The possibility is very real that the police may be misinformed . . . [t]hat possibility is itself a good reason for holding a tight rein against judicial approval of unannounced police entries into private homes. Innocent citizens should not suffer the shock, fright, or embarrassment attendant upon an unannounced police intrusion. Second, the requirement of awareness also serves to minimize the hazards of the officers’ dangerous calling.129

It appears that Brennan’s message is even more pertinent today, as paramilitary units have become increasingly involved in local police activity.130 The majority in Hudson found that the term we have reiterated that ‘it is beyond dispute that the home is entitled to special protection as the center of the private lives of our people.” Id. at 621 (quoting Georgia v. Randolph, 547 U.S. 103, 115 (2006)).

128 See id. at 624-28. In rejection of Segura v. United States, Breyer reasoned that the majority erred in its application of the “independent source doctrine.” Id. at 625-26. He found that, in Hudson, “[t]he search that produced the relevant evidence here is the very search that the knock-and-announce violation rendered unlawful. There simply is no ‘independent source.’” Id. at 625. Furthermore, Breyer found the facts of New York v. Harris inconsistent with Hudson. Id. at 626. In Harris, the parties agreed that any incriminating evidence that was obtained in Harris’ home as a result of his arrest should be excluded. Id. at 627. Thus, the only question at trial was whether the written statement made by Harris at the stationhouse was subject to suppression. Id. In Hudson, on the other hand, the evidence was obtained after the police did not knock and announce. Id. at 628. Breyer contends that “[t]he police's failure to knock and announce rendered the entire search unlawful, and that unlawful search led to the discovery of the evidence in petitioner’s home.” Id. (citation omitted). Lastly, Breyer distinguished Hudson from United States v. Ramirez by stating that the entry in Hudson was illegal, in contrast to that in Ramirez, which was not. Id. at 628.


130 See BALKO, supra note 10 (discussing the increasing militarism of civilian police officers and the substitution of SWAT team raids for common police work); see also Kraska & Cubellis, supra note 13, at 623 (discussing the growth of small local paramilitary police units with the local police force); Steven G. Brandl, Back to the Future: The Implications of September 11, 2001 on Law Enforcement Practice and Policy, 1 OHIO ST. J. CRIM. L. 133, 145-48 (2003) (discussing the history of policing and
potential deterrence benefits that would result from applying
the exclusionary rule to knock-and-announce violations did not
outweigh the social costs generated by its use. In justifying
this conclusion, the majority emphasized that the increased
professionalism of law enforcement officials serves to deter
police misconduct and guarantee that officers adhere to
governing procedures. In particular, the majority noted a
“new emphasis on internal police discipline,” reflected by
“increasing evidence that police forces across the United States
take the constitutional rights of citizens seriously. There have
been ‘wide-ranging reforms in education, training and supervision of police officers.’” This Note argues that the increased
internal discipline of today’s police force advocated by the
Court is not as evident as the majority suggests.

131 Hudson, 547 U.S. at 594. In Mapp v. Ohio the Court chose to extend the
exclusionary rule to the states, recognizing the importance of having effective deterrent
means to prevent police misconduct and uphold individuals’ constitutional rights. 367
U.S. 648, 660 (1961). In Mapp, the Court affirmed the assertion that “[t]he efforts of
the courts and their officials to bring the guilty to punishment, praiseworthy as they
are, are not to be aided by the sacrifice of those great principles established by years of
devote and suffering which have resulted in their embodiment in the fundamental
law.” Id. at 648 (quoting Weeks v. United States, 232 U.S. 383, 393 (1916)). Furthermore, the Court in Mapp stated that post Weeks, “This Court has ever since required of
federal law officers a strict adherence to that command which this Court has held to be
a clear, specific, and constitutionally required—even if judicially implied—deterrent
safeguard without insistence upon which the Fourth Amendment would have been
reduced to a ‘form of words.’” Id. (emphasis added) (quoting Silverthorne Lumber Co. v.
United States, 251 U.S. 385, 392 (1920)). Accordingly, in Wilson v. Arkansas the knock-
and-announce rule became constitutionally mandated that police officers must follow to
ensure an individual’s Fourth Amendment rights. See 514 U.S. at 937. Thus, it appears
that the Court in Hudson not only went against the reasoning of Mapp, but applied
deterrence mechanisms that appear ineffective. Hudson, 547 U.S. at 609-10 (Breyer, J.,
dissenting).

132 Hudson, 547 U.S. at 598.

133 Id. at 599 (quoting Samuel Walker, Taming the System: The Control
of Discretion in Criminal Justice, 1950-1990, at 51 (1993)).

134 Id.

135 See Sewell Chan, Mayor’s Report Finds a Safe and Healthy City, But One
with Troubles, Too, N.Y. TIMES, Sept. 15, 2006, at B3. For example, findings for New
York City’s Mayor’s Management Report for 2006 indicated that there were 7,373
civilian complaints against police officers for the year. This was a 16% increase from
2005, when 6,358 complaints were filed. See Graham Rayman, Cops in the Clear/Ex-
Additionally, in New York City a Civilian Complaint Review Board was established in
1993 to recommend sanctions on misconduct that is reported by the public. Id. The
agency receives a significant number of complaints regarding no-knock raids, and
interviews with former Review Board investigators reveal that no-knock complaints
have been a problem for some time. Id. Moreover, it was reported that “where
The difference in function between civilian law enforcement officials and the military is rooted in our country’s history.136 Whereas the military’s primary function was to protect our country from external enemies, law enforcement officials were charged with keeping domestic peace.137 Today, the line that separates the military and local police departments has been blurred as police officers execute a more militaristic style of law enforcement.138 This is a consequence of Special Weapons and Tactics (“SWAT”) teams or similar paramilitary units139 playing a more active role in addressing local community problems.140 As a result, local police forces

investigators tried to raise the issue, they encountered resistance from supervisors who believed such an inquiry was outside agency jurisdiction . . . .” Id. If civilians are told to report to the Review Board and the Review Board is given little power to enact change, what is the proper recourse if internal sanctions are not being—and possibly cannot be—given?

136 See Kopel & Blackman, supra note 130, at 649-55.

137 Id. In discussing the problems of the military in law enforcement functions, Col. Charles J. Dunlap noted, “military training is aimed at killing people and breaking things . . . [p]olice forces, on the other hand take an entirely different approach. They have to exercise the studied restraint that a judicial process requires; they gather evidence and arrest ‘suspects’ . . . [t]hese are two different views of the world.” Balko, supra note 10 (quoting Col. Charles J. Dunlap Jr., The Thick Green Line: The Growing Involvement of Military Forces in Domestic Law Enforcement, MILITARIZING THE AMERICAN CRIMINAL JUSTICE SYSTEM at 35).

138 Weber, supra note 13, at 11.

139 Other names for paramilitary units include: “Special Response Team (SRT), Emergency Response Team (ERT), Special Emergency Response Teams (SERT), Emergency Services Unit (ESU) . . . .” Karan R. Singh, Note, Trending the Thin Blue Line: Military Special-Operations Trained Police SWAT Teams and the Constitution, 9 WM. & MARY BILL RTS. J. 673, 680 (2001).

140 See supra note 130; see also Kopel & Blackman, supra note 130, at 649-55. The military has not always been permitted to take part in police functions. Weber, supra note 13, at 2-5. At first, the Posse Comitatus Act of 1878 was designed by Congress to prevent the intrusion of the military in civilian law enforcement affairs, and as amended in 1994, provides:

Whoever, except in cases and under circumstances expressly authorized by the Constitution or Act of Congress, willfully uses any part of the Army or the Air Force as a posse comitatus or otherwise to execute the laws shall be fined not more than $10,000 or imprisoned not more than two years, or both.

Brandl, supra note 130, at 146-47. Over time, this act was subsequently amended to include a variety of exceptions where it is proper for the military to aid the local police force. Specifically, in 1981 the “war on drugs” exception established that military personnel can be used in executing a search warrant at an individual’s home in pursuit of illegal contraband. As Congress was going through the process of amending the Posse Comitatus Act, military officials did try to warn of the harmful side effects that would result. The deputy assistant secretary for drug policy, Stephen G. Olmstead, argued to a U.S Senate subcommittee that “[o]ne of [America’s] greatest strengths is that the military is responsive to civilian authority and that we do not allow the Army, Navy, and the Marines and the Air Force to be a police force. History is replete with countries that allow that to happen. Disaster is the result.” Balko, supra note 10, at 16 (emphasis added) (alteration in original) (quoting Stephen G. Olmstead in George C.
are now conducting no-knock, military-style raids when carrying out search warrants.\textsuperscript{141} Moreover, statistics reflect that a rise in the use of these paramilitary units corresponds with an escalation of unnecessary violence and wrong-door raids.\textsuperscript{142} Accordingly, it appears the police discipline and professionalism emphasized by the Court in \textit{Hudson} is failing to limit this aggressive and violent behavior.

Wilson, Agencies Intensify Battle to Secure Key Roles In Anti Drug Effort, WASH. POST, Apr. 28, 1987). Nonetheless, the amendment was made and the military was permitted to take part in local police enforcement that involved the seizing of contraband.

Accordingly, although paramilitary units were first established in the 1960s in Los Angeles, it was not until the 1980s that these units gained momentum. Prior to the 1980s, paramilitary units were used only in the most volatile and high risk environments, such as hijackings or hostage situations. At this time, SWAT units consisted of small teams which greatly resembled police officers, only they had slightly better equipment. As the country became increasingly concerned about the proliferation of drugs since the 1970s, statistics show a corresponding rise in the use of these paramilitary units.

A study conducted by Peter Kraska on paramilitary units reflects their proliferation in today's civil law enforcement agencies: Today, 77\% of police departments surveyed have paramilitary units, which is approximately a 48\% increase since 1985. Kraska & Cubellis, \textit{supra} note 13, at 620. In 1996, Kraska and his colleague Victor Kappeler conducted a study of police departments that patrol cities of 50,000 or more citizens, and found that almost 90\% of the police departments surveyed had paramilitary units, compared to about 59\% in 1982. Brandl, \textit{supra} note 130, at 147 (548 agencies responded to the survey, a 79\% response rate). Moreover, Kraska and Kappeler documented the rise in call outs, or deployments, of these units and found that the number increased from thirteen call outs per unit in 1980 to fifty-three per unit in 1995. \textit{Id.} A call out or deployment of a unit is used in a variety of circumstances including: "barricaded persons, hostages, terrorists, civil disturbances, and the serving of a high-risk search and arrest warrant. [This] data does not included activities related to proactive patrol work by PPUs [or paramilitary units.]" Kraska & Cubellis, \textit{supra} note 13, at 614. What was even more startling was that 75\% of the call outs were for the execution of a search warrant, which typically took the form of a "no-knock" raid. Lastly, the study indicated that over the last twenty-five years, the use of SWAT team units for proactive patrol has increased nearly 300\%. \textit{Id.}

Additionally, Kraska conducted a study with Louis Cubellis to determine the increased use of paramilitary units among smaller police departments where there were approximately 25,000 to 50,000 citizens within each jurisdiction. Kraska & Cubellis, \textit{supra} note 13, at 611-12 (The study included 473 police departments, which is a response rate of 72\%). Additionally, there was an average of sixty-two police officers in each department.). These smaller jurisdictions indicated an even greater rise in the use of SWAT teams or similar agencies, as the number of paramilitary or similar units rose 157\% between 1985 and 1995. \textit{Id.} at 613. Additionally, the number of call outs increased from four and a half per year for each department in 1985, to over twelve by 1995. \textit{Id.} at 614. Furthermore, similar to the larger police departments surveyed, 66\% of the call outs were for warrant services, which typically took the form of a no-knock raid. This indicated a 342\% increase between 1985 and 1995 in the use of paramilitary units to execute search warrants. \textit{Id.} at 615.
A. What Are Paramilitary Units?

Paramilitary units are prestigious departments within the local police force with a sub culture all their own. Police officers desire to be a part of these units because the type of work these units perform is seen as exciting, high status, dangerous, and bolstering of male camaraderie. Paramilitary units are typically equipped in “black or urban camouflage BDUs (battle dress uniforms), lace-up combat boots, full-body armor, Kevlar helmets, and ninja-style hoods.” Additionally, the Department of Defense has armed these men and women with sophisticated military hardware including: “submachine guns, tactical shotguns, sniper rifles, percussion grenades, CS and OC gas (tear and pepper gas), surveillance equipment, and fortified personnel carriers.” Furthermore, these units function as “special military operation teams” that demand strict discipline and rigorous internal enforcement. In fact, elite military divisions, such as the Army Rangers and Navy Seals, commonly train paramilitary units and expose them to tactical and specialized military procedures.

The military weaponry and advanced technology that accompany paramilitary units have placed officers in a warrior-like mindset that is responsible for the militaristic attitude

---

143 Kraska & Cubellis, supra note 13, at 623.
144 Id. When observing the attitudes of police officers in paramilitary units, it becomes apparent that officers are attracted to the intense, high risk environment that is common in these units. In conducting his research, Peter Kraska interviewed two military reserve officers who were responsible for training civilian law enforcement personnel. These men stated that:

This shit [the creation of paramilitary units] is going on all over. Why serve an arrest warrant to some crack dealer with a .38? . . . With full armor, the right shit [pointing to a small case that contained a nine-millimeter Glock], and training, you can kick ass and have fun . . . Most of these guys just like to play war; they get a rush out of search-and-destroy missions instead of the bullshit they do regularly.

BALKO, supra note 10, at 17-18 (emphasis added) (alteration in original) (quoting Peter Kraska, Playing War, in MILITARIZING THE AMERICAN CRIMINAL JUSTICE SYSTEM at 143).

145 Kraska & Cubellis, supra note 13, at 610-11.
146 WEBER, supra note 13, at 7. For example, in 1997 alone, the Pentagon distributed 1.2 million pieces of military equipment to civilian police forces. BALKO, supra note 10, at 8.
147 Kraska & Cubellis, supra note 13, at 611.
148 Id. at 610.
149 See id.
that is becoming more apparent in today's police force. As one commentator notes, the use of these units advances an "over emphasis on the crime-fighting function of police work and promotes a warlike approach to crime and drug problems." For example, consider the typical method a paramilitary unit employs in executing a search warrant. In most cases, a "no-knock" raid occurs, in which police officers forcibly enter an individual's home without announcing their presence. Outfitted in military styled uniforms, multiple three-officer teams make a "dynamic entry" by using a battering ram, explosives, or similar device to forcibly break down a civilian's door. Upon entering, officers may dispense flashbang concussion grenades, break a window or possibly release a chemical spray to create a diversion. Additionally, officers are equipped with ammunition, such as automatic submachine guns, assault rifles, or 9-mm shotguns, to create fear and ensure compliance. These tactics, which turn police officers into "soldiers," produce grave consequences because "[t]he sharing of training and technology by the military and law enforcement agencies has produced a . . . mindset of the warrior . . . simply not appropriate for the civilian police officer charged with enforcing the law." Accordingly, officers forget that they are not

---

150 Weber, supra note 13, at 10. See Kraska & Cubellis, supra note 13, at 610-11, 619.
151 Kraska & Cubellis, supra note 13, at 609. Additionally, as "[d]epartmental SWAT teams have accepted the military as a model for their behavior and outlook . . . American streets are viewed as the 'front' and American citizens as the 'enemy.'" Weber, supra note 13, at 10.
152 See Balko, supra note 10, at 3-4. These no-knock raids can be pursuant to a no-knock warrant which many jurisdictions administer. Id.
153 The name "dynamic entry" is used due to the force and instantaneous matter of entry. Singh, supra note 139, at 682. Note that pursuant to Richards v. Wisconsin, a no-knock raid is lawful only if exigent circumstances are present. 520 U.S. 385, 394 (1997).
154 William Booth, Exploding Number of SWAT Teams Sets Off Alarms; Critics See Growing Role of Heavily Armed Police Units as 'Militarization' of Law Enforcement, WASH. POST, June 17, 1997, at A1 (discussing the rise of SWAT teams among local police departments); see Michael J. Bulzomi, Knock and Announce: A Fourth Amendment Standard, FBI LAW ENFORCEMENT BULL. (May 1997), available at http://www.fbi.gov/publications/leb/1997/may976.htm (discussing how law enforcement agencies can conduct "no-knock" entries in compliance with the Constitution). In cases where doors are reinforced shut, they may be removed by chaining the door to a tow truck that effectively yanks the door off. Singh, supra note 139, at 682.
155 See Booth, supra note 154.
156 See Balko, supra note 10, at 14.
157 Weber, supra note 13, at 10. See Balko, supra note 10, at 15. Balko argues that "[g]iven that civilian police now tote military equipment, get military training, and embrace military culture and values, it shouldn't be surprising when
challenging an enemy, but rather, citizens who are entitled to protection of their constitutional rights.158

B. A Consequence of Paramilitary Units: The Rise in Wrong-Door Raids

The consequences of civilian police officers acting like military personnel are untoward for both the citizens whose rights are impeded and for the officers themselves.159 The rise in the number of paramilitary units has led to an increase in “wrong-door raids,” which, as the name indicates, are forced, sometimes militaristic, entries perpetrated against innocent individuals due to police mistakes in executing a search warrant.160 Wrong-door raids are most commonly the result of misinformation by a police informant who exchanges confidential information for money or a reduced sentence in his or her own case.161 It is hard to determine just how many wrong-door

---

158 Weber, supra note 13, at 10.
159 See supra Part IV.A.
160 BALKO, supra note 10, at 26-29. See supra note 129 and accompanying text.
161 BALKO, supra note 10, at 21-25. Reports of recent wrong-door raids include the following: On September 12, 1996, seventy-year-old Ana Roman returned to her home to find police, with their guns pointed at the heads of her family, conducting a raid based on an informant’s false tip. Melissa Grace, Raid-Snafu Trial to Open Suit Blames Cops for Women’s Death 6 Yrs. Later, DAILY NEWS, suburban sec., June 1, 2004, at 3. As a result, Roman had a heart attack and spent two weeks in the cardiac unit of Lutheran Medical Center. Id. On June 25, 2003, police and FBI agents “broke down the door to an apartment of a frail man, Timothy Brockman, threw a stun grenade inside—setting a carpet on fire—then ordered him out of bed, and handcuffed him as he lay face down.” Jim Dwyer, Police Raid Gone Awry: A Muddled Path to the Wrong-door, N.Y. TIMES, June 29, 2003, at B27. The police and federal agents failed to discover that an informant had identified the wrong apartment in a housing complex. Id. A wrong-door raid took place at the apartment of Rosanna Samuel on the morning of May 24, 2001 as a result of mistaken information by an informant. Graham Rayman, Raid That Changed Her Life, NEWSDAY, June 29, 2003, at A2. Samuel was eating breakfast when officers “threw her apartment door from its hinges, dropped in a flash grenade that emitted a bright light and loud noise and burst in.” Id. Samuel, a heavyset woman, had trouble getting to the ground, so police removed her legs from under her. Samuel sustained bruises to her face, knees and elbows and was admitted to Kings County Hospital for treatment. Id. On February 27, 1998, police mistakenly raided the apartment of Shaunsia Patterson, who was eight months pregnant and living with her two- and three-year-old children. Bob Herbert, In America, Reprise of Terror, N.Y. TIMES, Mar. 12, 1998, at A27. Patterson’s fifteen-year-old sister was also home at the time. Id. The police entered her apartment in teams with their guns drawn, after a loud boom brought her door down. Id. Although visibly pregnant, Patterson stated they “threw me face down on the floor and handcuffed me behind my back . . . one of the cops stepped on the side of my face and pressed my face into the floor.” Id. She could...
raids have occurred, as police do not always keep track of their mistakes and homeowners do not consistently come forward. One study suggests, however, that almost 200 wrong-door cases have been reported in the last fifteen years, which resulted in at least forty innocent deaths. Police officials have rationalized these mistaken raids as an unavoidable outcome of the “war on drugs.”

One significant consequence of wrong-door raids is that unnecessary violence may ensue between police officers and homeowners. SWAT team raids often occur in the early morning or late evening when occupants are asleep. Moreover, officers typically do not knock and announce their presence, and diversionary tactics make it difficult for a homeowner to affirm who is at the door. Consequently, when not ask cops why they were there, because whenever she spoke, they told her to “Shut the [expletive] up!” Id. They destroyed her house in search of drugs, verbally abused her, kept her and her sister handcuffed for more than two hours, and never even showed her a warrant. Id.  

BALKO, supra note 10, at 28. In 1999, then Attorney General Janet Reno even affirmed that although the 1994 Crime Control Act required the federal government to compile information on police shootings and use of non-deadly forces, there was no law requiring that local police agencies hand over this information. Id.

Radley Balko & Joel Berger, Wrong Door, WALL ST. J., Sept. 2, 2006; see also Sure Let’s Open the Door for SWAT Teams, OAKLAND TRIB., June 22, 2006, Sports Turn at 2.

BALKO, supra note 10, at 21-26, 29. Reports of such justifications include the following: After a couple was hospitalized as a result of the police’s forcibly entry of their home thinking it was a methamphetamine laboratory, Police Chief Darryl Whaley said “[O]bviously, a mistake was made and it is regrettable . . . [b]ut I stand by my officers.”. Associated Press, Elderly Couple Hurt in Raid On Wrong House by Horn Lake Police, CLARION-LEDGER (Jackson, Miss.), Mar. 23, 2006. Whaley believed “they acted correctly and followed procedures when they entered [the couple’s] home.” Id. In Crown Heights, Brooklyn, N.Y., police mistakenly raided the apartment of Basil Shorter and his family. Michael Cooper, Scared Family Says Police Raided the Wrong Home, N.Y. TIMES, May 8, 1998, at B1. The police “broke down the door . . . tossed a stun grenade into the front hall and handcuffed everyone inside, including a mentally retarded 18-year-old girl who was taking a shower.” Id. Police asserted, however, that they did nothing wrong in following an informant’s tip that led them to the wrong apartment. Id.  

BALKO, supra note 10, at 19-20.

Id. at 19. These early morning or late evening raids contribute to violence because homeowners are more likely to think the police intrusion is a burglar attempting to rob their home. See infra notes 168-169.

BALKO, supra note 10, at 19-20. It appears that the Supreme Court is aware of the possibility that residents may act in self-defense when an unannounced entry occurs. The Court stated that one of the main purposes of the knock-and-announce rule is “protection of life and limb, because an unannounced entry may provoke violence in supposed self-defense by the surprised resident.” Hudson v. Michigan, 547 U.S. 586, 594 (2006); see also Miller v. United States, 357 U.S. 301, 313 n.12 (1958) (“Compliance is also a safeguard for the police themselves who might be mistaken for prowlers and be shot down by a fearful household.”).
police officers burst into an individual’s home the officers may be mistaken for robbers. A shoot-out then may transpire between the homeowner and officer as an individual reaches for his or her gun in an act of self-defense. As a result, innocent lives of both civilians and police officers are put at risk.

---

168 Balko, supra note 10. The following two examples illustrate instances where a resident thought police officers searching his or her home were intruders and fired at them in self-protection, leading the police to return fire. Lewis Cauthorne reportedly thought that a police search of her home for drugs was a burglary, and Cauthorne fired at four officers in an act of self-protection. Allison Klein, Courtroom Showman, a Champion of Defense, Balt. Sun, Nov. 23, 2002 at 1A. Police fired back, but luckily, no one was fatally injured. Id. In another case, police raided Ellis Elliot’s Bronx, N.Y., apartment based on information from a mistaken informant. Kit R. Roane, Bronx Man Recounts Abuse by Police in Mistaken Raid, N.Y. Times, Mar. 4, 1998, at B5. The raid occurred at eight a.m. when Elliot was sleeping. Id. Not knowing who was at the door, Elliot grabbed his gun, told whoever was at the door to move away and when they did not, he fired his gun. Id. Elliot alleges that nine police officers fired back twenty-six bullets into his apartment, destroyed his home, screamed racial epithets at him and beat him. Id. As a result, Elliot sued the city. Peter Noel, NYPD Storm Troopers, Village Voice, May 16, 2000, at 27.

169 Balko, supra note 10, at 19-29. The following occurrences depict the unfortunate fatalities that can occur as a result of wrong-door raids. On March 25, 1994, a wrong-door raid took place at the home of Rev. Accelyne Williams. Robyn E. Blumner, Court Signals Loosening of the Last Reins of Police, St. Petersburg Times, June 25, 2006, at 4. As a result of incorrect information given by an informant, thirteen members of the Boston SWAT team forcibly entered Williams’s apartment. Id. Williams, age seventy-five, subsequently died of a heart attack after struggling with the SWAT team members. Id. On March 26, 2001, police executed a warrant at a duplex building. Jerry Mitchell, Was There Justice For, Clarion-Ledger (Jackson, Miss.), Mar. 5, 2006, at A1. The warrant listed the name of Jamie Smith, who occupied one apartment, but not Cory Maye, who occupied the other. Id. Police alleged that they knocked and announced their presence, but received no answer and then broke down the door. Id. Maye was sleeping at the time with his eighteen-month-old daughter and asserted that he did not hear the announcement. Id. Maye grabbed his gun in an act of self-defense and fired at Officer Ron Jones, causing fatal injuries. Id. Maye, who had no prior criminal record, was sentenced to death. Id. On May 16, 2003, misinformation by an informant caused police to raid the Harlem, New York, home of fifty-seven-year-old Alberta Spruill. Diane Cardwell & William K. Rashbaum, City Officials Suggest a Shot in Police Raid was Accidental, N.Y. Times, Aug. 6, 2003, at B1. Suprill died of a heart attack after a police team used a flashbang grenade to forcibly enter her home. Id. Likewise, police mistakenly raided John Adams’ home, when they were in fact supposed to execute a search warrant at the home next door. Warren Duzak, Chief Addresses Fatal Errors, The Tennessean, Oct. 7, 2000, at A1. The officers apparently knocked but did not identify themselves, and then they forcibly entered Adams’s home. Id. Unaware of who was at the door, Adams met them with a shotgun. Id. The officers claimed that Adams fired shots at them and they retaliated, but the question of whether Adams used his gun was disputed. Id. Adams died as a result of the police gunshots. Id. On September 29, 1999, the Denver police executed a wrong-door raid on Ismael Mena’s home based on misinformation by an informant. Bruce Finley, Mena’s Farm Dreams Turned to Dust Talks Start Today in No-Knock Death, Denver Post, Mar. 16, 2000, at A1. Mena, awoken by the police forcibly entering his apartment, thought he was being robbed and fired three shots from his own gun. Id. As a result, police fired eight bullets into Mena’s face, chest and arms, killing him. Id.
Evidence suggests that the increased use of paramilitary units will continue in the future.\textsuperscript{170} When this disturbing trend is noted in light of \textit{Hudson}'s weakening of the knock-and-announce requirement, it certainly appears that the problems of wrong-door raids and excessive police violence will become more commonplace.\textsuperscript{171} Accordingly, it is imperative to have guidelines in check that will curb officers' behavior and guarantee that they act in accordance with the law. By not subjecting knock-and-announce violations to the exclusionary rule, there are no real incentives to ensure that officers comply with knock-and-announce procedures.\textsuperscript{172} It is not enough to rely on the internal discipline of law enforcement officials, as there is clear evidence that wrong-door raids are occurring and that innocent lives have been lost.\textsuperscript{173} \textit{Hudson} presented the Court with the opportunity to create a true deterrence mechanism to limit this aggressive and militaristic conduct. If officers simply knocked and announced, then homeowners would know who was at their door and would be better suited to deal with police presence. By failing to uphold the exclusionary rule for violations of the knock-and-announce rule, the Court gave officers an incentive to ignore its prohibition against unannounced searches. As a result, this Note argues that the exclusionary rule must be extended to knock-and-announce violations to make certain that an effective deterrence mechanism is in place that guarantees that police officers are respecting individuals' constitutional rights when executing search warrants.\textsuperscript{174}

\textsuperscript{170} Brandl, \textit{supra} note 130, at 146-47; \textit{see also} Kopel & Blackman, \textit{supra} note 129, at 649-55.


\textsuperscript{172} \textit{See infra} Part V.B.1.

\textsuperscript{173} \textit{See supra} note 169.

\textsuperscript{174} \textit{See} Kopel & Blackman, \textit{supra} note 130, at 649-55. In stressing the importance of protecting against the increased militarization of the police, Kopel and Blackman cited a powerful quote from the Eighth Circuit opinion, \textit{Bissonette v. Haig}:

Civilian Rule is basic to our system of government. The use of military forced to seize civilians can expose civilian government to the threat of military rule and suspension of constitutional liberties. \textit{On a lesser scale, military}
V. A LOOK INTO THE FUTURE

The studies above indicate that the professionalism of the police force is not as effective a deterrent as the Hudson majority suggests. Rather, there appears to be a rise in police misconduct stemming from the increased use of paramilitary units and resulting in an alarming number of wrong-door raids. Consequently, it is imperative that the Court take measures to rectify its decision in Hudson and create real incentives that will deter unlawful police behavior. This need for action becomes especially apparent when comparing the knock-and-announce rule to the Miranda warnings.

In 1966, in Miranda v. Arizona, the Supreme Court held that police officers must inform suspects of their right to remain silent before proceeding with custodial interrogation in order to ensure that confessions were voluntary and therefore admissible at trial. Due to a number of subsequent exceptions that the Court made to the Miranda rule, law enforcement officials felt that the original version of Miranda was substantially weakened. As a result, certain states began to train police to question outside Miranda by instructing officers to continue interrogating suspects even after they invoked their Miranda rights. Hence, Fifth Amendment protections

\[\text{enforcement of the civil law leaves protection of vital Fourth and Fifth Amendment rights in the hands of persons who are not trained to uphold these rights.}\]

Id. at 621 (emphasis added) (quoting Bissonette v. Haig, 776 F.2d 1384, 1387 (8th Cir. 1985)); see also Case Comment, Fourth Amendment—Exclusionary Rule—Seventh Circuit Holds that the Suppression of Evidence Is a Disproportionately Severe Sanction for Timing Violation of the Knock-and-Announce Requirement.—United States v. Espinoza, 256 F.3d 718 (7th Cir. 2001), 115 HARV. L. REV. 709, 713-15 (2001) (arguing that evidence seized in violation of the knock-and-announce requirement needs to be suppressed under the exclusionary rule).

175 See supra Part IV.
176 See supra Part IV.
177 384 U.S. 436 (1966). Before suspects in custody are interrogated, they are read their Miranda Rights, consisting of four warnings: that a suspect “has the right to remain silent, that anything he says can be used against him in a court of law, that he has the right to the presence of an attorney, and that if he can not afford an attorney one will be appointed for him if he so desires.” Id. at 479. Accordingly, if evidence is seized in violation of Miranda, its use is suppressed at trial. See United States v. Dickerson, 530 U.S. 428, 435 (2000).
178 See Weisselberg, Saving Miranda, supra note 15, at 109 (arguing that the original vision of Miranda has been transformed, and as a result, police departments have begun to promote a policy of questioning outside Miranda). Officers were being trained that “it is permissible to question suspects who invoked the right to counsel or the right to remain silent.” Id. at 132. For example a training video instructed police officers that:
dissolved and an individual’s right to remain silent became increasingly disrespected.179

Likewise, as noted by many commentators, the Court’s decision in *Hudson* has created incentives for police to violate knock-and-announce guidelines.180 It therefore appears vital

When you violate *Miranda*, you’re not violating the Constitution. *Miranda* is not in the Constitution. It’s a court-created decision that affects the admissibility of testimonial evidence and that’s all it is . . . [s]o you’re not doing anything unlawful, you’re not doing anything illegal, you’re not violating anybody’s civil rights, you’re doing nothing improper.

Id. at 110 (quoting Videotape: Questioning “Outside *Miranda*” (Greg Gulen Productions 1990)).

179 Id. at 149. The Fifth Amendment states that:

No person shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a Grand Jury, except in cases arising in the land or naval forces, or in the Militia, when in actual service in time of War or public danger; nor shall any person be subject for the same offence to be twice put in jeopardy of life or limb; nor shall be compelled in any criminal case to be a witness against himself, nor be deprived of life, liberty, or property, without due process of law; nor shall private property be taken for public use, without just compensation.

U.S. CONST. amend. V. This section of the Note will focus on the Fifth Amendment’s self-incrimination clause which prohibits police from compelling a suspect to incriminate him or herself.

180 See, e.g., Milton Hirsch, *Hudson v. Michigan: Whose Fourth Amendment Is It, Anyway?*, THE CHAMPION, Aug. 2006, at 50-51 (stating that, as a result of the majority’s opinion, “if the police do come crashing through your front door without giving the householder a fair chance, or any chance, to open the door, the exclusionary rule doesn’t apply”). Hirsch argues that there is no remedy when police officers violate the knock-and-announce rule because

[the conclusion is inescapable that but one remedy exists to deter violation of the search and seizure clause. That is the rule which excludes illegally obtained evidence. Only by exclusion can we impress . . . that violation of the Constitution will . . . do no . . . good. And only when that point is driven home can the prosecutor be expected to emphasize the importance of observing constitutional demands in his instructions to the police.

Id. (quoting Wolf v. Colorado, 338 U.S. 25, 44 (1949) (Murphy, J., dissenting)). Hirsch is not alone in holding these sentiments. See Cathy Young, Editorial, *Exclusionary Rule Sends Dangerous Message*, SEATTLE POST-INTELLIGENCER, July 9, 2006, at D2 (arguing that “leaving the exclusionary rule intact but exempting no-knock searches from its scope sends a dangerous message that for police to burst into a citizen’s house unannounced is no big deal.”); Elaine Silvestrini & Valerie Kalfin, *Ruling Unlikely to Alter Police Searches*, TAMPA TRIBUNE, June 16, 2006 at 1 (quoting defense attorney John Fitzgibbons, “Police now [after *Hudson*] have a great deal of leeway and this simply continues the trend in weakening the Fourth Amendment”); Stephanie Francis Ward, *Court Backs Evidence Found in “Knock-Announce” Case: Justices’ 5-4 Decision Narrows Exclusionary Rule in Police Searches*, ABA J. E-REP., June 16, 2006, at 1 (quoting Timothy Lynch, Director of the Cato Institute’s Project on Criminal Justice, “Here was an opportunity for the court to put the brake on [paramilitary style raids], and say, ‘slow things down,’ but they didn’t . . . . This is a kind of weakening of the exclusionary rule, and our fear is that the brake which was needed has not been applied.”); see also *Hudson v. Michigan*, 547 U.S. 586, 605 (2006) (Breyer, J.,
that effective deterrence mechanisms be put in place before police training changes in response to a significantly watered down knock-and-announce rule. The institutionalization of questioning outside *Miranda* serves as a warning that if nothing is done, officers may be taught how to effectively get around the knock-and-announce requirement.

A. Potential Problems If Hudson Is Not Rectified:
A Comparison to Questioning Outside Miranda

In 1995, training materials from several California police departments revealed the promotion of a “new vision of *Miranda*” that encouraged officers to continue questioning suspects even after they invoked their Fifth Amendment rights. This technique was referred to as questioning “outside *Miranda*” and was advocated by police officials who believed that Court exceptions had transformed the original version of the *Miranda* rule. For example, in *Harris v. New York*, the Court held that statements obtained in violation of the *Miranda* warnings could still be used at trial for impeachment purposes, as long as trustworthiness was shown. As a result of this exception, officers realized that even if they violated *Miranda*, a suspect’s statements and other valuable evidence could still be admissible at trial, just not in the prosecution’s case-in-chief. Accordingly, exceptions like *Harris* led police to disregard and circumvent the *Miranda* warnings by creating new guidelines that would allow officers to question suspects even after the suspects invoked their *Miranda* rights.

---

181 In addition to those in California, police officers in Arizona, Colorado, the District of Columbia and Missouri have all been found to question suspects “outside *Miranda*.” Sandra Guerra Thompson, *Evading Miranda: How Seibert and *Patane* Failed to “Save” *Miranda*,* 40 VAL. U. L. REV. 645, 670 (2006) (“[E]xact numbers are not available, but evidence suggests the practice has been spreading.”).


183 Thompson, *supra* note 181, at 649.


185 Evidence could not be used in the prosecution’s direct case, but could be used for other purposes such as to impeach or cross-examine a witness. Weisselberg, *Saving Miranda*, *supra* note 15, at 134.

Professor Charles Weisselberg drew attention to this new practice by publicizing training manuals that informed officers that “the warning and waiver components of *Miranda* were simply a court-created ‘series of recommended procedural safeguards that were not themselves rights protected in the Constitution.” Moreover, officers received instructions that as long as they “avoid overbearing tactics that offend the Fourteenth Amendment due process, the mere fact of deliberate noncompliance with *Miranda* does not affect admissibility for impeachment . . . officers risk no civil liability for ‘benign’ questioning outside *Miranda*. Instead, they have ‘little to lose and perhaps something to gain.” Thus, this “new vision of *Miranda*” taught officers that adherence to the *Miranda* warnings was optional. Officers realized they had a choice. They could abide by the *Miranda* rule and cease questioning suspects once they invoked their Fifth Amendment rights, or they could violate the rule and the prosecution would only receive a minimal penalty of having the evidence excluded from their case-in-chief. As officers began to do the latter, the protections of the Fifth Amendment were significantly undermined and little respect was given to individual autonomy and Supreme Court authority. As Weisselberg noted, “the Court could not have intended to give police grounds to disobey this

---

187 Id. at 133 (quoting a California District Attorneys Association training bulletin). Weisselberg’s article gives examples from training bulletins, manuals, and videos that reflected California’s widespread practice of training officers to go outside *Miranda*. See id. at 133-36.

188 Id. at 133-34.


190 Id.; see Charles D. Weisselberg, *In the Stationhouse After Dickerson*, 99 Mich. L. Rev. 1121, 1133 (2001) [hereinafter Weisselberg, *After Dickerson*]. When being trained outside *Miranda*, officers are taught “there is nothing legally or morally wrong in interrogating a suspect who has invoked the right to counsel or the right to remain silent. Questioning over an invocation merely has an evidentiary consequence at trial.” *Id.* at 1124-25. Additionally, even though the statement seized can not be used in the prosecution’s case-in-chief, officers are trained that “you can accomplish all of these legitimate purposes that don’t have anything to do with the prosecution of the case, and some that do, by talking to the guy ‘outside *Miranda*.’” *Id.* at 1125 (quoting Videotape: Questioning: ‘Outside *Miranda*’ (Greg Gulen Productions 1990)). As Weisselberg suggests, “Officers trained in this fashion perceive no downside to questioning ‘outside *Miranda*.’ Investigators who respect an invocation . . . will obtain no information from a suspect. On the other hand, questioning over an invocation may yield useful information, even if that information has a limited use at trial.” *Id.*

191 Weisselberg, *Saving Miranda*, supra note 15, at 168; Thompson, supra note 181, at 648 (discussing how “the police increasingly ignore *Miranda*”).
portion of Miranda deliberately, but this disregard is the natural consequence of these decisions.⁷¹⁹²

Although the Court did not intend to create a framework in which officers would be trained on how to circumvent the Miranda warnings, through its decisions the Court transformed the established rule. As a result of the Court’s Miranda exceptions, the “unintended consequence . . . [of] these restrictions sent[ted] a clear message that many constitutionally defective evidence-gathering acts will go unpunished. Some police departments have internalized this news as conferring a ’green light’ to lawless action.” ¹⁹³ Likewise, as the Court’s decision in Hudson has significantly weakened the knock-and-announce rule, there is now the legitimate possibility that officers will be trained in a new manner that will allow them to circumvent knock-and-announce guidelines completely. When the Court weakened the link of the Miranda warnings to the Constitution, commentators noted that “by alienating Miranda’s rule from the Fifth Amendment, the Justices have undermined its legitimacy.” ¹⁹⁴ Accordingly, the same holds true in the knock-and-announce context.

¹⁹² Weisselberg, Saving Miranda, supra note 15, at 132 (emphasis added).
¹⁹³ Davies, supra note 189, at 1319. For some time, the Miranda warnings were considered a non-constitutional rule of evidence. Weisselberg argues that the Court’s opinion in United States v. Dickerson, which reaffirmed Miranda as a constitutional rule, has resulted in a change of police policies and training as some police departments in California are now being trained not to violate the Miranda rule by questioning outside Miranda. Weisselberg, After Dickerson, supra note 190, at 1122. But see Thompson, supra note 181, at 653. Thompson’s article seems to suggest that going outside Miranda is still a technique used by law enforcement officials. Id. She argues that Missouri v. Seibert and United States v. Patane, both decided in 2004 after Dickerson, have resulted in police questioning outside Miranda. Id. (stating that this just adds to the already “diminished . . . possibility that Miranda might play even a moderately effective role in reducing the coercive atmosphere in the interrogation room”). Seibert involved the practice of “questioning first.” See id. at 646. Under this practice, an officer does not give the Miranda warnings until after a defendant confesses, and then the officer attempts to obtain the confession a second time, which would be admissible because it followed the warnings. See id. The officer in Seibert admitted that he purposely withheld the Miranda warnings, and the Court acknowledged the practice of police circumventing Miranda guidelines. Missouri v. Seibert, 542 U.S. 600, 605-06 (2004). A divided Court held that the statements were inadmissible, and Justice Kennedy’s concurrence (which provides the holding of the case) devised a two-part test which appears to do little to solve the problem. Id. at 618-22; Thompson, supra note 181, at 672, 677-81. Additionally, in United States v. Patane, the Court held that “the ‘physical fruit’ of a voluntary statement—in this case, a gun—is admissible even if the statement is given without a sufficient Miranda warning[].” Susan R. Klein, Enhancing the Judicial Role in Criminal Plea and Sentence Bargaining, 84 TEX. L. REV. 2023, 2032 n.43 (2006) (citing United States v. Patane, 542 U.S. 630, 643 (2004)).
¹⁹⁴ Weisselberg, Saving Miranda, supra note 15, at 130. As Weisselberg notes:
This Note argues that, just as officers saw a weakened *Miranda* rule as the “go ahead” to circumvent the established law, officers may now be influenced to evade the knock-and-announce rule since violations will not always result in the suppression of evidence. Instructive similarities can be drawn between the knock-and-announce rule and the *Miranda* warnings. First, the knock-and-announce rule and the *Miranda* rule are both common law principles that are not explicitly written in the Constitution, but are directly linked to the Constitution through the Fourth and Fifth Amendments respectively. Because the Court has held that violations of the Fourth, Fifth, and Sixth Amendment are subject to the exclusionary rule, it is in the Court’s discretion to suppress the evidence that is obtained in violation of the knock-and-announce rule and the *Miranda* rule at trial. Currently, only

---

Id. at 168-69 (alteration in original) (quoting Commission on Peace Officer Standards & Training, Cal. Dep’t of Justice, Interrogation Law Instructors’ Outline, at 21 (1996)).

195 See *Miranda* v. Arizona, 384 U.S. 436, 443 (1966); *Wilson* v. Arkansas, 514 U.S. 927, 936 (1995) (establishing that the knock-and-announce rule is mandated by the Constitution and part of the Fourth Amendment); *Weisselberg, Saving Miranda*, supra note 15, at 110 (suggesting the link of the *Miranda* rules to the Fifth Amendment self-incrimination clause); see also *Dickerson* v. United States, 530 U.S. 428, 434-35 (2000) (reaffirming that *Miranda* is and has always been a constitutional rule). For some time, *Miranda* warnings were considered a “prophylactic,” non-constitutional rule, which further supported the idea that officers could go outside *Miranda* and continue to interrogate suspects after they invoked their Fifth Amendment rights. *Weisselberg, Saving Miranda*, supra note 15, at 128. However, in *Dickerson*, the Court noted this concern and held that “*Miranda* is a constitutional decision.” 530 U.S. at 437.

196 The *Exclusionary Rule*, supra note 42, at 186-87 (“Under the exclusionary rule, evidence obtained in violation of the Fourth, Fifth, or Sixth Amendments may not be introduced at trial for the purpose of proving the defendant’s guilt.”). The Supreme Court recognized in *Boyd v. United States* the similarities between the Fourth and Fifth Amendments, stating that they “run almost into each other.” 116 U.S. 616, 630 (1886) (quoted in *Mapp v. Ohio*, 367 U.S. 643, 646 (1961). *Boyd* also noted, “The principles laid down in this opinion . . . . apply to all invasions on the part of the government and its employés of the sanctity of a man’s home and the privacies of life.” *Id.*
what commentators refer to as a “weak exclusionary rule” is attached to Miranda. Second, in establishing the necessity of both the knock-and-announce requirement and the Miranda warnings, the Court set forth limitations on the implementation of these rules. Over time, the Court has expanded these limitations and continued to carve out exceptions to both rules, which has weakened their application and created incentives for police to violate their guidelines.

197 Weisselberg, After Dickerson, supra note 190, at 113 (arguing that a weaker exclusionary rule is applied to Miranda violations than to Fourth Amendment violations and “other types of compelled testimony and . . . statements that violate the Fourteenth Amendment”); see Thompson, supra note 181, at 648 (suggesting that the Court does not strictly apply the exclusionary rule to Miranda violations).

198 In Wilson, the Court determined that certain circumstances would reasonably justify a police officer’s unannounced entry, and left the lower courts to decide when those conditions were present. Wilson, 514 U.S. at 936. Likewise, the Court did not apply the Miranda rules to all interrogations without counsel, just those that took place while a suspect was in custody. Moreover, the Miranda Court permitted states to implement their own or additional procedures as long as they would protect an individual’s right against self-incrimination. Miranda, 384 U.S. at 490.

199 In regard to the Miranda rules, the Court’s first exception was made in Harris v. New York, in which the Court held that “prosecutors can use for impeachment purposes a statement from a suspect in custody who did not receive Miranda warnings so long as ‘the trustworthiness of evidence satisfies legal standards’ . . . .” Weisselberg, Saving Miranda, supra note 15, at 127. This allows officers to use statements that were obtained without proper Miranda warnings to prove a defendant’s guilt for all purposes, except the government’s case-in-chief. For example, the statement can be used against a defendant for impeachment or cross-examination purposes. Id. Additionally, the Court held in Michigan v. Tucker “that Miranda’s exclusionary rule does not apply to the ‘fruits,’ evidence derived from statements obtained in violation of Miranda.” Thompson, supra note 181, at 648. This allows the fruits of statements to be admissible where questioning preceded Miranda, but trial followed. Moreover, in New York v. Quarles, the Court created a “public safety exception” under which “warnings are not necessary when officers ask a suspect questions arising from a reasonable concern for public safety.” Weisselberg, Saving Miranda, supra note 15, at 129. The Court asserted that a cost-benefit analysis should be applied here, and if the circumstances are such that the need for answers in order to protect public safety outweighs the protection guaranteed by the Fifth Amendment, the failure to give warnings is justified. Id. These three exceptions led police to train outside Miranda, which is important for the purposes of this Note. However, other exceptions were established which further weakened the Miranda rule, such as that in United States v. Patane, which “broadens the long-standing rule that physical evidence derived from Miranda violations is freely admissible even as part of the government’s case in chief, clarifying that even intentional violations may yield admissible fruits.” Thompson, supra note 181, at 648. In Patane, the Court held that “the ‘physical fruit’ of a voluntary statement—in this case, a gun—is admissible even if the statement is given without a sufficient Miranda warnings.” Klein, supra note 193, at 2032 n.43 (citing United States v. Patane, 542 U.S. 630, 643 (2004)).

The Supreme Court has also created exceptions to the knock-and-announce rule that have weakened its application and use. See supra Part II.A. In Richards v. Wisconsin, the Court held that officers are not required to knock and announce when exigent circumstances are present. Thus, where police have a reasonable suspicion that knocking and announcing would be “dangerous or futile,” or would “inhibit the effective investigation of the crime by, for example, allowing the destruction of evidence,” a no-
As a result of the Court's decision not to apply the exclusionary rule to knock-and-announce violations, the Court effectively took away the only major legal incentive that police had to abide by its procedures. Officers are now aware that whether or not they obey the law and follow the knock-and-announce requirement, the evidence seized will still be admissible at trial. Thus, the need for a legal incentive to deter police misconduct and guarantee compliance with the law is imperative. In order to ensure that officers abide by the Fourth Amendment guarantees and protections, the exclusionary rule must be applied to knock-and-announce violations.

B. The Only Plausible Solution: The Exclusionary Rule as an Effective Remedy

It is necessary for the Court to apply the exclusionary rule to knock-and-announce violations in order to ensure that an effective deterrence mechanism is in place. Although the Court in *Hudson* determined that the costs of applying the exclusionary rule outweigh the benefits, the increased militarization of the police force and the very real possibility that

knock entry is lawful. 529 U.S. 385, 394 (1997). Additionally, in *United States v. Banks*, the Court determined that when police believe that destruction of the evidence is possible, fifteen seconds is a reasonable amount of time before entering an individual's home. 540 U.S. 31, 40 (2003). Although this may seem like a long time, for individuals who are sleeping or on the couch in another room, this may not present them with enough time to get to the door before police forcibly enter. See BALKO, supra note 10, at 32. Lastly, the Court's decision in *Hudson* appears to be the most significant decision concerning the knock-and-announce rule, as it established that evidence that is seized in violation of the knock-and-announce rule can be used at trial to determine a defendant's guilt. Hudson v. Michigan, 547 U.S. 586, 594 (2006). Additionally, in *Hudson* the police only waited three to five seconds before entering Hudson's home and although Hudson appealed his conviction based on this premature entry, the Court determined that suppression was not required. *Id.* at 2162. This seems to suggest that in future cases where police only wait three to five seconds before entering, no penalties will be applied.

Before the Court's decision in *Hudson*, courts did apply the exclusionary rule to knock-and-announce violations. For example, in *United States v. Dice*, the Sixth Circuit stated that the exclusionary rule is the appropriate remedy for knock-and-announce violations. In coming to this conclusion, the court determined that "[t]o remove the exclusionary bar from this type of knock-and-announce violation whenever officers possess a valid warrant would in one swift move gut the constitution's regulation of how officers execute such warrants." *United States v. Dice*, 200 F.3d 978, 986 (6th Cir. 2000) (emphasis added). Moreover, the Eighth Circuit opinion in *United States v. Marts* also applied the exclusionary rule to knock-and-announce violations when officers did not wait a reasonable amount of time before entering. 986 F.2d 1216, 1220 (8th Cir. 1993).

See infra notes 212-221 and accompanying text.

200 *Hudson*, 547 U.S. at 594-95.
police training could change as a result of weakened knock-and-announce guidelines make clear that the costs of not applying the exclusionary rule are indeed significant. Thus, the Court must act to rectify its decision in *Hudson* and reinstate the authority of the knock-and-announce rule.203

1. Other Deterrence Mechanisms Do Not Deter

As the need to preserve individuals’ Fourth Amendment rights and promote lawful police activity is imperative, a forceful deterrent must be put in place. The only effective way to create such a deterrent is through the legal threat of the exclusionary rule. In *Hudson*, the majority relied on implausible and ineffective deterrence mechanisms, leaving officers little incentive to comply with knock-and-announce procedures.204 Specifically, the majority determined that the increased professionalism of law enforcement officials and the threat of civil suits will prevent officers from violating the knock-and-announce rule.205 Although these safeguards may be available in theory as the Court suggests, in reality they do not serve to deter police misconduct and unlawful behavior.

First, as already explained, the rise of paramilitary units within local police departments has resulted in a militaristic style of law enforcement and detrimental behavior that debunks the Court’s notion that police departments have become increasingly professionalized.206 Second, evidence suggests that a civil suit against a police officer is not a viable remedy when an individual’s rights are violated by an officer’s failure to knock and announce.207 Not only do the obstacles faced by one who seeks to bring a suit against a police officer make it unlikely that a victim of police misconduct will succeed

203 Although stated in regards to *Miranda*, this statement proves worthwhile here: “If supervisors wish to imbue respect for *Miranda* [in this case the knock-and-announce rule], they must themselves take [it] seriously and signal that they prefer their officers to honor an invocation, even when doing so means losing an opportunity to gain useful information or evidence.” Weisselberg, *After Dickerson*, supra note 190, at 1156.

204 See *Hudson*, 547 U.S. at 605 (Breyer, J., dissenting).

205 *Id.* at 595, 598-99 (majority opinion). See supra text accompanying notes 92-95.

206 See supra Part IV.

207 See *Hudson*, 547 U.S. at 611-14 (Breyer, J., dissenting). This too has proven to be an ineffective remedy to ensure that officers follow knock-and-announce guidelines because there are many obstacles that a lawyer and his client will encounter when bringing a civil rights suit under 42 U.S.C. § 1983 against law enforcement officials.
in court,\textsuperscript{208} but even in the rare case that a civil suit survives summary judgment, it is very unlikely that substantial monetary sanctions will be granted.\textsuperscript{209} With only these weak

\textsuperscript{208} See generally Criminal Practice Guide, supra note 123 (discussing how lawyers should proceed in criminal cases as a result of the Court’s decision in \textit{Hudson}). First, most citizens lack the knowledge that civil rights suits are available and, moreover, they do not have the necessary resources to initiate such a suit. \textit{Id.} Even if an individual has the means to initiate a suit, consider a lawyer’s argument when asking the jury to find in favor of his or her client. \textit{See} Hirsch, supra note 180. Hirsch offers the following hypothetical opening statement to illustrate the difficulties a lawyer would face when his or her guilty client seeks to bring a civil rights suit for a knock-and-announce-violation:

\begin{quote}
\textit{In this case, the police obtained a warrant to search my client’s home for drugs. They searched . . . found the drugs . . . my client was convicted for possession of those drugs . . . which is why he appears before us in court today in handcuffs, leg manacles . . . Pay no attention to those things. The fact is that the police, in executing their search warrant, smashed my client’s door down. Those door hinges were completely destroyed! They were really expensive hinges . . . . [s]o we’re going to ask for one million dollars in damages, to teach those police officer to respect [] constitutional rights . . . .}
\end{quote}
\textit{Id.} As this example suggests, lawyers will have a difficult time convincing a jury to find in favor of their otherwise guilty client. The qualified immunity defense is another hurdle that must be overcome to succeed in a civil rights claim against law enforcement officials for violating the knock-and-announce rule. See \textit{supra} note 123 (explaining the qualified immunity defense). In knock-and-announce suits, a court will often find in favor of an officer on the qualified immunity defense, concluding that the officer had a reasonable belief that a forced entry was necessary to protect him- or herself against harm. Urbonya, \textit{supra} note 123, at 104; \textit{see, e.g.}, Leaf v. Shelnutt, 400 F.3d 1070, 1080-83 (7th Cir. 2006) (granting qualified immunity to an officer after he did not knock and announce and used guns and tactical lights to gain entry into plaintiff’s apartment). Moreover, courts have found that when the question is a close one, “the officers are entitled to the benefit of doubt under the qualified immunity standard.” \textit{Id.} (emphasis added) (internal quotation marks omitted) (citing Dickerson v. McLellan, 101 F.3d 1151, 1160 (6th Cir. 1996)).

\textsuperscript{209} See Brief for Petitioner at *37, \textit{Hudson}, 547 U.S. 586 (No. 04-1360) (alleging that there are no cases cited in the opposition briefs that show the award of tangible damages as a result of knock-and-announce violations); \textit{see, e.g.}, Doran v. Eckold, 409 F.3d 958, 960-72 (8th Cir. 2004) (reversing a jury award of over $2 million for an officer’s illegal no-knock entry into the defendant’s home, holding that the entry was reasonable under the Fourth Amendment). In an effort to emphasize this point, it is helpful to examine Michigan’s state law prior to \textit{Hudson}. In 1999, the Michigan Supreme Court held, in \textit{People v. Stevens}, that the exclusionary rule is not a valid remedy for knock-and-announce violations. 597 N.W.2d 53, 64 (1999). Rather, Michigan relied on federal and state incentives that prohibit police misconduct and allow individuals to bring private claims against officers if a civil rights violation occurs. \textit{Id. As Hudson was being argued, Michigan’s law was examined to determine if damages were in fact granted for knock-and-announce violations. Hudson}, 547 U.S. at 610 (Breyer, J., dissenting). Not only did a brief in support of the petitioner assert that it could not find any decisions that awarded actual damages in knock-and-announce violations, but the respondent’s lawyer, Mr. Timothy Baughman, conceded this point during oral argument before the Supreme Court. See Brief for Petitioner at *59, \textit{Hudson}, 547 U.S. 586 (No. 04-1360); Transcript of Oral Argument, Jan. 9, 2006, at *31-32, \textit{Hudson}, 547 U.S. 586 (No. 04-1360) [hereinafter \textit{Hudson} Transcript]. Accordingly, Mr. Baughman stated that he was not aware of any successful § 1983 actions in Michigan since the \textit{Stevens} decision, which took place five years prior. \textit{Hudson}
deterrence mechanisms in effect, officers know that whether or not they follow knock-and-announce procedures the likelihood of sanction is slight and the evidence seized will still be admissible at trial.

2. Why the Exclusionary Rule? Because It Actually Deters

It is crucial that the Court reconsider Hudson and put deterrence mechanisms into effect that will actually prevent unlawful police behavior. In holding that evidence should not be subject to suppression and, subsequently, not condemning officers for forcibly breaking into Hudson’s home after only a three-to-five second waiting period, the Hudson majority sent the message that knocking-and-announcing is just a futile gesture that does not need to be respected. Thus, as this Note argues, only through the legal sanction of the exclusionary rule

**Transcript, supra, at *31-32.** Subsequently, in oral arguments before the Supreme Court in May 2006, Mr. Baughman restated that he was still not aware of any Michigan case where a civil judgment was rendered against law enforcement officials for knock-and-announce violations. **Id. at *36-38.** Additionally, an extensive search of case law revealed only one unpublished decision where monetary damages were actually awarded. **See** Buss v. Quigg, No. 01-CV-3908, 2000 LEXIS 19324 (E.D. Pa. Oct. 9, 2002). In Buss, it appears that the plaintiff collected nominal damages of $1 after officers failed to knock and announce. Thus, even in the limited cases in which damages are in fact awarded, they seem to be nominal. This finding is supported by Justice Breyer in his dissent, condemning the majority’s reliance on civil suits: “The majority . . . has failed to cite a single reported case in which a plaintiff has collected more than nominal damages solely as a result of a knock-and-announce violation.” Hudson, 547 U.S. at 610 (Breyer, J., dissenting). Breyer continues, “[T]he majority, as it candidly admits, has simply ‘assumed’ that ‘[a]s far as [it] know[s], civil liability is an effective deterrent,’ a support-free assumption that Mapp and subsequent cases make clear does not embody the Court’s normal approach to difficult questions of Fourth Amendment law.” **Id. at 611** (citation omitted).

**210** The decision seems to suggest that in future cases where police only wait three to five seconds before entering an individual’s home, no penalties will be applied. Accordingly, the whole purpose of the knock-and-announce rule appears to be undercut as the Court has stated that two of the main purposes of the knock-and-announce rule are

…the protection of human life and limb, because an unannounced entry may provoke violence in supposed self-defense by the surprised resident. . . . [and the protection of] those elements of privacy and dignity that can be destroyed by a sudden entrance. It gives residents the “opportunity to prepare themselves for” the entry of the police. “The brief interlude between announcement and entry with a warrant may be the opportunity that an individual has to pull on clothes or get out of bed.”

**Hudson, 547 U.S. at 594** (citations omitted) (quoting Richards v. Wisconsin, 520 U.S. 385, 393 n.5 (1997)). Surely, three to five seconds is not a sufficient amount of time to guarantee that an individual receives the protections afforded by the knock-and-announce rule.
will officers be compelled to abide by knock-and-announce procedures for fear that vital evidence will be suppressed at trial.\footnote{See infra notes 212-221 and accompanying text.}

In determining the effectiveness of the exclusionary rule, some research suggests that a majority of officers believe the exclusionary rule deters misconduct, and many are convinced that the threat of suppression results in better law enforcement. For example, a study conducted of Chicago’s judges, police officers, public defenders, and prosecutors by Professor Myron W. Orfield supports the idea that the exclusionary rule does deter police misconduct.\footnote{See Myron W. Orfield, Jr., Deterrence, Perjury, and the Heater Factor: An Exclusionary Rule in the Chicago Criminal Courts, 63 U. COLO. L. REV. 75 (1992). For additional studies on the deterrence effect of the exclusionary rule, see H. Mitchell Caldwell, Fixing the Constable's Blunder: Can One Trial Judge in One County in One State Nudge a Nation Beyond the Exclusionary Rule, BYU L. REV. 2006, at 1 (discussing different studies on the effectiveness of the exclusionary rule); see also William C. Heffernan & Richard W. Lovely, Evaluating the Fourth Amendment Exclusionary Rule: The Problem of Police Compliance with the Law, 24 U. MICH. J.L. REFORM 311 (1991); L. Timothy Perrin et al., If It's Broken, Fix It: Moving Beyond the Exclusionary Rule: A New and Extensive Empirical Study of the Exclusionary Rule and a Call for Civil Administrative Remedy to Replace the Rule, 83 IOWA L. REV. 669 (1998).}

First, Orfield studied the impact of suppression on the rate of conviction and concluded that the exclusionary rule does not often affect the prosecution of crimes because in most cases where evidence was suppressed, a defendant was still convicted based on alternative forms of evidence.\footnote{Orfield, supra note 212, at 78. Over the course of his research, Orfield determined that “[i]n Chicago, a jurisdiction with a comparatively high rate of suppression, unconstitutionally obtained evidence is suppressed under the exclusionary rule in only 0.9% of armed robbery cases, 0.5% of residential burglary cases, and 0.5% of cases involving violent crimes. Moreover, in many of the cases where evidence was suppressed, convictions were still obtained on the basis of other evidence.” Id.} Accordingly, the majority’s fear in \textit{Hudson} that suppression of the evidence is a “get-out-of-jail-free card” is not as accurate as the Court suggests.\footnote{See \textit{Hudson}, 547 U.S. at 595-96 (discussing the costs of applying the exclusionary rule).} Second, Orfield’s research of police department members established that they view the exclusionary rule as an advantageous institutional deterrent,\footnote{Ninety-eight percent of Orfield’s respondents believed that the exclusionary rule effectively deterred police misconduct. Orfield, \textit{supra} note 212, at 77, 84-85.} which they believe “does little harm to police work, and instead makes them more professional.”\footnote{\textit{Id.} at 81. Orfield’s research for his “Police Study” was gathered when he was a law student at the University of Chicago. Orfield “interviewed twenty-six of approximately one hundred officers in the Narcotics Section of the Organized Crime Division of the Chicago Police Department . . . using a 26-page standardized}
Accordingly, Orfield determined that application of the exclusionary rule actually promotes professionalism and internal discipline—the very things that the Hudson majority believed made the exclusionary rule unnecessary.\textsuperscript{217} His study showed that police and prosecutors design programs and create incentives to avoid the risk of suppression of evidence, which results in better compliance with the Fourth Amendment.\textsuperscript{218} Likewise, Orfield noted a similar positive reaction to the exclusionary rule in his study of judges, assistant public defenders, and assistant state attorneys.\textsuperscript{219} Orfield’s interviewees suggested that because “officers care about convictions and experience adverse personal reactions when they lose evidence . . .[,] police change their behavior in response to suppression of the evidence.”\textsuperscript{220} The impact of suppression on the police force in terms of professionalism and conduct led court respondents to proclaim, “that there is no more effective a remedy for Fourth Amendment violations . . . [and] they believe the rule should be retained.”\textsuperscript{221} Therefore, it is only through the threat of suppression that officers will curb their unlawful behavior and comply with the established law. As studies show
that the exclusionary rule prevents police misconduct and promotes the increased professionalism of law enforcement officials, it appears to be the only effective remedy to deter knock-and-announce violations.

VI. CONCLUSION

The knock-and-announce rule has withstood the test of time, with its common law roots dating back to thirteenth century England.\(^{222}\) In Wilson v. Arkansas, the Court recognized the rule’s resilience and importance by holding that it is a constitutional command necessary to safeguard an individual’s Fourth Amendment rights.\(^{223}\) Despite its ultimate holding in Hudson v. Michigan, the Court reaffirmed the main purposes of the knock-and-announce rule: “the protection of human life and limb” and the protection of “those elements of privacy and dignity that can be destroyed by a sudden entrance.”\(^{224}\) Accordingly, the security that the knock-and-announce rule affords cannot be overstated.

The Court’s decision in Hudson substantially weakens the knock-and-announce requirement by inviting the police to disobey it:

As repeated players, police have an incentive to comply with what the established laws allow them to do, rather than what the law explicitly instructs them to do. Through repeated interaction with the courts (and police disciplinary boards), police officers learn of—and respond to—judicially created incentives.\(^{225}\)

By mitigating the threat that potential evidence may be suppressed at trial, the Court has undermined the main legal incentive for officers to actually knock and announce.\(^{226}\) As the police response to weakened Miranda jurisprudence illustrates, when an original rule is severely undercut by exceptions, officers discover ways to get around the law.\(^{227}\) After Hudson, officers know that whether or not they abide by the knock-and-announce rule, evidence gathered will still be admissible at


\(^{225}\) Case Comment, supra note 174, at 714-15 (emphasis added).

\(^{226}\) See Hudson, 547 U.S. at 605, 610-14 (Breyer, J., dissenting). As paramilitary raids result in violence and injury, the Court’s decision in Hudson leaves the door open for an increase in this behavior.

\(^{227}\) See supra Part V.1.A.
trial. It is not enough for the Court to state that the knock-and-announce rule is still in effect.228 To prevent unlawful police behavior, it is crucial that the Court re-establish the importance of the knock-and-announce requirement. The only effective way to do so is through legal sanction, and experience has shown that the only effective legal sanction against police misconduct of this sort is that provided by the exclusionary rule.

Jessica M. Weitzman

---

228 Hudson, 547 U.S. at 602-03 (Kennedy, J., concurring). In his concurrence, Justice Kennedy states “The Court’s decision should not be interpreted as suggesting that violations of the requirement are trivial or beyond the law’s concern.” Id.

† J.D. Candidate, Brooklyn Law School, 2008; B.A., Emory University, 2005. Thanks to everyone at the Brooklyn Law Review for their hard work and to Professor Susan Herman for her guidance and insight, and a special thanks to my family for their endless love and support.